

Applying Messick's Framework
to the Evaluation Data of Distance/Distributed
Instructional Programs

by

Valerie Ruhe

B.Ed. Concordia University, 1977

M.A. The University of British Columbia, 1989

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Centre for the Study of Curriculum and Instruction)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October, 2002

© Valerie Ruhe, 2002

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Center for Study of Curricula & Instruction

The University of British Columbia
Vancouver, Canada

Date January 27, 2003.

Abstract

In the past twenty years, the literature of evaluation in distance education has evolved largely independently of the literature of program evaluation. A survey of evaluation models for distance instructional programs shows that these models have not included unintended consequences or value implications as explicit evaluation criteria. Consequently, using these models in program evaluation studies may tend to produce findings which are incomplete. Because it does include unintended consequences and value implications, Messick's (1989) framework on validity can be used to guide evaluation studies of distance instructional programs.

In this mixed methods study, I will take an adapted version of Messick's (1989) framework for a "test-drive" by applying it to authentic evaluation data from three BC post-secondary courses—Modern Languages 400, Psychology 101 and MCSE (Microsoft Certified Systems Engineer). Qualitative findings will then be compared with survey findings to obtain an in-depth understanding of the workings of the three implementation systems. My findings demonstrate that the adapted Messick's (1989) framework can be very useful in guiding the evaluation of distance programs because it provides a comprehensive assessment of merit and worth. Moreover, the application of this framework resonates with Stake's (1995) responsive approach to evaluation, so that applying the framework brings an easy-to-use and reputable approach to program evaluation into the field of evaluation of distance education.

Table of Contents

Abstract.....	ii
Table of Contents.....	iii
List of Tables.....	viii
List of Figures.....	x
Acknowledgements.....	xi
CHAPTER 1: Overview and Summary.....	1
Rationale.....	2
Purpose.....	4
Definitions.....	5
A Glossary of Definitions.....	7
CHAPTER 2: Literature Review.....	8
Program Evaluation: Definitions, Issues and Approaches.....	8
Definitions of Program Evaluation.....	8
The Role of Values.....	9
Epistemology.....	10
Stakeholders.....	11
An Overview of Program Evaluation Models.....	12
Stufflebeam's (2001) Overview.....	12
Pseudo-evaluations.....	12
Question- and methods-oriented approaches.....	13
Improvement/accountability-oriented approaches.....	16
Social Advocacy Approaches.....	18
The responsive approach.....	19
The constructivist approach.....	20
Deliberative Democratic.....	20
Utilization.....	21
Summary.....	21
A History of Evaluation of Distance/Distributed Courses.....	22
Quantitative Approaches.....	22
Problems with Quantitative Methodologies.....	23
Qualitative Studies.....	24
A history of qualitative studies.....	25
Mixed Methodologies.....	26
Strengths of mixed methodology studies.....	26
A history of mixed methodology studies in distance education.....	26
Cost-benefit Analysis.....	29
Accreditation Models.....	30
Quality assurance models.....	31
Summary.....	34
Unintended Consequences of Technology.....	34
Unintended consequences in educational contexts.....	35
Unintended social consequences.....	38
Positive unintended consequences.....	39

Evaluation Models for Distance/Distributed Instructional Programs	40
Overview	40
Gooler (1979).....	43
Rumble (1981)	44
Collis (1993)	45
Clark (1994a)	46
Bates' (1995) ACTION model.....	47
Van Slyke, Kitter & Belanger (1998)	47
Belanger and Jordan (2000).....	48
The CIAO Model	50
Weaknesses of Distance Education Evaluation Frameworks	51
Summary	53
CHAPTER 3: A Comprehensive Framework for Evaluation.....	55
Overview	55
Contribution of This Research	55
The Overlap between Validity and Evaluation.....	56
Validity and Validation.....	59
Messick's (1989) Unified Conception of Validity.....	62
The evidential basis and construct validity	64
The consequential basis: value implications	65
Moss's (1998a) reply	69
The Controversy over Unintended Consequences	71
Messick's response—Defining the term “unintended consequences”.....	75
Summary	76
Applying the Adapted Messick's (1989) Framework to the Evaluation of Distance/distributed Courses.....	77
Contributions of the Adapted Messick's (1989) Framework.....	77
The evidential basis of evaluation.....	78
The consequential basis of evaluation	80
Summary	81
CHAPTER 4: The Methodology for an Empirical Application	83
Overview	83
Background—The Response of Adult Learners Project.....	84
Design	85
Method Sampling	86
The Response of Adult Learners project	86
Sampling in this research	87
Participants.....	89
Ethical Issues	89
Consent.....	89
Stakeholder influence.....	90
Instrumentation	90
Interview Schedules	91
Procedures.....	92
Data Collection.....	92
Applying the Framework to the Data Using a Mixed Methodology	93

Summary	95
Validity Issues in Evaluation Research.....	96
Triangulation	96
Generalizability	98
Utility	98
Summary and Conclusion	98
CHAPTER 5: Results	100
Overview	100
Modern Languages 400.....	102
Learner Response	105
Flexibility.....	105
Materials	106
Cost/Benefit and Relevance	112
Cost	112
Relevance	115
Unintended Consequences	118
Insufficient flexibility	118
Mismatch between technology and subject matter	119
Conclusion.....	120
Psychology 101 (Print Distance Version).....	121
Course Overview	121
Course Materials	122
Learner Characteristics.....	122
Value Implications	123
Ideology.....	123
Course Objectives	124
Learner Response.....	124
Flexibility	124
Materials.....	125
Interaction.....	126
Support Services.....	127
Relevance and Cost/Benefit	129
Relevance	129
Costs.....	130
Benefits.....	131
Unintended Consequences	132
Updating the course materials.....	132
Interruptions in support services.....	132
Unintended Social Consequences	136
Traditional print courses as ghettos	136
Mismatch between the Online Forum and Enrolment Policies.....	138
Conclusion.....	140
Microsoft Certified Systems Engineer (MCSE)	141
Course Overview	142
Value Implications	143
Microsoft's ideology.....	143

Course objectives	143
Learner Response	146
Flexibility	146
Materials	147
Interaction	149
Support services	149
Relevance and Cost/Benefit	151
Relevance	151
Cost	153
Benefits	154
Unintended Consequences	157
Relevance--Mismatch with the needs of the market.....	157
Lack of access and flexibility	158
Adding in face-to-face components	159
Short product cycles and updating of the curriculum and tests	161
Eliminating qualified instructors.....	163
Continuous Enrolment.....	166
Conclusion.....	167
CHAPTER 6: Discussion.....	169
The Contribution of this Research	169
What Have We Learned from Applying the Adapted Messick's Framework?	170
The Novel Contribution of this Research.....	170
Borrowing from Assessment and Evaluation.....	171
Bringing Issues from the Background to the Foreground	171
Summary of the Findings	173
Unintended consequences.....	173
Value implications	174
Construct labels.....	174
Ideology	175
Theory.....	176
The multiple value bases of distance/distributed education	177
The tension between facts and values.....	178
How do Evaluators Deal with Multiple Values?.....	181
Summary	181
Insights from Social Advocacy Approaches and Responsive Evaluation	182
Social Advocacy and Program Evaluation.....	182
A responsive approach to evaluation of distance programs.....	182
Where Do We Go From Here?	184
The Need for Further Research	184
Is the adapted Messick's framework the last word?	186
Dissemination of the findings	186
Conclusion	187
References.....	189
Appendix A: Letter of Permission from Tony Bates	211
Appendix B: UBC Letter of Ethics Review.....	212
Appendix C: Informed Consent Form for Student Interview	213

Appendix D: Informed Consent for Faculty/Staff Interview	215
Appendix E: Questionnaire.....	217
Appendix F: Semi-structured Student Interview: Sample Questions	225
Appendix G: Semi-structured Faculty/Staff Interview Schedule	226
Appendix H: Sample Memo - February 24, 2001.....	227
Appendix I: Letter of Permission from Adnan Qayyum	229

List of Tables

Table 1	Overview of Distance Education Evaluation Models.....	42
Table 2	Overview of Case Studies.....	86
Table 3	Overview of Case Studies Selected for this Research	87
Table 4	Response to “I like this delivery mode because it gives me flexibility in my studies (e.g. time, place, location).”.....	105
	Response to “If this course was not offered in this delivery mode, I would not be able to complete it.”.....	105
Table 5	Response to “How do you rate the course materials?.....	106
Table 6	Support Services—Recommended Improvements	108
Table 7	Response to “I am not satisfied with the software used in this course.”	109
	Response to “The technology increases my motivation to work on the course.”.....	110
	Response to “The technology increases my motivation to work on the course.”.....	110
	Response to “I can learn better using print materials than by working on a computer”.....	110
Table 8	Response to “In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want.”.....	111
	Response to “In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want.”	111
Table 9	Response to “This course is not worth the money it costs.”.....	113
	Response to “I would not take another course using this delivery mode.”.....	113
Table 10	Benefits and Drawbacks of the Delivery Mode.....	114
Table 11	Response to “The course materials are relevant to my personal or professional needs.”.....	116
	Response to “Using technology in this course helps me to learn more relevant information.”.....	116
	Response to “Using technology in this course helps me to learn with greater depth of understanding”.....	116
	Response to “In this course, the interaction with the instructor is relevant to my learning”.....	116
	Response to “In this course, the interaction with the other students is relevant to my learning.”.....	117
Table 12	I like this delivery mode because it gives me flexibility in my studies, (e.g. time, place, location).	124
	Responses to “If this course were not offered in this delivery mode, I would be unable to take it.”	125
Table 13	Response to “How do you rate the course materials?.....	125
	Response to “I can learn better using print materials than by working on a computer	126
Table 14	Response to “In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want.”.....	126
	Response to “In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want.”.....	127
Table 15	Response to “Support services for this course are unsatisfactory.”.....	127

Table 16	Support Services—Recommended Improvements	128
Table 17	Response to “The course materials are relevant to my personal or professional needs.”	129
	Response to “In this course, the interaction with the instructor is relevant to my learning.”	129
	Response to “In this course, the interaction with the other students is relevant to my learning”.	
Table 18	Response to “This course is not worth the money it costs.”	130
	Response to “I would not take another course using this delivery mode.”	130
Table 19	Benefits and Drawbacks of the Delivery Mode.....	131
Table 20	Response to “I like this delivery mode because it gives me flexibility (e.g. time, place, location).”	146
Table 21	Response to “How do you rate the course materials?	147
	Response to “The technology increases my motivation to work on the course.” ...	147
Table 22	Response to “I am not satisfied with the software used in this course.”	148
	Response to “I can learn better using print materials than by working on a computer.”	148
Table 23	Response to “In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want”	149
	Response to “In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want.”	149
Table 24	Response to “Support services for this course are unsatisfactory.”	150
Table 25	Support Services—Recommended Improvements	150
Table 26	Response to “The course materials are relevant to my personal or professional needs.”	152
	Response to “Using technology in this course helps me to learn more relevant information.”	152
	Response to “Using technology in this course helps me to learn with greater depth of understanding”	152
	Response to “In this course, the interaction with the instructor is relevant to my learning”.	152
	Response to “In this course, the interaction with the other students is relevant to my learning”.	153
Table 27	Response to “This course is not worth the money it costs.”	154
	Response to “I would not take another course using this delivery mode.”	154
Table 28	Benefits and Drawbacks of the Delivery Mode.....	155
Table 29	Student Problems with the Course.....	156

List of Figures

Figure 1. Belanger and Jordan's framework of evaluation.....	49
Figure 2. The CIAO model of evaluation	51
Figure 3. Messick's (1989) unified conception of validity.....	63
Figure 4. An adapted framework for the evaluation of distributed courses.....	78

Acknowledgements

With his exceptional kindness and supervisory skills, and unparalleled expertise in all areas of measurement and evaluation, Professor Bruno Zumbo guided this dissertation to an efficient and speedy conclusion. Dr. Zumbo has been the most wonderful supervisor anyone could ever have. I would also like to express my heartfelt thanks to Professors Karen Meyer and Carl Leggo, whose post-modern vision of education has informed not only my understanding of program evaluation, but more importantly, of everyday life. Next, I would like to thank Dr. Stephen Petrina, who supported me in the beginning, by suggesting I re-analyze the OLT data, and in the end, as university examiner. I would also like to thank Dr. Hillel Goelman for reading this lengthy work and for his very detailed and helpful comments. Finally, I would like to express my deepest thanks to my parents, Lawrence and Phyllis Oszust, for their constant encouragement, home-cooked meals and many other sources of support. Without the kindness and support of all these people, this dissertation would never have been completed.

CHAPTER 1: Overview and Summary

In response to the recent expansion of distance education, it is time to critically assess traditional evaluation models for distributed instructional programs and to adopt a new model to guide the comprehensive assessment of the merit and worth. In the past 20 years, evaluation models have mostly consisted of checklists which have not included unintended consequences. Yet in the literature of program evaluation, educational technology, and quality assurance, unintended consequences are often mentioned as important components of quality in evaluating complex implementation systems. Although the distance education literature is replete with specific exemplars of how unintended consequences play themselves out in implementation systems, distance educators have tended to avoid using the term “unintended consequences” as an evaluation criterion.

Given the increasing complexity of distance instructional systems based on multiple technologies (Rumble, 1981), there is a pressing need to adopt a rigorous model of evaluation based on the science of program evaluation. Messick’s (1989) four-faceted conception of validity was developed to provide a comprehensive assessment of the merit and worth of standardized tests. This model includes features common to evaluation models for distance education courses (e.g., Bates, 1995; Gooler, 1979; Van Slyke, Kittner & Belanger, 1998), as well as a new category implied but seldom explicitly stated in the literature of distance education--unintended consequences.

The purpose of this study is to apply an adapted version of Messick’s (1989) framework to the datasets from three post-secondary distributed courses, that is, Psychology 100, Modern Languages 400 and Microsoft Certified Engineer, to demonstrate how the model works to guide the evaluation of distance/distributed

instructional programs. These three datasets, which consist of both qualitative and quantitative data, had previously been analyzed using a traditional evaluation model, Bates' (1995) ACTION model. In this study, I will write three case reports based on the adapted Messick's (1989) framework, and then compare the findings from this analysis to the findings from the original case reports based on Bates' (1995) ACTION model. I will demonstrate how specific findings in the background of the first case reports emerged to the foreground in the second analysis based on the adapted Messick's (1989) model. In this way, I will demonstrate how the adapted Messick's (1989) framework re-positions the findings to provide a comprehensive assessment of merit and worth.

Rationale

In distance education, there is a long history of questions/methods-driven evaluation studies focussed on evaluation categories such as technical quality of the course components and completion rates. In fact, evaluation models in distance education remain largely confined to the questions/methods-driven approach, which is just one category of approaches to program evaluation (Stufflebeam, 2001). In the contemporary educational *landscape* of distance education, there is a need for "proper evaluative studies" (Academic Committee for the Creative Use of Learning Technologies, 2000, p. 20) based on approaches and principles from the rich literature of program evaluation.

As we enter the new century, there is a proliferation of new and blended technologies for media-based instruction, an explosion of online enrolments, a blurring of conventional and distance education, the emergence of new for-profit educational providers and an emerging culture of learner choice (Frank, 2000; Scolfield, 1999). "Distance education has

come out of the closet,” and universities which do not “catch the wave” will be left behind (Frank, 2000, p. 12). In this new environment, there is a need for a new evaluation model which takes a comprehensive approach to investigating the merit and worth of distributed instructional programs.

With technology-based systems becoming increasingly complex (Rumble, 1981, Tenner, 1996), Rumble’s (1981) recommendation to analyze the gap between the ideal and the actual implementation is in effect a call to analyze unintended consequences. In the new context of global marketing of education, the challenge for contemporary distance course evaluators is to identify unintended consequences, so that course quality can be controlled and improved. Yet the term “unintended consequences” is hardly ever used in traditional evaluation models for distance education programs. A comprehensive assessment of merit and worth requires that unintended consequences be included as an evaluation criterion in models for distance education programs, and a cross-disciplinary approach to an appraisal of value based on the literature of educational measurement and program evaluation can bring new insights into the literature of distance education.

Messick’s (1989) four-faceted framework of validity was developed to guide the comprehensive assessments of the merit or worth of standardized tests. When viewed in the context of the rich literature of program evaluation, this model can be adapted and used to evaluate distance and distributed courses. An adapted framework can include categories commonly found in previous evaluation models for distance/distributed courses, and also fills a gap in the literature by including unintended consequences as an evaluation criterion.

Purpose

The purpose of my dissertation is to introduce and apply an adapted version of Messick's (1989) four-faceted conception of validity to the data from three distributed post-secondary courses in BC collected under the *Learning through New Technologies: The Response of Adult Learners (RALP)* project (Ruhe & Qayyum, 1999), and discuss emerging issues and implications, including how this approach works, what kinds of themes emerged and whether the framework needs to be adapted or refined. The contribution of my research is to demonstrate the benefits of applying the adapted Messick's (1989) framework to authentic data.

In Chapter 1, I will present the rationale for a new approach to evaluation and give an overview of my research. In Chapter 2, I will present the theoretical context, that is, key concepts in program evaluation, quality assurance, the unintended consequences of educational technology, the history of evaluation studies in distance education and the commonalities and shortcomings of distance education evaluation frameworks.

In Chapter 3, I will introduce Messick's (1989) framework on validity. First, I will discuss the four facets—construct validity, relevance and cost/benefit, value implications and unintended consequences. Next, I will discuss the philosophical underpinnings and issues which emerge from applying the framework in its original context, standardized testing. After that, I will identify the kinds of issues which might emerge from applying this framework to the distance/distributed learning context.

In Chapter 4, I will discuss the methodology involved in applying the adapted Messick's (1989) framework to the data collected for the Response of Adult Learners project, funded by the Office of Learning Technologies (OLT), Human Resources Canada.

Using a mixed methodology, I will cycle through the four constructs in the adapted Messick's (1989) framework, which will be operationalized through selected Response of Adult Learners questionnaire items and also used as coding categories to apply to the interview data. The focus is on the issues which emerge from applying the adapted Messick's (1989) framework to the data from three BC post-secondary distance courses—Modern Languages 400, Psychology 101 and MSCE (Microsoft Certified Engineer). In the Response of Adult Learners project, the data for these courses had been analyzed using Bates' (1995) ACTION framework, a traditional distance education evaluation model. As a researcher for that project, I wrote the case reports for a foreign language course which will be referred to as Modern Languages 400 (Ruhe, 1999a), and Psychology 101 (Ruhe, 1999b). The MSCE (Microsoft Certified Engineer) report was written by Adnan Qayyum (1999).

In Chapter 5, I will present three case reports based on applying the adapted Messick's (1989) framework to the data from Modern Languages 400, Psychology 101 and MSCE (Microsoft Certified Engineer). In Chapter 6, I will discuss the issues and implications which emerge from applying the adapted Messick's (1989) framework to the data. I will identify findings which moved from the background to the foreground and the specific ways in which the underlying assumptions of the adapted model worked to enrich the findings.

Definitions

In this section, I will define some key terms used in this research. First, though, I would like to discuss the use of two terms: 1) "distance/distributed" and 2) "model", as contrasted with "approach" and "framework". In this research, I tend to use

“distance/distributed” throughout because there is no umbrella term which encompasses both “distance” and “distributed”. The difference between the two terms is that distance courses use technology to bring flexible learning to off-campus learners, while distributed learning uses technology to bring flexible learning to on-campus learners (Oblinger & Maruyama, 1996). In this study, Psychology 101 is a distance course, while Modern Languages 400 and MCES are distributed courses. I have not used the term “technology-based education” because it is both too broad and does not carry the connotation of innovations conveyed by the terms “distance” and “distributed.”

In this research, Messick’s (1989) four-faceted conception of validity will be referred to as a “framework” rather than as a “model”. Madaus & Kellaghan (2000) define the term “model” as authors’ beliefs about “the main concepts and structure of evaluation work” (p. 19) which provide guidelines for arriving “at defensible descriptions, judgements and recommendations” (p. 20). “They are idealized or “model” views.” (p. 20). In contrast, Stufflebeam (2001) prefers the term “evaluation approach” to “evaluation model” because “the former is broad enough to cover illicit as well as laudatory practices” (p. 9). The implication is that “model” covers only laudatory practices. In their debate in Social Indicators Research, Moss (1998a), Messick (1998) and Markus (1998) seem to use the terms “model”, “framework” and “approach” interchangeably. For the sake of consistency and to avoid confusion, the term “framework” will be used in this research because as a set of concepts organized in a schematic diagram, it more closely describes Messick’s (1989) four-faceted conception of validity, and also avoids the positive connotations of the term “model” mentioned by Stufflebeam (2001).

A Glossary of Definitions

Educational evaluation—a systematic or “formal appraisal of the quality of educational phenomena” (Popham, 1993, p. 7).

Educational program--educational phenomena including “curriculum materials and other replicable instructional sequences” (Popham, 1993, p. 8). By this definition, a post-secondary distributed course is a kind of program and the principles of program evaluation apply.

Formative evaluation—“appraisals of quality focused on instructional programs that are still capable of being modified” (Popham, 1993, p. 13).

Normative theory—description of the program goals, intended consequences, program components and rationale; beliefs about how the program should work (Chen, 1990).

Program—“a set of resources and activities directed toward one or more common goals, typically under the direction of a single manager or a management team” (Wholey, 1987, p. 78).

Quality assurance—evaluation procedures to maintain high standards of quality

Stakeholders—People associated with or affected by a program, whether or not they have a say in its future, e.g. school administrators, teachers, parents, students and community groups (Weiss, 1986).

Summative evaluation—“appraisals of quality focused on completed instructional programs” (Popham, 1993, p. 13).

Validity--“an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (Messick, 1989, p. 13).

CHAPTER 2: Literature Review

In Chapter 1, I presented the rationale for a new approach to evaluation in distance education and gave an overview of my research. In Chapter 2, I will discuss program evaluation models, quality assurance approaches, the unintended consequences of technology, and the history of evaluation theory and practice in distance education. This review of the literature will also discuss the shortcomings of traditional evaluation models, the emphasis in the quality assurance literature on unintended consequences, and the need for an evaluation framework in distance education which, by including unintended consequences, provides a comprehensive assessment of merit and worth.

Program Evaluation: Definitions, Issues and Approaches

Definitions of Program Evaluation

Evaluation is an attempt to judge the worth, value or quality of something (Coldeway, 1988). Educational evaluation is “an enquiry which sets out to explore some educational programme, system, project or event in order to focus on its worthiness” (Bassey, 1999, p. 63). The “heart” of evaluation is a judgement of the overall value or worth of an endeavour (Wolf, 1987). Evaluation studies of educational programs investigate one or more of the following issues: a) program process, b) program outcomes, c) attributing outcomes to the program, d) links between processes and outcomes, and e) explanations (Weiss, 1998).

Program evaluation is applied research, and context is an important consideration in determining the type and extent of evidence which meets acceptable standards of proof (Davidson, 2000). Moreover, these applied studies are conducted in diverse disciplines and the literature of program evaluation reflects this diversity. The field of program evaluation, then, is a “trans-discipline” (Scriven, 1991); that is, it overlaps with other

disciplines. For this reason, program evaluation can transform evaluation theory and practice in other disciplines and in turn be transformed by them (Scriven, 1991).

There are two types of evaluation studies: *formative* and *summative* (Flagg, 1990). In a formative evaluation, value judgements are made during program development for the purpose of program improvement (Bassey, 1999; Melton, 1995). The evaluator gathers evidence on the efficiency of various components of the instructional program in order to isolate problems and remedy deficiencies in the program (Popham, 1993). In contrast, with summative evaluation, value judgements are made at the end of a program to determine success or failure (Popham, 1993). The focus, then, is on “appraisals of quality focused on completed instructional programs” (p. 13). In actual practice, however, the distinction between formative and summative evaluation is often unclear (Chen, 1990), especially for distance courses, where evaluation may be done at different stages of project development (Caulder, 1994a; Tennyson, 1997b).

The Role of Values

Values are central to evaluation studies, and programs are based on sets of values. (Popham, 1993). In its early days, the goal of educational evaluation was to determine whether educational objectives were being achieved (Worthen, Sanders, & Fitzpatrick, 1997). This approach, which originated with the work of Tyler (1942) and Hammond (1973), is referred to as a *goal-attainment* or *objectives-driven* approach. To measure educational outcomes, Tyler (1942) recommended an objective- and performance-based measures, with goals being translated into measurable objectives. Hammond (1973) recommends a detailed analysis of the impact of contextual (institutional and instructional) factors which are relevant to the attainment of objectives. Goals which were not achieved

were considered inadequacies in the program (Popham, 1993). Evaluators working within an objectivist view of underlying reality tend to seek definitive, unequivocal answers to evaluation questions (Stufflebeam, 2001). With the accreditation approach, for example, evaluators rate aspects of the program using lists of objective, predetermined criteria. Although the objectives-driven approach appeals to “common sense”, the disadvantages are that it does not focus on the process, especially side effects, and that the findings are too narrow to provide an assessment of overall merit and worth (Stufflebeam, 2001).

In the 1960's, educational evaluation evolved from an objectives-driven focus to include naturalistic approaches and value pluralism, with the evaluator working as a negotiator with stakeholders to interpret and use evaluation findings (Ross & Morrison, 1997). Today, evaluation takes place in a complex and diverse social context, with evaluators holding “different paradigms, perspectives and values”, conducting evaluations for different purposes, and taking on different roles in “a diverse array of practices” (Caracelli, 2000, p. 100). In the contemporary “landscape of values”, there is a need to bring moral discourse back into the evaluation of educational programs (Schwandt, 2000, p. 25).

Epistemology

In contrast to the traditional, objectivist epistemology underlying the scientific method, alternative epistemologies have informed contemporary approaches to program evaluation. Gadamer (1981) holds that there is no knowledge without preconceptions, and no single correct interpretation. Moss (1998a) believes that there are multiple interpretations which may have equal merit and which are “contextualized and perspectival” (p. 58). Complexity science holds that outcomes are uncertain, and that

unpredicted outcomes can “emerge” despite our best efforts to control them (Davis, 2002). Cook’s (1985) critical multiplism is based on the assumption that scientific knowledge is uncertain, and that the validity of inquiry can be enhanced by building diversity into the research process through the triangulation of diverse theories, values, measures and methods.

As for evaluation literature, several contemporary approaches to evaluation are grounded in postmodern values of subjectivist epistemologies, the validity of multiple perspectives and a “naturalistic” epistemology, rather than an epistemology based on prediction and control. Of these alternative approaches, perhaps the most notable is Stake’s (1995) responsive evaluation. Stake (1995) believes that every program is a case, and that cases should be explored qualitatively in a comprehensive manner in their unique contexts. The goal of evaluators is to ascertain the complexity of the particular relationships within each case (Stronach, 2001). The research process is an “art” which involves a tension between direction, that is, theories and methods, and indirection (Stronach, 2001), that is, the “ineffable nature” of the research task with unexpected questions which can “pop up” and send the inquiry in new directions.

Stakeholders

The Joint Committee on Standards for Educational Evaluation (1994), which sets standards of applied ethics of professional conduct, recommends that values be dealt with in the value-neutral framework of scientific detachment. In actual practice, evaluation research takes place within a political context, and the values underlying a program may or may not be shared by all stakeholders, who may have conflicting goals, values and expectations (Chelimsky, 1998). For this reason, responsiveness to stakeholders is an

important component of most evaluation approaches (Stufflebeam, 2001). When stakeholders disagree, the evaluator's choices include presenting conflicting views, working towards a consensus, and balancing the values of stakeholders with the views expressed in the literature (Chen, 1990).

An Overview of Program Evaluation Models

Stufflebeam's (2001) Overview

The following overview of contemporary program evaluation approaches is based on Stufflebeam's (2001) ranking of 22 different evaluation models, which represents the increasing diversity in evaluation approaches since the 1960's. The author classifies these approaches into four categories: 1) pseudo-evaluations 2) question- and methods-oriented approach 3) improvement accountability oriented approach and 4) social agenda/advocacy approaches. Except for pseudo-evaluations, these approaches represent "an increasingly balanced quest for rigor, relevance and justice... a strong orientation towards stakeholder involvement and the use of multiple methods" (Stufflebeam, 2001, p. 89). Using a checklist keyed to the Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation Standards, 1994), Stufflebeam (2001) includes three Social Advocacy approaches as "best" approaches for the new century.

Pseudo-evaluations

A "pseudo-evaluation" is a study which fails "to produce and report valid assessments of merit and worth to all right-to-know audiences" (Stufflebeam, 2001, p. 13). There are two types of pseudo-evaluations: public relations studies and politically controlled studies. Public relations studies present program strengths but not weaknesses, often giving false impressions. With politically controlled studies, information can be suppressed, polls and

files kept private and information withheld for political reasons. Both types tend to be “motivated by political objectives” (p. 13), including obtaining funding or hiding potentially damaging information. Stufflebeam (2001) lists the research methods which are used in pseudo-evaluations. These methods include

biased surveys; inappropriate use of norms tables; biased selection of testimonials and anecdotes; “massaging” of obtained information; selective release of only the positive findings; reporting central tendency, but not variation; cover-up of embarrassing incidents; and the use of “expert” advocate consultants” (p. 14).

The result is that findings may be shaded, selectively released or even falsified (Stufflebeam, 2001); in any case, they help the program “put its best foot forward” (p. 14). Pseudo-evaluations “mislead taxpayers, constituents and other stakeholders concerning the programs’ true value and what issues need to be addressed to make it better” (p. 14). Formative studies may work against program improvement, while summative studies may result in more funds being invested in unsound programs. In either case, pseudo-evaluations discredit the field of program evaluation, lower confidence in the evaluation profession, mislead decision-making and ignore or support injustice (Stufflebeam, 2001).

Question- and methods-oriented approaches

A question-oriented approach is focused on specific research questions, while a methods-oriented approach is focused on a specific methodology (Stufflebeam, 2001). These approaches include experimental studies, case studies, mixed method studies, cost-benefit analysis and theory-based evaluation. Question- and method-oriented approaches are “quasi-evaluation” models because the emphasis is on the question or method of evaluation, not on an assessment of the overall merit and worth of a program. The focus of

quasi-evaluation studies, then, is narrow or tangential to an assessment of overall merit and worth (Stufflebeam, 2001).

The experimental approach contrasts outcomes between experimental and control groups to assess the effects of a treatment, that is, the program. This approach was common during the 1960s and 1970s, when the U.S. government required an evaluation of federally funded social programs. With the “paradigm wars” between the advocates of quantitative and qualitative methods, this approach gradually fell into disfavour in part because it was perceived by some to be unethical, narrow and problematic because of the difficulty in controlling for intervening variables in educational contexts (Stufflebeam, 2001).

Another method-based approach is the evaluative case study, which is “a single case or collection of cases studied in depth to provide decision-makers with information on the worth of policies, programmes or institutions” (Stenhouse, 1988, p. 49). The purpose of a case study is to “delineate and illuminate a program, not necessarily to guide its development or to assess and judge its merit and worth” (Stufflebeam, 2001, p. 34). In case research, the evaluator uses coding categories to make the conceptual connections which constitute theory building (Miles & Huberman, 1994). Various strategies, such as triangulation, audit trails, and appropriate sampling methods, are used to enhance credibility. A single case can provide an in-depth, stand-alone “picture” of a specific program, or several cases be sampled from a range of scenarios to enhance generalizability (Reigeluth, 1999).

Mixed methodology studies are an outgrowth of the “paradigm wars” (Tashakkori & Teddlie, 1998). The underlying assumption is that combining qualitative and quantitative

approaches will result in greater validity, generalizability, and usefulness (Stufflebeam, 2001). A list of strategies for blending the epistemologies and value systems of the two paradigms is given by Tashakkori & Teddlie (1998). According to Stufflebeam (2001), qualitative and quantitative methods “can complement each other in ways that are important to the evaluation’s audience” (p. 41) and the consideration of mixed methods is “almost always appropriate” (p. 41).

A cost-benefit analysis is a set of quantitative procedures to determine the ratio of investments to social benefit. Cost-effectiveness combines relative measures of outcome and cost so that alternative programs or policies can be compared (Levin, 1983). It differs from cost-benefit analysis in that monetary units are not required, and therefore is more appropriate to the context of education (Levin, 2001). Costs can be determined by summing various “ingredients” which make up total costs, including facilities, equipment and client inputs (Levin, 2001). Cost is defined as the value of alternative uses of resources which are given up; for example, the value of a distance teacher is the value of the face-to-face section the teacher would have taught (Levin, 1983). According to Simpson (1991), outputs are “proxies” for benefits, and include quantity components such as the number graduating in different categories and quality components such as acquisition of “basic transferable skills, “appreciation of cultural diversity” (p. 25). A benefit is gained from the output to the extent that a goal is attained, and benefits which are intangible or difficult to measure are no less important than tangible benefits (Simpson, 1991).

Finally, the theory-based approach to evaluation is based on the belief that program evaluation should begin with a theory of how the program is supposed to work. The program theory is then used to guide the evaluation (Bickman, 1987; Chen, 1990; Rogers,

Petrosino, Huebner & Hacsí, 2000). A program implementation theory, for example, might focus on why the program was or was not delivered as intended or it might highlight areas for program improvement (Scheirer, 1994). Over the last thirty years, there has been growing interest in theory-based evaluations in health sciences and social work (Chen, 1990).

Two benefits of theory-based evaluations are: 1) bringing the attention of evaluators to unintended consequences and 2) using logic models of the program theory to involve stakeholders in the evaluation process (Chen, 1990). On the negative side, some theorists believe that many programs are not based on theories, that the conceptualisation of a program theory requires “a lot of muddling around” and that the approach can lead to conflict among stakeholders committed to different models (Stake, personal communication, April 27, 2000). Stufflebeam (2001) also believes that the theory-based approach has little to recommend it, and in education, there are few theory-based evaluation studies (Lipsey, Crosse, Dunkle, Pollard, and Stobart, 1985).

Improvement/accountability-oriented approaches

The third category is the improvement/accountability-oriented approach. The goal of the evaluation studies in this group is to “provide a knowledge and value base for making and being accountable for decisions that result in developing, delivering, and making use of cost-effective services” (Stufflebeam, 2001, p. 56). Various quantitative and qualitative assessment tools and multiple data sources are used to obtain a comprehensive assessment of overall worth or merit of the program in which all outcomes, including unintended ones, are identified and assessed. These evaluation approaches are based on an objectivist view of underlying reality, seek definitive, unequivocal answers to evaluation questions, and

focus on using the findings for program improvement (Stufflebeam, 2001). The approach stresses independent and objective assessment and is grounded in ethical notions of the common good and benefit to society (Stufflebeam, 2001).

The three approaches in this category are decision/accountability, consumer-orientation and accreditation (Stufflebeam, 2001). Respectively, these approaches emphasize improvement of services, consumer reports of optional programs and the merits of programs offered by competing institutions (Stufflebeam, 2001). Stufflebeam, Foley, Gephart, Guba, Hammond, Merriam and Provos's (1971) CIPP approach, for example, focuses on obtaining information on *context*, *input*, *process* and *product* for judging decision alternatives. With the consumer approach, the consumer's welfare is the ultimate value, and that evaluators should make a judgement about the relative merit of competing products and services. Scriven's (1972) goal-free evaluation is one example of a consumer approach. Scriven recommends that evaluators ignore statements of intended effects and focus only on actual effects. The reason is that statements of intent constitute a "rhetoric of intent," often couched in the fashionable jargon of current trends, which are often used "as a substitute for evidence of success" (p. 7).

The concept of quality assurance has been adopted from industrial models of evaluation into evaluation models of face-to-face courses (Forsyth, Jolliffe & Stevens, 1995). In educational contexts, quality assurance is linked with quality control, which refers to tools designed to collect information such as accreditation checklists (Forsyth et al., 1995). Accreditation is a process "whereby an organization grants approval of institutions such as schools, universities and hospitals" (Worthen et al., 1997, p. 173). Accreditation studies address the question of whether the program is meeting established

standards of quality, and often provide recommendations for program improvement. The purpose is to determine whether programs should be certified or whether institutions should be approved to deliver these programs. Accreditation studies help individuals to make informed judgements about the quality of educational services from competing providers. An accreditation approach stresses the professional judgement of the evaluator, often in a formal professional review system using guidelines and criteria developed by a professional accrediting body (Stufflebeam, 2001). Traditionally, accreditation focused on quantitative measures of facilities, staff qualifications and appropriate process (Worthen et al., 1997), and several current systems “aspire to justify their criteria on the basis of empirical links of inputs and processes to outcomes” (p. 123).

Social Advocacy Approaches

Stufflebeam (2001) recommends four social advocacy approaches as among the strongest and most promising for the 21st century. Social advocacy approaches are based on a subjectivist epistemology, which holds that there are no best answers or clearly preferable values. These approaches

favor a constructivist orientation and the use of qualitative methods. For the most part, they eschew the possibility of finding right or best answers and reflect the philosophy of postmodernism, with its attendant stress on cultural pluralism, moral relativity, and multiple realities. They provide for democratic engagement of stakeholders in obtaining and interpreting findings (p. 62).

The role of the evaluator, then, is to document the multiple realities of all participants with first-hand experience of the program, including teachers, administrators and taxpayers. With a pluralistic approach, the evaluator’s role ranges from facilitating the reconciliation

of different perspectives to taking a “hands off” approach by giving stakeholders control over the study and letting them decide how to handle values. “Clients must also be receptive to ambiguous findings, multiple interpretations, the employment of competing value perspectives, and the heavy involvement of stakeholders in interpreting and using findings” (Stufflebeam, 2001, p. 70).

Based on a subjectivist epistemology, evaluators may be less interested in finding the “right answer” than in gathering multiple perspectives, all of which may be equally valid. By rejecting an “unquestioned, singular value base” (Stufflebeam, 2001, p. 91), these approaches “reflect the philosophy of postmodernism, with its attendant stress on cultural pluralism, moral relativity, and multiple realities” (p. 62). Not only do diverse value perspectives reflect the complexity and diversity of contemporary social realities, but they also enhance the credibility of findings obtained in political environments, and help to ensure that these findings are used correctly and not misused (Chelimsky, 1998). This umbrella category includes Stake’s (1983) responsive approach, constructivism (Preskill & Torres, 2000), House and Howe’s (2000) deliberative democratic model and Patton’s (1997) utilization model.

The responsive approach

With Stake’s (1975) client-centered or responsive approach, the role of the evaluator is to document the multiple realities of all participants involved with the program, including teachers, administrators and taxpayers. The evaluator is also committed to identifying intended outcomes and comparing them to actual outcomes, thereby uncovering side effects. Parlett and Hamilton’s (1977) illuminative approach also uses a qualitative approach to evaluate innovative educational programs. The emphasis in both approaches is

on using qualitative data to illuminate both intended and unintended consequences of program implementation.

The constructivist approach

The constructivist approach is based on the assumption that evaluations are never value-free. Knowledge is constructed by individuals from diverse perspectives, and is believed to be problematic, subjective and changing (Stufflebeam, 2001). A constructivist epistemology involves an iterative process grounded in constructivist communities of evaluation practice (Preskill & Torres, 2000), in which stakeholders “play a key role in determining the evaluation questions and variables” (Stufflebeam, 2001, p. 71). The evaluator’s role is to present the diverse constructions of various stakeholders, make sense of them and work towards a consensus. The ultimate goal of evaluation is to empower the disenfranchised and change society for the better.

Deliberative Democratic

With deliberative democratic evaluations, the focus is on democratic participation, dialogue to assess stakeholders’ views and the negotiation of a credible assessment of overall worth (House & Howe, 2000). The equitable participation of stakeholders at all stages in the process is critical and power imbalances are unacceptable (Stufflebeam, 2001). Multiple methods such as discussions, surveys, debates and negotiation are used to obtain stakeholder participation and reach a defensible assessment of the program. Unlike other social advocacy approaches, this one recommends that evaluators reject input if it is invalid or unethical (Stufflebeam, 2001). Finally, the evaluator is responsible for reaching a final judgment on the worth of a program.

Utilization

Utilization evaluations are based on the assumption that evaluations are conducted to provide information for use by decision-makers (Caracelli, 2000), and the focus of the utilization approach is on assuring that program evaluations make an impact (Patton, 1997). The evaluator works with a select group of stakeholder representatives to clarify values, determine questions, investigate contextual dynamics, triangulate findings from different sources and determine the uses to be made of the findings. All aspects of the evaluation are geared towards maximizing the chances of applying the findings to their intended uses, and stakeholder consultation is important in furthering the change process (Stufflebeam, 2001). When evaluation evidence is presented to funding agents to justify funds given or to support a request for more funds, an investigation of unintended program and social side effects is necessary (Henry, 2000).

The term *utilization* embraces value pluralism (Galston, 1999) and multiple methods and roles for evaluators (Caracelli, 2000). One use of evaluation is process use--the learning in individuals, teams and organizations which results from participation in the evaluation process (Patton, 1997). For Henry (2000), the purpose of evaluation is social betterment, defined as improved social conditions, fewer social problems and a reduction in human distress and suffering. Stufflebeam (2001), however, notes that the utilization approach does not necessarily advocate any particular social or moral agenda.

Summary

Educational program evaluation is a study to determine the worth or value of educational services. In the last half of the twentieth century, diverse evaluation philosophies, theories, values and practices have resulted in a more pluralistic

understanding of evaluation use (Caracelli, 2000). In his overview of 22 approaches, Stufflebeam (2001) notes that all approaches have strengths and weaknesses, but recommends four social advocacy approaches, based on a subjectivist epistemology and multiple perspectives, as the strongest and most promising for the 21st century.

A History of Evaluation of Distance/Distributed Courses

Quantitative Approaches

In traditional outcomes-based studies, the focus was on quantitative methodology, that is, using statistical hypothesis testing to compare outcomes between face-to-face classes and distance classes (Organization for Economic Co-operation and Development, 1999). Along with a quantitative methodology, there are also many surveys for evaluating distributed courses (e.g. Cheung, 1998; Robson, 2000; Tessmer, 1993). The Flashlight Project, conducted under the auspices of the American Association for Higher Education (AAHE), for example, comprises a range of survey items and assessment tools to help institutions evaluate technology-based educational practices (TLT Group, 2000).

The goal of *equivalence* studies is to demonstrate that the value provided by a distance course is equivalent to the face-to-face version of the same course. In experiments comparing the two versions, questionnaires are used to measure variables such as outcomes, learner satisfaction, perceptions, study habits and attitudes towards technology; statistical significance tests are then performed to determine equivalence (Russell, 1999). Upon finding no statistically significant differences, the authors of these studies conclude that the distance method is “equivalent” in value to the classroom method (Organization for Economic Co-operation and Development, 1999; Russell, 1999). The contemporary version of this argument is that powerful, modern telecommunications systems provide an

even closer “replication” of face-to-face classroom interaction (Simonson, Schlosser, and Hanson, 1999).

Problems with Quantitative Methodologies

There is a plethora of methodological problems with a quantitative approach to evaluating distance programs, including lack of true control groups, random assignment, pre-tests to equalize individual differences and control for confounding variables (Institute for Higher Education Policy, 1999). Other problems are small sample sizes, misinterpretation of the results of significance testing, a proliferation of Type I errors, novelty effects (Russell, 1999; Suen & Stevens, 1993). When a priori power analyses are omitted, as they typically are in these studies, findings of no statistically significant differences do not justify the inference that the treatments are equivalent (Suen & Stevens, 1993). Because power is a function of sample size, the design may have had insufficient statistical power to detect differences, but a design with a larger sample size might have detected a difference (Rosnow & Rosenthal, 1988).

Another issue, which is relevant to distance education research, is the design of survey items and the reporting of the psychometric properties of scales (see Meier & Davis, 1990). Most distance education research do not mention reliability or validity coefficients, and Cheung’s (1998) survey is one of few that have performed well on tests of validity and reliability. A related issue is that checklists evaluating software are distinct from how the technology is actually used by learners (Jones, Scanlon, Tosunoglu, Ross, Butcher, Murphy & Greenberg, 1996).

For these reasons, it is difficult to establish that computer media are the cause of any enhancement of student learning (Lookatch, 1997; Ungerleider & Burns, 2002, April-May). Because of the complexity of the distance learning context and process, both construct validity and the meaning of the findings are jeopardized and the conclusions of many of these kinds of studies are unwarranted (Gibson, 1998). According to Jones et al. (1996), different rationales and different types and uses of computer-assisted learning “will require entirely different ways of evaluation which will address the extent to which the innovation has achieved what it set out to do” (p. 12). These authors recommend an investigation of the context and learning process in addition to outcomes.

Qualitative Studies

There are few qualitative studies of distance education, although Selwyn (1997) has called for more qualitative studies to shed light on how distance courses actually work, as opposed to how they should work. In contrast to outcomes-based evaluations, qualitative studies investigate learning from the perspectives of the participants, e.g. the course designers, instructors and learners (McCulloch, 1997). Questions to participants can focus on learner response to course components which helped or hindered learning (Reigeluth and Frick, 1999).

They should ask ... what they did and did not like about the various elements of the instance, what helped them, what did not help them, whether they felt that the materials and activities were appropriate for their needs, what changes they would make if they could and whether they felt they had attained the objectives (p. 641).

Qualitative techniques can help to identify key strengths and weaknesses in the implementation process which might not surface in quantitative studies. Herrman, Fox and

Boyd (1999), for example, did a case study of unintended consequences of technology to counter “the long history” of “uncritical introduction of educational technology into the education context” (p. 3). By bringing in multiple perspectives, qualitative research can lend both breadth and depth to the discussion of value.

A history of qualitative studies

There are few published qualitative evaluation studies in distance education, and Selwyn (1997) suggests there is a need for more of them. Melton’s (1995) case study of the Open Junior High School system in Indonesia is a process evaluation in the naturalistic tradition. Andrusyszyn and Davie (1997) conducted a qualitative study to examine the thinking of learners who engaged in interactive reflective journal writing with a course instructor. The authors analyzed electronic transcripts of online discussions and conducted interviews with learners after course completion. The authors found that reflection through journal writing offered a valuable means for the transformation of knowledge to occur.

McCulloch (1997) did a case study of participatory evaluation, which embedded evaluation in the learners’ experiences of their tutorial activities. McAlister’s (1998) ethnographic study of 36 mature learners at the Open University (OU) explored the dynamics among individual, social and institutional factors which affected the outcomes for learners with “low” qualifications. Finally, using unstructured interviews, Henderson and Putt (1999) did a case study of different uses of audio-conferencing, including the “effectiveness of implementation strategies and the various roles of the participants in a cross-cultural context” (p.25).

Mixed Methodologies

Strengths of mixed methodology studies

Studies based on mixed methodologies employ both quantitative and qualitative methods as equal and parallel or as dominant/subordinate methods (Creeve & Caracelli, 1997; Tashakkori & Teddlie, 1998). Mixed methodologies are effective because they provide different tools to study different aspects of the distance/distributed context. Studies based on mixed methodologies generate more and different types of data, which can be used to investigate both the macro- and the micro-context. The following studies are examples of evaluation studies in distance education using mixed methodologies.

A history of mixed methodology studies in distance education

Kanuka and Anderson (1998) conducted an exploratory multi-method study and transcript analysis of an online forum using a constructivist interaction analysis model. A constant comparison method was used to recode transcript messages. These authors found that time engaged in social discourse tended to generate social discord, which served as a catalyst to the knowledge construction process. In addition, a survey was used to assess learners' perceptions of whether, and if so, to what extent, online discussion groups reflected a constructivist model of learning communities.

The British Open University (OU) has a long and impressive track record of developing a broad range of diverse combinations of mixed methodologies, and evaluation studies based on a variety of mixed methodologies (Jones et al., 1996). In 1973, the OU launched a five-year project to evaluate the success of 250 qualified and 250 unqualified school leavers in first-year university courses (Woodley & McIntosh, 1980). The study focused on variables related to learner characteristics, e.g. motivation, self-discipline and

environmental context, e.g. employment, and domestic environment. Interviews, questionnaires and a battery of psychometric tests were used. The goal of the study was to determine how younger learners who persisted fared and how others fared and why they withdrew. This study found that younger learners faced financial difficulties and time pressures and found it difficult to “play the system” or to put their studies ahead of their other commitments. They found that the younger learners did less well in the first year, but that those who were successful in year 1 usually went on to complete. There were no differences in ability across age groups, but for younger learners, success was related to ease of access, support of friends and certain personality characteristics.

The first coordinated attempt to evaluate the OU’s computer-assisted learning component was undertaken in 1979 (Jones, Scanlon, Tosunoglu, Ross, Butcher, Murphy, & Greenberg, 1998). Data collected through interviews and journals were used to construct items in a survey distributed to 2,000 respondents. The purpose was to determine to what extent the interview findings could be generalized to a larger population. The study found that student interviewees had three reasons for not using optional tutorials: 1) fear of looking “stupid,” 2) fear of breaking the software and 3) fear of being spied on. Results of the questionnaire confirmed the researchers’ suspicion that these fears were widespread.

In the 1980s, the OU adopted the Home Computing Policy whereby learners could use personal computers from their homes to access their courses (Jones, et al., 1996). A large multi-dimensional evaluation involving linked projects was done to determine whether the costs to learners of purchasing and maintaining personal computers were worth the benefits. Again, a mixed methodology was used and the data was collected through

interviews, surveys and records of student usage patterns. Because of low response rates, however, data from journals and diaries were not used.

The OU conducted a 1989 formative evaluation project for a Living with Technology course (Jones et al., 1996). Learners using the materials were observed and interviewed, materials were revised and learners using the revised materials were surveyed three times during the year. The response from learners were consistently positive and learners felt that the time they had spent on learning how to use the computer had been worthwhile in terms of the benefits received.

Jones and Petre (1994) studied learners in a Computers and Learning course. This study used a mixed methodology, with questionnaires and student journal entries about significant events. These authors found unanticipated consequences such as learners reading the manual only as a last resort and a mismatch between learners' working styles and the assumptions of the instructional designers. They also had problems locating material on audiotapes. In this study, it was found that some means of tracking learners' work in progress, either through ongoing observations, interviews or journals, was essential.

Jones et al. (1996) also did a study to determine the kinds of problems being experienced by learners as they were using the materials on phase diagrams with Works Metallurgist software. A preliminary questionnaire to measure attitudes and prior knowledge of phase diagrams was sent to 110 learners. Of these, 50 learners were then sent 1) a special software evaluation disk which recorded dates of use, sections used, time taken and number of errors and 2) a follow-up questionnaire containing post-attitude and achievement tests. The other 60 learners received an extended follow-up questionnaire but

no disk. The results of the two groups were then compared. The study found that beneficial effects could not be attributed to the program in isolation from a much larger learning environment. A second finding was that some learners were arriving at the right answer through the wrong method (Scanlon et al., 1998).

Scanlon, Jones, Barnard, Thompson & Calder (2000) conducted two studies: 1) the Driven Pendulum, a physics simulation of chaotic motion and 2) the Galapagos tutorial, a multimedia CD-Rom. The purpose, to determine the role played by simulations in student learning, was both formative and summative. Observations, questionnaires, mini-quizzes, records of interactions with the software and interviews were used. This study found that it was important that the narrative structure be kept constant over the tutorial activities. Perhaps the most important outcome of these Open University studies for our purposes, however, is that they were the basis of an evaluation model, which will be discussed further on.

Cost-benefit Analysis

Cost-benefit analysis is an evaluation tool designed to make decisions by comparing the benefits or outcome for each alternative with its costs. Costs can be categorized as fixed (e.g. technology and course development) or variable (e.g. tutor marking and Internet connections), and start-up costs for instructional technology can be considerably higher than with face-to-face formats (Knapper, 1980). As for benefits, the monetary value of educational benefits to society such as a university degree can be difficult to quantify (Levin, 2001).

Despite their usefulness, there are few cost-benefit studies in education (Levin, 2001). Evaluators need to develop new costing models for the distributed learning environment to determine the impact on infrastructure requirements when learners do not

occupy the 'seat' used in traditional cost estimation models (Belanger & Jordan, 2000). Bates (1995) proposes calculating average cost per student study hour for a given technology times the number of students over the life of a course, the grades, learner satisfaction ratings or the number of students who complete the course. Another method is to use a detailed break-down of costs at the student level (Bartolic-Zlomislic & Bates, 1999). Levin's (1983) "ingredients" approach could also be applied to distance education courses.

At a time of shrinking budgets, productivity improvements are important aspects of added value. A cost-benefit analysis can determine whether productivity gains should be made by enhancing quality of small-scale delivery methods or by increasing access at reduced cost (Garrison & Anderson, 1999). A related notion is the "replace-ability challenge", that is, are there other media or set of media attributes which yield similar learning outcomes with comparable cost structures and opportunities for access (Keegan, 1993)? For this purpose, baseline measures of the set of conditions which are being replaced by technology should be used (Clark, 1994a). From here, I will go on to discuss the quality assurance approaches, which emphasize control of quality, costs and outcomes.

Accreditation Models

Accreditation is carried out by associations of schools whose representatives visit and evaluate the program on the basis of a checklist (Popham, 1993). For some time, regional accrediting associations have been encouraging their members to devise plans for establishing and evaluating distance programs (Dasher-Aston & Patton, 1998).

Accreditation has served as a mechanism to evaluate and validate an institution's commitment to planning, continuous improvement, program integrity, and educational effectiveness. Institutions can also use the interregional guidelines as a starting point towards quality assurance. (p. 14)

The ACE Distance Learning Evaluation Guide (American Council on Education, 2002) provides a detailed checklist for quality of the following five categories: (1) learning design, (2) learning outcomes, (3) technology, (4) learner support, and (5) organizational commitment. Distance learning should be consistent with the mission statement and policies of the institution. Learning design includes fit between the learning activities and the context, including the needs of learners. Support services should be comprehensive and readily accessible, and a reward system for faculty should encourage continuous professional development.

The Western Interstate Commission for Higher Education has established several principles of good practice for distance education (Johnstone & Krauth, 1996). First, student capability to succeed in distance education programs should be related to admissions and recruiting policies and decisions. Secondly, the outcomes of distance education programs such as learning outcomes, student retention, and student satisfaction should be comparable with those of campus-based programs. Thirdly, the integrity of student work and the credibility of the degrees and credits awarded should be ensured.

Quality assurance models

The quality assurance approach encompasses issues such as standardization of products and services, fit among the course and context variables and relevance, currency and transfer of the learning materials to authentic contexts outside of the classroom

(Forsyth, et al., 1995). According to Mann (1998), quality assurance models include quality of curriculum, quality of interaction, customer satisfaction, independent and external evaluation, and turn-around time (time between assignment submission and receipt of feedback).

One aspect of quality assurance is standardization of products, services and modes of assessment. Universitas 21 (2000), a consortium of 17 leading research universities in 10 countries with 500,000 learners, offers a highly recognized pre-eminent brand name for educational products supported by a proven quality assurance capability. Neilsen (1997) recommends a quality assurance approach to distance teacher education. The Commonwealth of Learning's Writing Effectively for UNHCR course is an example of a globally delivered course which uses quality assurance procedures to standardize the quality of tutor marking.

Another issue from a quality assurance perspective is the "fit" among the various components and the environment. Distance education has unique characteristics including: 1) market and cost analysis, 2) student support system, 3) media selection, 4) delivery and 5) student assessment (Bourdeau & Bates, 1997). The distance context includes society, the organization, the target group, the course characteristics, and the technology characteristics; these elements can vary considerably and affect course implementation in various ways (Bates, 2000). Because contextual variations may affect the choice of media, cost structures and teaching approach (Bourdeau & Bates, 1997), the "fit" among these contextual variables is an important aspect of the quality of distance programs (American Council on Education, 2002).

Some evaluation studies have been done which have focused on this notion of fit among contextual variables. For example, Mann (1998) attributed the low drop out rate of the Surrey MA (TESOL) program at the University of Surrey (UK) to the success of their admissions procedures and process. Similarly, Smith and Smith (1999) investigated the “fit” between the cultural characteristics of learner groups and the teaching approach. Finally, Woodley and McIntosh (1980) studied the relationship between learner characteristics and success rates.

Another aspect of quality is relevance to the needs of society. In *Charting a New Course*, the Ministry of Advanced Education, Training and Technology, Government of British Columbia (2000) expressed concern that the workforce is not undergoing the constant skills retraining and upgrading which are needed for BC to make the transition to a knowledge-based economy. “Employers, in particular, are concerned that workers lack the transferable skills to adapt quickly to new work opportunities, and that there appears to be a mismatch between the skills of current workers and those required to obtain better jobs in the changing economy” (p. 10).

Constructivism is a new approach to teaching and learning which address issues of relevance, authenticity and transfer (Duffy & Jonassen, 1992; Willis, 2000). In educational technology contexts, constructivism provides the foundations for cognitive complexity, meaningfulness (authenticity), transfer, generalizability (Linn, Baker, & Dunbar, 1997) and cognitive efficiency (Cobb, 1997). Two examples of using technology to implement an innovative constructivist approach to learning is Harasim, Hiltz, Teles & Turoff’s (1996) learning networks and De Jong and van Joolingen’s (1998) scientific discovery learning with computer simulations. The concept of

“distributed cognition” can also provide many insights into how computer technology can be used to usher in new forms of teaching and learning (Salomon, 1993).

Summary

As the previous history demonstrates, evaluation studies in distance/distributed learning largely reflect a questions- or method-oriented approach to evaluation. The history of evaluation studies is mostly quantitative, followed by a recent shift towards qualitative studies. Cost-benefit, quality assurance and accreditation models are also used to guide evaluation studies of distance programs.

Unintended Consequences of Technology

Tenner (1996) makes the case that complex systems will inevitably generate unintended consequences, of which there are four kinds: *bugs*, *productivity paradoxes*, *side effects*, and *revenge effects*. A *bug* is a small mechanical glitch which is often experienced by users. A *productivity paradox* is a situation where huge investments in technology produce little, no or even negative gains in productivity, even though costs tend to be high, stable or increasing (Fahy, 1999). One example is the unrealized dream that computers would lead to the paperless office (Tenner, 1996). A *side effect* is an unanticipated consequence that is less central to the desired effects. A *revenge effect* is some negative outcome of technology which undoes the predicted benefits, for example, carpal tunnel syndrome. A revenge effect is not produced by technology alone. Only when technology is anchored “in laws, regulations, customs and habits does the irony reach its full potential” (p. 7).

There are five kinds of revenge effects: *rearranging*, *repeat*, *recomplicating*, *regenerating*, and *recongesting effects* (Tenner, 1996). A rearranging effect occurs when technology is applied to improve some condition, but has the effect of degrading related conditions, thereby outweighing any gains. For example, air conditioning on subway cars merely displaces heat onto the subway platform, thereby increasing temperatures for waiting passengers. Similarly, MacIntosh (1992) describes how the installation of private telephone service in the Australian outback lead to the opposite of what was intended— a severe loss of a sense of community. This happened when medium wave radio, which had allowed people in remote areas to speak in groups for short, specified times of the day, was replaced by more private, one-on-one telephone service.

A *re-complicating* effect occurs when technology which was originally designed to make a task easier actually makes the task more complicated, e.g. moving from rotary to keypunch phones made dialing easier at first, but then lead to a proliferation of phone numbers, phone numbers of increasing length, and with voice mail, push button sequences of increasing length and complexity (Tenner, 1996). A *regenerating* effect is when a technological solution to a problem actually increases the probability of a negative outcome. Finally, a *re-congesting* effect is when the congestion from an increasing number of users has the effect of reducing access and efficiency (Tenner, 1996).

Unintended consequences in educational contexts

Educational situations have been conceptualised as “systems” (Banathy, 1998; Romizowski, 1981; Tennyson, 1997a). In face-to-face classes, unintended consequences include disruptive behaviour, negative response to the learning event and inability to transfer learning (Forsyth et al., 1995). Innovations also tend to produce unintended

consequences (Parlett & Hamilton, 1977). "The introduction of an innovation sets off a chain of repercussions in the learning milieu. In turn, these unintended consequences are likely to affect the innovation itself, changing its form and moderating its impact" (p. 12).

Different kinds of unintended consequences can emerge when innovative educational technology is implemented in innovative classrooms (Fabos & Young, 1999). These authors refer to "overly optimistic claims" (p. 218) and "larger corporate motives" (p. 218), eurocentric notions of other cultures, lack of quality control in email-based writing lessons, and malfunctioning classroom email exchanges. Braun (1994) mentions unintended changes in the teacher-student relationship after technology is introduced. Another example is the misapplication of innovative technology so that it merely replicates traditional classroom activities (Gray & O'Grady, 1994).

The presence of the visual link per se does not necessarily improve the lesson's educational effectiveness. Often, it simply served to demonstrate that old practices which are ineffective in a mainstream classroom can be just as ineffective using this technology ... this style of teaching only served to simulate a classroom environment in which the technology generated a safety net for producing comfortable images of mediocre classroom practices. (p. 668)

Distance education systems are increasingly perceived as complex systems with a variety of components which are dynamic and interactive (Farhad, 1997; Moore & Kearsley, 1996). To evaluate a distance course, the gap between the ideal and the actual implementation should be assessed (Rumble, 1981). This "gap" is another name for unintended consequences. In fact, high attrition rates and bipolar distribution of grades have characterized some North American distance programs since the days of Personalized

Systems of Instruction (PSI) and teaching machines (Keegan, 1993; Knapper, 1980).

“Indeed, one significant advantage that appeared to exist for technology-based systems—that of cost savings—turned into a potential disadvantage when it transpired that most systems were ‘added on’ as enrichment for conventional education, rather than supplanting it” (Knapper, 1980, p. 137). Jones et al.’s (1996) ethnographic evaluation of software “in use” found that learners used pathways through knowledge systems in different ways from those intended by course designers. In technology-based systems, then, there is evidence for side effects and productivity paradoxes.

In addition, Herrmann, Fox, & Boyd (1999) documented unintended effects when a World Wide Web CMC system was introduced in the first phase of a project called the Curtin Learning Link. These effects were classified using Tenner’s (1996) categories of *re-arranging*, *repeat*, *re-complicating*, *re-generating*, and *re-congesting effects*. Under rearranging effects, more time was spent in learning HTML programming skills and developing and maintaining web pages than was saved by creating them. Another example of rearranging effects has to do with “socio-economic factors which reduce the level of access for many learners” (p. 6). Although learners could afford the set-up costs of a computer, their ongoing access was limited by an inability to pay ongoing maintenance costs. “A number of isolated learners were asked to monitor these line costs which ranged from up to \$2.32 for five minutes plus 44 cents per minute during business hours to \$.95 for 5 minutes plus 17 cents per minute after hours” (p. 6). Repeat effects were also noted, e.g. while the unit sites facilitated increased access to online information, “learners spent the ‘saved’ time in surfing for more information (often of doubtful or marginal use)” (p. 6).”

An example of repeat effects is the proliferation of email messages constructed so quickly that further 'remedial' messages are needed (Herrmann et al., 1999). An example of a re-complicating effect is the proliferation of user IDs, passwords and PINs, and the time-consuming search for new ways to reduce information. Regenerating effects include transferring telecommunications costs to learners, and the loss of control resulting from sophisticated hardware and software requirements. Finally, an example of a re-congesting effect is when increasing numbers of Internet users create Internet gridlock, correspondence workload and reduced access. Jones and Petre (1994) also uncovered examples of revenge effects, including file management problems, problems running applications and snags with printing. Their learners found the tutorials frustrating because "the business of following instructions left them too busy to assimilate the material" (p.32). These learners read the help manual only as a last resort.

According to Hannafin, Hannafin, Land and Oliver (1997), the mismatch between the rhetoric of what should happen, and the design practices of what really happens in technology-based environments is most serious in emergent constructivist environments. Klinger (2002), for example, found unintended consequences in a study of an online discussion group. Instead of responding to a BC Ministry of Education policy document, teachers used an online forum to deconstruct the underlying value implications of the ministry document. Because this was not the ministry's intention, Klinger's (2002) study is in effect a study of unintended consequences in an online discussion group.

Unintended social consequences

Unintended social consequences can occur when technology is adopted in face-to-face classrooms. Wilson, Qayyum & Boshier (1998) provide compelling evidence for the

domination of the World Wide Web by American sites and search engines. Another example is the unconscious assimilation of American cultural values by learners (Fabos & Young, 1999). The larger corporate motives behind the adoption of technology in the schools is an important underlying issue (Fabos & Young, 1999). Given the profits to be made from the sales of computers to schools, it is hardly surprising that little attention may be paid to pedagogy (Fabos & Young, 1999). Stoll (2000) claims that technology is used mostly for word processing and games, and provides several examples where the benefits associated with computers in schools are over-rated. One example of adverse social consequences is the misdirection of public funds into flashy multimedia of untested pedagogical quality often at considerable expense (Lookatch, 1997).

Unintended social consequences also occur in distance education courses, one of the more well-known examples being the isolation of distance learners. Gooler's (1979) list of "important social consequences" includes forced alteration of university policies, continuous registration, faculty acceptance, internal rewards for faculty, newness to learners, and the impact on health care and poverty-oriented programs. Clark (1994a) recommends a questionnaire to participants as an "early warning system" (p. 69) to identify negative effects such as isolation and communication problems among participants.

Positive unintended consequences

According to Bates (2000), instructional technologies also result in unintended positive consequences. Sometimes side-effects can provide important benefits which are not captured by research designs which primarily consider the ability of technology to replicate classroom teaching. At Stanford, a distance education course resulted in the spontaneous emergence of networks or communities of practice, a positive consequence

which was not anticipated by the course designers (Gibbons, Pannoni, & Orlin, 1996). Engineers watching a videotape of lectures would stop the tape and discuss at regular intervals before continuing with the lecture. To everyone's surprise, even though they had low credentials when entering the course, they consistently outperformed the classroom group when tested on the course material. According to Seely Brown and Duguid (2000), "this finding has proved remarkably robust and the courses using this 'TVI' [tutored video instruction] method have had either comparative success" (p. 222).

Another example from education is the unintended "spill-over" benefits of a program, such as when children who learn reading skills become more cooperative and less disruptive (Weiss, 1998). Unanticipated positive contagious effects include children teaching to others the skills they have learned in an innovative program (Weiss, 1998). Ruhe (1998) found positive unintended consequences of using email to teach ESL to foreign college students, including affective benefits and increased knowledge of cultural differences. However, Weiss (1998) notes that positive unintended consequences are less likely than negative unintended consequences in innovative programs because it is likely that program reformers will have listed and exhausted all likely results.

Evaluation Models for Distance/Distributed Instructional Programs

Overview

As we have seen, the literature of program evaluation often mentions unintended consequences as an important evaluation criterion. In effect, the term "quality control" refers to the control of unintended consequences. However, it is difficult to find the term "unintended consequences" in evaluation frameworks in the field of distance education over the last 20 years. In fact, the term itself is avoided as a primary or secondary criterion

in evaluation frameworks for distance education. Instead, these models may refer to “goodness of fit” in the course components, “incongruities in the system”, fit among the elements (which implies lack of fit), “negative effects” and “the gap between the ideal and the real”. In effect, these terms are euphemisms for “unintended consequences”. This conclusion flowed from the following analysis.

First , I made a list of evaluation frameworks in distance education. My criterion for selection was that the model be recommended for “generic” application across technologies and subject areas, instead of being recommended for specific applications such as video (e.g. Lane, 1989). I identified eight evaluation frameworks which met this criterion. Next, I listed the concepts which were present as explicit evaluation criteria in these models. I used a “Yes” to indicate that these concepts were present in the framework or “No” to indicate which concepts were present in which models. I then wrote up my results in a “concept by framework” matrix (Table 1).

Table 1
Overview of Distance Education Evaluation Models

<u>Author</u>	<u>Year</u>	<u>Outcomes</u>	<u>Relevance</u>	<u>Cost</u>	<u>UC</u>	<u>Values</u>	<u>Interaction</u>	<u>P.E.</u>
Gooler	1979	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rumble	1981	Yes	Yes	Yes	Yes	No	Yes	No
Collis	1993	Yes	No	No	Yes	Yes	Yes	Yes
Clark	1994	Yes	Yes	Yes	Yes	Yes	Yes	No
Bates	1995	Yes	No	Yes	No	No	No	No
Van Slyke, Kittner & Belanger	1998	Yes	Yes	No	No	No	Yes	No
Belanger & Jordan	2000	Yes	Yes	Yes	No	No	Yes	No
Scanlon et al.	2000	Yes	No	No	No	No	No	No

In Table 1, the *Author* and *Year* indicate the source. *Outcomes* refers to types of outcomes for the learners such as learning outcomes and learner satisfaction ratings. *Relevance* refers to the link between the curriculum and the contemporary needs of society, and to the authenticity, and transferability of the learning materials. *Cost* refers to cost-benefit, cost-efficiency or cost-effectiveness. Under *Unintended consequences (UC)*, I coded a "Yes" if the model contained either 1) a euphemism for unintended consequences or implied their presence (e.g. "Fit" implies "lack of fit") or 2) examples of specific unintended consequences sprinkled in the background or mentioned in passing. I coded a

“No” if there was no reference of any kind to the unintended effects of distance courses.

Values refers to whether the model includes an investigation of values such as the underlying theory, ideology and values and/or the values of stakeholders. A recommendation to describe course objectives is not a criteria for inclusion in this table.

Interaction refers to whether the model is dynamic, that is, whether there is a recommendation to analyse the overlap or interaction among the components of the implementation system. *Program Evaluation* refers to whether the model is based on references to the literature of program evaluation. From here, I will now discuss the evaluation models.

Gooler (1979)

Gooler's (1979) evaluation framework includes purpose, audience, issues, resources, evidence, data gathering, analysis and reporting. Nested within this framework, his evaluation framework includes the following elements: 1) access and equality of opportunity, 2) relevancy to needs and expectations, with the recognition that shifts in needs occur, 3) quality of academic program offerings, 4) learner outcomes, including unanticipated program outcomes, 5) cost-effectiveness, 6) impact on individuals, the institution and society, and 7) generation of knowledge. Although Gooler avoids using the term “unintended consequences”, he gives many varied and specific examples of unintended consequences as they relate to the above elements. Under “access” for example, he mentions delivery problems. Under “outcomes”, he mentions changes in learner attitudes and sophistication. Gooler's (1979) extensive list of “important social consequences” has previously been mentioned.

Rumble (1981)

Rumble (1981) recommends that evaluation take place on two levels. The first level is a comparison of objectives or intended outcomes with actual outcomes, in other words, “the overall performance of the system under evaluation, in terms of output relative to its aims and objectives” (p. 67). The objectives are the context or the “ideal” which “has been defined and which can be used as a benchmark against which actual performance can be compared” (p. 66). The second level is a delineation of the various sub-systems, that is, an analysis of the coordination of the sub-components in the day-to-day operations.

Evaluation takes place at two distinct levels. At the first level, the concern is the overall performance of the system under evaluation, in terms of output relative to its aims and objectives. At the second it is with the internal functioning of the system—the assessment of the effectiveness and efficiency of the various sub-systems (p. 67). In effect, this two-level approach, which is central to Rumble’s (1981) model, is a system-wide analysis of the gap between the standards of the ideal and the actual program implementation.

For Rumble, “outcomes” includes the following: 1) number of graduates in the shortest possible time, 2) ratio of the number of graduates to total number of learners admitted, 3) response to needs of learners and society, 4) cost efficiency, and 5) effectiveness. Although Rumble (1981) does not use the term “unintended consequences”, this two-level approach is an analysis of both intended and unintended outcomes as they play themselves out in the day-to-day operations of the distance education system. In effect, Rumble is calling for an in-depth investigation of the implementation system, including unintended consequences.

Collis (1993)

Collis's (1993) five-stage evaluation framework is based on Stake's (1967) countenance model. Stage 1 is an analysis of assumptions, intentions and planning of the project. Stage 2 is an assessment of logical contingencies, that is, the likelihood of how the assumptions, intentions, plans for execution and success indicators will be interrelated. The third stage involves making observations about the dynamics driving the project, for example, "personal ambition, or the project's desire to perpetuate itself, as much as it may relate to the stated goals" (p. 270). The fourth stage is "an assessment of the goodness-of-fit between what was planned and what is observed to be happening" (p. 270). The fifth stage is "the interpretation of the incongruities in the system which usually involve "a complex of reasons" (p. 270). The final output of the analysis is a set of recommendations for changes in the system.

Although it is a framework for analyzing the implementation system, Collis (1993) avoids using the term "unintended consequences", and instead uses the terms "goodness-of-fit" and "incongruities in the system". In applying her framework to the evaluation data from a course using three technologies to deliver professional training to engineers and managers in electronics industries, Collis (1993) showed in considerable detail the many diverse and specific ways in which unintended consequences played themselves out in the course implementation.

The problem with Collis's (1993) framework is that it is both cumbersome, and incomplete because it excludes cost and relevance as evaluation criteria. Because there are five different versions of the model, one for each stage of the process, the approach is confusing. The fifth framework has no fewer than eight boxes and eleven paths between

them, which would make it difficult to apply. To analyze “unintended consequences”, for example requires tracing through various paths in the five models. Finally, being based on Stake’s (1967) model, Collis’s (1993) framework needs to be updated to reflect the insights in the recent literature of program evaluation.

Clark (1994a)

Clark (1994a) proposed a two-level framework consisting of participant reactions and achievement of program objectives. Surveys can be used to uncover learners’ perceptions of changes in their learning and unanticipated consequences. Because Clark believes that media can no more influence the quality of learning than a delivery truck can influence the quality of nutrition, he recommends that program objectives be divided into “at least two categories: those associated with delivery and those associated with instruction” (p. 69). Delivery technology includes “equipment, machines and media” (p. 64) while instructional technology includes “lessons, examples, practices and tests” (p. 64). For Clark (1994b), the effects of media should be considered separately from instruction. Delivery technologies should be evaluated for their abilities to provide access and technical quality, while instructional technology should be evaluated for changes in learning, transfer, motivation and application of knowledge outside of the classroom (Clark, 1994a).

Clark (1994a) recommends evaluation in the early stages of course implementation to identify “negative effects” which can then be corrected before the program ends. A questionnaire to participants can identify negative effects such as the unmet social needs of high school students. Clark notes that unexpected effects can also be beneficial. He also (1994a) recommends a cost-effectiveness analysis, with time, especially the donated time of volunteers, being included as a cost. This analysis should also include opportunity costs

Acknowledgements

There are no words adequate enough to express my thanks to Professor Bruno Zumbo. With his exceptional kindness and supervisory skills, and unparalleled expertise in all areas of measurement and evaluation, he guided this dissertation to an efficient and speedy conclusion. Dr. Zumbo has been the most wonderful supervisor anyone could ever have. I would also like to express my heartfelt thanks to Professors Karen Meyer and Carl Leggo, whose post-modern vision of education has informed not only my understanding of program evaluation, but more importantly, of everyday life. Next, I would like to thank Dr. Stephen Petrina, who supported me in the beginning, by suggesting I re-analyze the OLT data, and in the end, as university examiner. Finally, I would like to express my deepest thanks to my parents, Lawrence and Phyllis Oszust, for their constant encouragement, home-cooked meals and many other sources of support. Without the kindness and support of all these people, this dissertation would never have been completed.

by comparing the costs of the program being evaluated to the cost of an alternative delivery method for the same program. Finally, Clark recommends an investigation and consideration of the views of stakeholders.

Bates' (1995) ACTION model

Because Bates' (1995) ACTION model is called a "model" by its author, the term "model" will be used in this description. The ACTION model consists of the following evaluation criteria:

- A** Access: How accessible and flexible is the technology?
- C** Costs: What is the cost structure? Unit cost per learner?
- T** Teaching and learning: What learning, instructional approaches and technologies are best?
- I** Interactivity and User-friendliness: What kind of interaction is provided? How easy is it? How reliable is the technology? Are there frequent crashes or break-downs?
- O** Organizational Issues: What are the organizational requirements and barriers?
- N** Novelty: How new is the technology?
- S** Speed: How quickly can the course be changed to accommodate revisions and updates?

In effect, this is a two-level model because access, teaching and learning, interactivity and costs are evaluated at the level of the individual, while novelty and speed are evaluated at the level of the organization.

Van Slyke, Kitter & Belanger (1998)

Van Slyke et al. (1998) proposed a two-level framework of evaluation consisting of predictor variables and outcome variables. The predictors are: 1) learner characteristics, 2)

Acknowledgements

There are no words adequate enough to express my thanks to Professor Bruno Zumbo. With his exceptional kindness and supervisory skills, and unparalleled expertise in all areas of measurement and evaluation, he guided this dissertation to an efficient and speedy conclusion. Dr. Zumbo has been the most wonderful supervisor anyone could ever have. I would also like to express my heartfelt thanks to Professors Karen Meyer and Carl Leggo, whose post-modern vision of education has informed not only my understanding of program evaluation, but more importantly, of everyday life. Next, I would like to thank Dr. Stephen Petrina, who supported me in the beginning, by suggesting I re-analyze the OLT data, and in the end, as university examiner. Finally, I would like to express my deepest thanks to my parents, Lawrence and Phyllis Oszust, for their constant encouragement, home-cooked meals and many other sources of support. Without the kindness and support of all these people, this dissertation would never have been completed.

course characteristics 3) distance learning characteristics and 4) institutional characteristics (including objectives, delivery methods and support structure). As for outcome variables, there are two levels: 1) institutions and 2) learners. Institutional outcomes include lower costs, better productivity of instructors, shared resources with other institutions and increased geographical reach. Learner outcomes include technical awareness and skills. These authors believe that these variables interact in a complex system, but they do not hypothesize any cause/effect relationship among the variables.

Belanger and Jordan (2000)

Belanger and Jordan (2000) proposed a framework consisting of similar predictor variables to those of Van Slyke et al.'s (1998) model. However, they include not two, but four levels of outcome variables impacted by distance learning: learners, instructors, institutions and society (Figure 1).

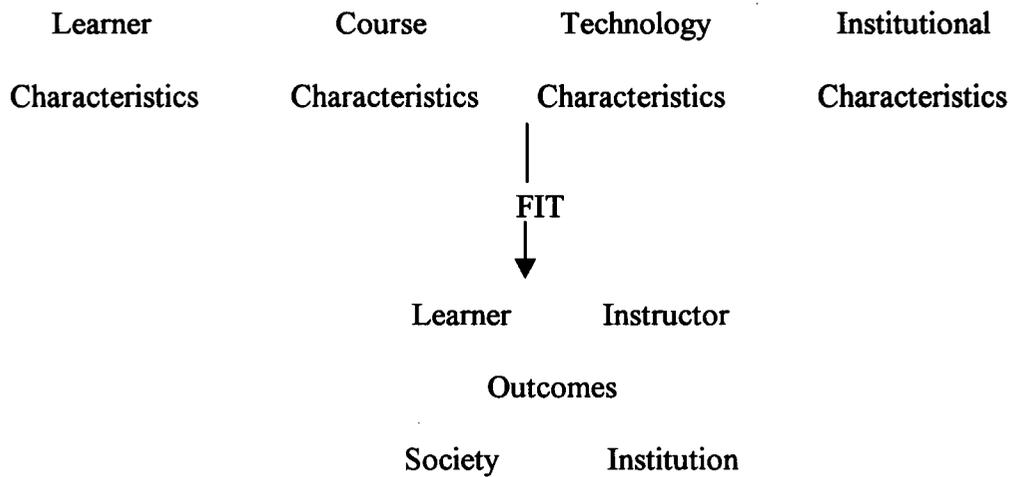


Figure 1. Belanger and Jordan's framework of evaluation

Learner characteristics includes the learners' objectives, personal skills such as self-sufficiency, computer proficiency, time management, interpersonal communication, problem-solving and planning, previous technology experience, and expectations and attitudes. *Course characteristics* include group projects, evaluation methods and hands-on components of the course, e.g. a series of computer-mediated technologies to support collaboration tasks. *Technology characteristics* include the "transition to an "anytime, anywhere" environment [which] provides no inherent guarantees for quiet, comfort or ease of learning" (p. 189).

The lower part of the framework shows four levels of outcomes: *Learner, Instructor, Society and Institution*. *Learner outcomes* include increased technology awareness and skills, and higher quality of interaction with or better access to the instructor. Outcomes for one learner can be different than for another learner, depending on how the characteristics on the upper level of the framework interact. *Institutional outcomes* include lower costs, increased geographical reach, increased productivity among instructors and the sharing of instructional resources with other institutions. Finally, *Social outcomes* include a more

professional workforce, increased quality of life, and increased access to education (regardless of culture, class or financial status of the learner).

One major contribution of Belanger and Jordan's (2000) framework is the concept of fit. The course variables are not modelled as isolated elements, but as complex, dynamic systems with multiple components interacting as a system. An arrow labelled "Fit" shows that "all of these course characteristics and contextual variables must be carefully examined, not in isolation, but together" (p. 189). One example is when the capabilities of younger learners to succeed in distance education programs are reflected in adjustments in admissions and recruiting policies and decisions in a way which enhances the overall efficiency and success of the instructional system (Johnstone & Krauth, 1996).

The CIAO Model

Based on twenty years of doing evaluation in course teams, the Open University produced the CIAO model for evaluation (Scanlon et al., 2000). (As previously, the term "model" will be used, because this is the term used by its authors.) The CIAO model considers context, interactions and outcomes (Figure 2).

	Context	Interactions	Outcomes
Rationale	Aims and Context of use of CAL	Process data to understand whether, how and why some element works	Cognitive and affective learning outcomes; attribution of outcomes to CAL is difficult
Data	Designers and course team aims Policy documents and meeting records	Records of student interaction, diaries, online logs	Measures of learning Changes in learners' attitudes and perceptions
Methods	Interview designers and course team Analyse policy documents	Observation Diaries Video/audio and computer recording	Interviews Questionnaires Tests

Figure 2. The CIAO model of evaluation

The CIAO model recommends an analysis of course team objectives, or intended consequences, by analyzing policy documents and meeting records of the course team. The model includes learning outcomes and acknowledges the difficulty of attributing learning outcomes to CAL. Several methods for data collection are also included.

Weaknesses of Distance Education Evaluation Frameworks

In comparing the preceding conceptual frameworks, there are four recurring elements: 1) outcomes, including learner satisfaction, 2) relevance to the needs of society, 3) costs, 4) unintended consequences, 5) values and 6) interaction. Other elements found in only one framework include novelty, speed of updating, increase in funding possibilities (Bates, 1995) and increase in the generation of knowledge (Gooler, 1979).

One weakness in all of these models is that none of these authors use the term “unintended consequences” as a primary evaluation criteria. In fact, all of them avoid even using the term “unintended consequences”, which hardly appears in the evaluation

literature of distance education. However, in almost all of the models, the concept is *implied* or *expressed* in euphemisms such as the “goodness-of-fit” (Collis, 1993), the gap between the ideal and the real (Rumble, 1981), the “fit” among the components (Belanger and Jordan, 2000), and “why and how some element works in addition to whether it works or not” in the CIAO framework (Scanlon et al., 2000, p. 4). Moreover, as demonstrated in this overview, the literature is replete with a very broad range of specific exemplars of unintended effects associated with every aspect of the implementation system. Yet the use of the term “unintended consequences” as an explicit evaluation criterion has been virtually “taboo” in the distance education literature for the last twenty years.

Secondly, most models do not recommend an analysis of the underlying values, goals or ideologies or the ways in which these values are reflected in the course implementation. Collis’s (1993) framework refers to underlying values and the roles of stakeholders, but does not go into much detail. Clark’s (1994) framework also recommends an investigation of stakeholder values, but does not explain how this should be done. Gooler (1979) notes that multiple stakeholders have different value systems, which underscore “the need to consider pluralistic purposes of distance education programs” (p. 47).

Only Gooler’s (1979) and Collis’s (1993) frameworks have any basis in program evaluation. Collis’s (1993) framework resembles Gooler’s (1979), Rumble’s (1981) and Belanger and Jordan’s (2000) frameworks in the sense that all three analyze the workings of the implementation system. Rumble’s (1981) framework resembles a program evaluation model, but does not include values. Collis’s (1993) framework is incomplete, cumbersome and based on Stake’s (1967) countenance approach. According to Gooler (1979), “it makes sense to apply the notions of evaluation to distributed educational

program” (p. 43), but added that distance instructional programs have unique characteristics which require different criteria and designs.

Summary

Over the last 20 years, approaches to evaluation in distance education have developed largely independently from the literature of program evaluation (e.g. Caulder, 1994b, Knapper, 1980; Thorpe, 1988). Being largely confined to questions- and method-driven approaches, these evaluation models lack the depth and diversity of contemporary evaluation models in program evaluation. In the literature of evaluation models for distance courses, there have been calls for investigating the unintended consequences of distance instructional programs. Despite the many and varied exemplars of unintended consequences in the literature, the use of the term “unintended consequences” is avoided. Yet program evaluation models often include explicit reference to unintended consequences and quality assurance and control is concerned with strategies to minimize them. Tenner (1996) conceptualized four categories of the unintended consequences of technology. In the literature of educational technology, and to a lesser extent distance education, there are many examples of unintended consequences which emerge when technology is implemented.

Given the “new order of complexity” of distance educational programs based on multiple technologies (Rumble, 1981, p. 65), the time has come to include unintended consequences as a category in distance education evaluation models. However, in the literature of evaluation of distance education, there is no clear sense of what “unintended consequences” means. Does the term “unintended consequences” refer to improper use, “misuse” or trivial misapplications or does it refer to something else? If unintended

consequences can be positive or negative, then they must be somehow linked with values. Yet there is no mention of linking unintended consequences with the values underlying distance courses. The absence of a discussion of the meaning of the term “unintended consequences” is an important gap in the distance education literature which can be filled by bringing in insights from the fields of assessment and program evaluation.

CHAPTER 3: A Comprehensive Framework for Evaluation

Overview

In Chapter 2, I presented the theoretical context, which begins with an overview of program evaluation and the role of values. Next, I discussed the history of evaluation in distance education, distance education evaluation models (with a summary of common elements and shortcomings), quality assurance models and the unintended consequences of educational technology. In Chapter 3, I will describe the contribution of my research, define validity and validation and discuss Messick's (1989) framework of validity, including its four facets, philosophical underpinnings, and the controversy over unintended consequences. Next, I will present a model of evaluation for distance courses adapted from Messick's (1989) framework and discuss its four facets, which will be "fleshed out" with concepts and examples from distributed learning. Finally, I will give examples from the literature of educational technology to illustrate the kinds of issues which can arise from applying the adapted model to a new context—the evaluation of distance instructional programs.

Contribution of This Research

The contribution of my research is to present insights which emerge from using an adapted version of Messick's (1989) framework as a model to evaluate distance/distributed instructional programs, and from applying this model to the data from three post-secondary courses in BC. As I have shown in Chapter 2, traditional evaluation models for distributed instructional programs are incomplete because they do not include unintended consequences as an explicit evaluation criterion. In contrast, the adapted Messick's (1989) framework provides a comprehensive analysis of merit and worth because it includes

traditional evaluation categories, and also adds two new categories not found in traditional models, that is, unintended consequences and value implications. Throughout this research, I will refer to my model as “the adapted Messick’s (1989) model” because my model is almost identical to Messick’s model, except for some re-labelling, and the underlying assumptions are also the same.

My research responds to calls in the literature of evaluation in distance education to investigate unintended consequences (Collis, 1993; Gooler, 1979; Rumble, 1981), even though the term “unintended consequences” is avoided. To fill in this gap, I will introduce the adapted Messick’s (1989) four-faceted conception of validity, which is applied in the context of evaluating the worth of standardized tests. Why have I chosen to adapt Messick’s (1989) model? The answer is that Messick (1989) has already done a very in-depth and fine-grained exploration of unintended consequences, and has many insights which emerge when the model is used to evaluate distance courses.

The Overlap between Validity and Evaluation

To begin this section, I would like to assert the credibility, feasibility and relevance of taking a conceptual framework from measurement and applying it to a new context—the evaluation of distance programs. While some readers might be taken aback by this application of Messick’s (1989) framework, I would like to remind them that I am not asserting that psychometrics is the same as program evaluation. Instead, I am asserting that there are several benefits associated with using concepts from the adapted Messick’s (1989) framework to guide evaluation studies in distance education.

Academic disciplines are not isolated fortresses. The concepts and principles associated with one discipline are routinely borrowed by others, and there is considerable cross-fertilization across subject areas. Ross and Morrison (1997), for example, present a long history of paradigms in educational testing and measurement being applied to evaluation in instructional design. The authors state that the fields of measurement and testing have provided some of the conceptual foundations of instructional design. As paradigms in measurement shift, so do the foci of assessment and evaluation in instructional design. They conclude that "if historical trends continue, we should expect measurement and evaluation emphases to mirror the prevailing paradigms in educational (instructional) theory and research.... Domain-specific interests will dictate more focused methodologies, particularly where the evaluator is a content or domain expert rather than a measurement-evaluation specialist" (p. 337).

In adapting Messick's (1989) framework into an evaluation model for distance education courses, am I implying that test validity and program evaluation are the same thing? Not exactly. Fifty years ago, when the fields of program evaluation and assessment were based on experimental methodologies, there was considerable overlap between them. However, with the adoption of qualitative methodologies and a proliferation of new approaches to program evaluation, assessment and program evaluation later emerged as distinct fields. Yet although they may often be thought of as distinct, these two fields share a common conceptual core, which is determining the worth, merit or value of educational activities. It is true that Messick's (1989) framework is a framework for assessing the value of worth of standardized tests, not educational programs. But just as Mabry (2001) points out the common core between personalized assessment and evaluation, Messick's

(1989) framework of test validity and program evaluation models share a common conceptual core—the appraisal of the value or worth of educational activities.

Common definitions of the term “program” are general enough for tests and courses to qualify as “programs”. If we use Wholey’s (1987) definition of the term “program” as a set of resources and activities directed toward one or more common goals, then a standardized test is a kind of program, because a test is a set of resources and activities directed toward a common goal. Consequently, it is not inconceivable that a model for determining the merit or worth of one kind of program, that is, a test, could also be useful in determining the merit or worth of another kind of program, that is distance instructional programs.

In hindsight, it is not surprising to discover that Messick treats a test *as if* it were a program. Most of the categories of his model overlap with categories commonly used for evaluating programs, e.g. cost-benefit, relevance, values and unintended consequences. Moreover, his approach to validating tests was an abrupt departure with traditional “fragmented” methods. Messick’s contribution to test validation is a comprehensive argument-based approach, using multiple methods, which is an approach commonly used in program evaluation studies. Nor was Messick alone in linking test validity and program evaluation. Some eminent scholars work in both program evaluation and test validity (Cronbach, 1982; 1989; 1990; Cronbach & Meehl, 1955). The fact that such theorists worked in both areas is evidence for the overlap among the two fields.

In summary, the contribution of my dissertation is to answer the recurring call in the distance education literature to investigate unintended consequences. By applying the adapted Messick’s (1989) framework to the evaluation of distance courses, I will share

Messick's insights into the meaning of the term on "unintended consequences", which is absent from the literature of distance education. My research will provide empirical evidence of the kinds of important insights which can be obtained from applying the adapted Messick's (1989) framework to the evaluation of distance courses, and from using unintended consequences" as a bridge between the fields of distance education and program evaluation.

Validity and Validation

Validity is a property of a good test, questionnaire or observation. Validity refers to trustworthiness or accuracy, and *validation* refers to the process of analyzing data to assess the validity of a measure. Validity is often assessed along with reliability, which refers to consistency and stability. An instrument can be reliable, but not valid, but to be valid, it must first be reliable. Reliability is a necessary, but not a sufficient condition for validity (Hubley & Zumbo, 1996).

Our theoretical conception of validity and validation practices have changed appreciably over the last sixty years (Angoff, 1988). According to Curreton (1951), the essential feature of validity was "how well a test does the job it was employed to do" (p. 621). In the American Psychological Association's (1954) Technical recommendations for psychological tests and diagnostic techniques, there were four distinct types of validity: *construct validity*, *content validity*, *criterion validity* and *concurrent validity*. *Construct validity* refers to how well a particular test can be shown to assess the construct that it is said to measure. *Content validity* refers to how well test scores adequately represent the content domain that these scores are said to measure. *Predictive validity* is the degree to which the predictions made by a test are confirmed by the later behaviour of the tested

individuals. *Concurrent validity* is defined as the extent to which individuals' scores on a new test correspond to their scores on an established test of the same construct that is administered shortly before or after the new test.

In the American Psychological Association's (1966) Standards for educational and psychological tests and manuals, criterion validity and predictive validity were collapsed into criterion-related validity, thereby reducing the four validity types into three types: criterion-related, content validity and construct validity. These three *aspects* of validity were referred to as the Holy Trinity (Guion, 1980), "meaning that at least one type of validity is needed but one has three chances to get it" (Hublely & Zumbo, 1996, p. 210). As early as 1957, however, Loevinger (1957) argued that construct validity was the whole of validity, anticipating a shift away from multiple types to a single type of validity.

Moreover, in the early days, validity was viewed as a property of tests, but the focus later shifted to the validity of a test in a specific context or application (Angoff, 1988). The authors of the 1974 Standards for Educational and Psychological Tests (APA, 1974) shifted the focus of content validity from a representative sample of content knowledge to a representative sample of behaviours in a domain (Messick, 1989). Moreover, standards for test use, including bias, adverse impact and the social consequences of the use of tests were also included (Messick, 1989). A major shift occurred when the focus for validation shifted from validating the test to validating responses to validating inferences and actions based on test scores (Angoff, 1988). In the 1985 *Standards* (APA, 1985), validity was redefined as the "appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (p. 9). In this version, validity is a property of "test-based inferences", i.e., judgements made on the basis of test score evidence. Validity is not a

characteristic of the instrument. Instead, validity was based on the notion of construct validity, i.e. "the extent to which a particular test can be used to assess the construct which it claims to measure" (p. 249). Professional standards were established for a number of applied testing areas such as "counseling, licensure, certification and program evaluation" (Messick, 1989, p. 18).

Validation practice is "disciplined inquiry to disprove alternative inferences from scores or other observations (Hubley and Zumbo, 1996). Traditionally, the validation process has consisted of calculations of measures of content representativeness, factor analysis and correlations with other measures (American Psychological Association, 1966). Each of these procedures, which is performed on the test responses, yields a score which is taken to be a measure of a single aspect of validity, e.g. content validity or predictive validity. Many of these procedures are based on logical or mathematical models that date from the early 20th century (Crocker & Algina, 1986). Messick (1989) describes such procedures as fragmented, unitary approaches to validation. Hubley and Zumbo (1996) describe them as "scanty, disconnected bits of evidence...to make a two-point decision about the validity of a test" (p. 214). In contrast, Cronbach (1982) recommends an argument-based approach to validation, which considered all sources of evidence bearing on validity.

In sum, the emphasis in successive conceptions of validity shifted from many types to one type of validity, from prediction to explanation and from validity being a property of tests to validity being a property of inferences (Messick, 1989). In addition, our conceptions of validation processes have evolved from a fragmented approach to a comprehensive unified approach in which multiple sources of data are used to establish the

case for validity. As for actual validation practice, however, checklists of procedures continue to be used to establish a “holy trinity” of validity, when what is needed is a combination of methods to “bridge the gap between psychometric theory and research practice” (Hubley and Zumbo, 1996, p. 215).

Messick’s (1989) Unified Conception of Validity

In 1989, Messick’s (1989) conception of validity brought many of the concepts previously mentioned together into a unitary conception of validity. Validity is “an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (Messick, 1989, p. 13). The basic question for validation is whether test scores should be interpreted and used in the manner proposed. Validation practice consists of building an argument based on multiple sources of evidence, including statistical calculations, qualitative data, a weighing of one’s own values against different value perspectives and an investigation of unintended consequences. Validity is an evaluative summary of the evidence for both the potential and actual consequences of score meaning and use (Messick, 1995b). “It is the argument that justifies the inferences as valid, rather than the inherent quality of the test” (Markus, 1998, p. 80). For this reason, “it is important to note that validity is a matter of degree, not all or none” (Messick, 1989, p. 13).

Validity is a unified concept and validation is a scientific activity (Messick, 1989). The emphasis is on scores instead of tests because “inferences are drawn from *scores*” (Messick, 1989, p. 14). *Scores* are broadly defined as “any coding or summarization of observed consistencies on a test, questionnaire, observation procedure, or other

performance on assessment device” (Messick, 1989, p. 14). Moreover, because test scores are a product of the persons’ responses and the implementation environment, validity needs to be assessed for the specific context in which test scores were obtained. According to Reckase (1998b), “each different definition of an application defines a particular line in multidimensional space, and a different validity for a measurement instrument” (p. 52).

For Messick (1989), validity is a unified concept consisting of four facets which are formed by crossing the evidential with the consequential bases of test interpretation and use (Figure 3).

		<u>Outcomes</u>	
		Test Interpretation	Test Use
<u>Justification</u>			
Evidential Basis		Construct Validity	CV + Relevance/Utility
Consequential Basis		Value Implications	Social Consequences

Figure 3. Messick’s (1989) unified conception of validity

The *outcomes* dimension includes the categories of *test interpretation* and *test use* (the ways in which the test is actually used). The *justification* dimension includes the *evidential* and *consequential bases*. The *evidential basis* is an appraisal of psychometric data, while the consequential basis is an appraisal of value implications and social consequences. In contrast, the *consequential basis of validity* is the value implications of the inferences based on scores as a basis for action and the actual and potential effects of test use, especially issues of bias, adverse impact, and distributive justice (Messick, 1989).

In the consequential basis, both value implications and unintended consequences are included because values are intertwined with the significance (either positive or negative) which we assign to unintended consequences. *Value implications* is included because the meaning or interpretation of scores depends on values. Finally, *social consequences* is an appraisal of the unintended social consequences of testing, in other words, an analysis of the indirect effects—both actual/potential and positive/negative—of using the test on the overall educational system (Messick, 1995). The tension between the evidential and the consequential basis is the tension between facts and values, which underlies all scientific inquiry.

The evidential basis and construct validity

The *evidential basis for test interpretation* refers to an appraisal of the evidence for construct validity. A *construct* is “a definition of skills and knowledge included in the domain to be measured by a tool such as a test” (Reckase, 1998b, p. 45). Construct validation begins with an appraisal of what to name a construct, how to represent it and what relations to include in a nomological net (Shepard, 1997). A demonstration of content validity is a necessary, but not a sufficient condition for construct validity (Sireci, 1998). Multiple types of evidence including content-related evidence, predictive evidence and/or concurrent evidence need to be gathered, with the relation between the evidence and the inferences determining the validation focus. Convergent and discriminant evidence can be used in a multi-trait multi-method matrix to establish nomological validity (Messick, 1989).

Messick’s (1989) framework is a “progressive” matrix, which means that construct validity overlaps with all four of the other cells, which also overlap with each other. “The fuzziness—or rather messiness—of these distinctions derives from the fact that we are trying

to cut through a unitary concept. (p. 21). The evidence for construct validity needs to be supported by evidence of relevance and utility (cost/benefit) in applied settings such as the workplace. The *evidential basis for test use* includes an appraisal of measures of criterion validity (e.g. correlations with external measures in particular settings) and utility, that is, a cost-benefit analysis. An appraisal of the evidence for relevance and utility buttresses the evidence for construct validity, and is another aspect of the evidence for construct validity (Messick, 1989).

Construct validity is the “unifying force” in a progressive matrix, with highly intertwined and overlapping facets. It is construct validity which makes validity a unitary concept and which ties together the elements of the four cells (Messick, 1989). Because the meaning of scores depends on values, an appraisal of psychometric evidence for construct validity is intertwined with an appraisal of value implications. The evidential basis is intertwined with the consequential basis because facts and values are intertwined. The unitary emphasis is on combining multiple lines of evidence to support the interpretation and use of scores (Markus, 1998). Validity, then, is construct validity, which is multifaceted and value-laden. It follows that validation practice is the construction of a validity argument based on multiple sources of evidence to support the adequacy, appropriateness and meaningfulness of inferences and actions based on test scores (Markus, 1998).

The consequential basis: value implications

An appraisal of *value implications* requires an investigation of three components: 1) the values of the construct labels, 2) the theories underlying their meaning, and 3) the “still broader ideologies that give the theories their perspective and purpose” (Messick, 1989, p.

62). In validation practice, labels given to test scores should be evaluated to determine whether they are accurate descriptions of knowledge and skills said to be assessed by a test, and whether they are value loaded. For example, to say that a mathematics test is “world class” because it was reviewed by a few international experts is misleading and violates this aspect of validity (Reckase, 1998a).

The second component of the category of *value implications* is an appraisal of the theory underlying the test. *Theory* refers to the underlying assumptions or logic of how a program is supposed to work (Chen, 1990). A theory connotes a body of knowledge that organizes, categorizes, describes, predicts, explains and otherwise aids in understanding phenomenon and organizing and directing thoughts, observations and actions (Sidani & Sechrest, 1999). For over thirty years, evaluators have recommended making explicit the underlying assumptions of how programs are supposed to work, and using this theory to guide evaluations (Rogers, Petrosino, Huebner, & Hacsí, 2000).

The third component is an appraisal of the “broader ideologies that give theories their perspective and purpose” (Messick, 1989, p. 62). An *ideology* is “a complex configuration of shared values, affects, and beliefs that provides, among other things, an existential framework for interpreting the world” (Messick, 1989, p. 62). One example is the view that persons with achievement scores below a certain level on a test of moral knowledge and skills are incapable of making moral judgements (Reckase, 1998a). Value implications are a “socially relevant part [of score meaning] that often triggers score-based actions and serves to link the construct measured to questions of social policy” (Messick, 1989, p. 63).

Finally, the consequential basis of test use refers to the unintended social consequences of the use of the test. According to Messick (1989), the functional worth of the test should

take into account all intended and unintended consequences of the test application, including individual, institutional, systemic and societal effects. Messick (1998) says that the side effects of test misuse are not within the scope of the appraisal of validity; instead, the appraisal of validity is confined to the unintended consequences or side effects of *legitimate* test use. Some examples of the unintended side effects of legitimate test use are narrowing the curriculum to teach to the test, placement decisions, prerequisite or minimum consequence decisions, coachability, gender and ethnic differences in score distributions (Shepard, 1997).

The Controversy over Values

From the positivist perspective, the scientific method is held to be value-free and experimental research is designed to lead to controlled and predictable outcomes. Traditionally, the assessment of validity has been based on statistical calculations which were held to be value-free (Crocker & Algina, 1986). In Messick's (1989) framework on validity, however, these kinds of calculations are a part but not the whole of validity. The assumption of the value neutrality of science is "perverse" because the underlying principles, such as predictive accuracy, internal coherence and parsimony, are value judgements (Messick, 1989). Moreover, psychometric evidence can never be value-free but is embedded in social practice, specifically, "in the meanings and values implicit in the social practices which give rise to it" (Markus, 1998, p. 12). Constructivism is neither peripheral nor external to test validation, but "lies at the very heart of validity and as a result, so too do problems of meaning and values" (Markus, 1998, p. 11). The failure to recognize this fact results in validity studies which "are imprecise because types of validity typically remain implicit and undefined" (MacPhail, 1998, p. 137). For these reasons,

Messick's (1989) conception of validity has an evidential basis (facts), which is intertwined with a consequential basis (values).

Messick's (1989) conception of validity has generated intense controversy in the literature of educational measurement. To clarify his thinking, I will now summarize some of this controversy which appeared in a recent issue of Social Indicators Research.

According to Markus (1998), Messick's (1989) conception of validity is based upon an inherent tension between the evidential basis (EB) and consequential basis (CB).

Tension: Validity is value-independent. (EB)

Validity is value-dependent. (CB)

Validity is dependent on values, and values vary widely depending on the specific contexts, applications, stakeholders and researchers. Given that values are diverse, Markus (1998) argues that there must be multiple validities, that is, a matrix of value-dependent validities. However, he sees this notion as problematic because it is at odds with the notion of validity as a unified concept. To resolve this contradiction, Markus (1989) calls for a completion of the synthesis between the evidential basis (facts) and consequential basis (values).

According to Markus (1998), the completion of a synthesis between facts and values is achieved by the development of a value justification which produces a single best justified validity for a given context or application. At the very least, "test users should be prepared to provide justification for the values inherent in their validity arguments or else accept that their validity argument is not uniquely justified" (p. 80). As Moss (1998a) points out, Messick stops short of accepting this view, and states that the tension between EB (facts) and CB (values) needs to be carefully negotiated in validation practice.

In response to Markus (1998), Reckase (1998b) questions whether validity is a unitary concept and whether the validity of a test depends on the values of the researcher. Validation can be conceptualised as comparing lines in multidimensional space; these lines correspond to varying constructs, tests and applications. A researcher's values enter into the construction of a measurement instrument, but a measurement instrument has a single validity for each construct definition and each context or application. Consequently, values may influence the validation process, the quality of data, and the interpretation of results, but not the measurement instrument per se. There may be different validities for different applications and different inferences, validity is not dependent on the values of researchers.

Moss's (1998a) reply

Moss (1998a) posits a dialectical view of rationality in which validity theory is not a completed project, but an "ongoing critical reflection about our interpretations and theories in light of challenges from alternative perspectives" (p. 55). In the four-faceted conception of validity, "value implications are not ancillary, but rather, integral to score meaning" (Messick, 1994, p. 20). Moreover, Messick's (1989) framework is founded on Singer's (1959) view of rationality, which Messick (1989) describes as one system of inquiry being observed by another to open "their underlying scientific and value assumptions to public scrutiny and critique" (p. 61-62). Messick (1989) continues, "The reality of an inquiring system depends on its being 'observed' by another inquiring system. Indeed he suggests that what is fundamental or 'given' in one system is an issue to be deliberated by another inquiring system that is observing the first one in its problem-solving activities" (p. 32).

In the contemporary post-modern world of diverse values, value-neutral data are problematic (Messick, 1998). Instead, there is "a multiplicity of values including the

decision-maker's values, the enhancement of individual welfare, equality and enhancement of the common good" (p. 18). Moss (1998a) recommends that evaluators assess the validity of tests "from multiple value perspectives to address a broad range of potential social consequences and to identify side effects likely to be seen as adverse by other value positions" (p. 80). Multiple value perspectives, then, can deepen and enrich the knowledge gained from the evaluation exercise by opening up conceptual "spaces" which would otherwise remain closed and "problematizing" issues which would otherwise remain hidden. "The issue is not really about what's *possible* within different perspectives (as Bernstein, 1979 notes), it's about what's relegated to the background as unimportant or impractical; and what the impact of these prevailing emphases is on the actual practices of social scientists and the communities they study and serve" (Moss, 1998a, p. 56).

According to Moss (1998a), a completed synthesis is not necessary, and validation practice is open to multiple perspectives which "illuminate taken-for-granted assumptions, values, and practices that alternative perspectives can provoke" (p. 65). According to Moss (1998), a pluralistic approach to values is central to Messick's theory and brings to the foreground knowledge which would otherwise be "disqualified" against the claims of a unitary body of theory which would filter, hierarchies and order them in the name of some true knowledge and some arbitrary idea of what constitutes science and its objects" (Foucault, 1980, p. 82). This emphasis on the importance of an outside perspective to illuminate what is taken for granted (as natural, normal, the 'way things are done') and thereby to provoke critical self-reflection is a theme that resonates across multiple philosophies of social science" (Moss, 1998a, p. 62). Moss (1998a) claims that this

profound insight is one of the most important insights which Messick has brought to the field of educational measurement.

In sum, Messick argues that values are integral to validity, and that in a post-modern world, diverse perspectives must be taken into account (Moss, 1998a). Validation practice needs to reflect a multiplist view of values, which characterize a contemporary, post-modern world. Indeed, this perspective is found in many contemporary evaluation approaches such as responsive evaluation (Abma & Stake, 2001), realist evaluation (Henry & Julnes, 1998) and constructivist evaluation (Caracelli, 2000). The value of approaching a question from these diverse value perspectives is to illuminate and probe the emergent issues and to make these issues explicit, thereby enriching our knowledge (Moss, 1998a). Moreover, evaluation in the context of multiple perspectives ensures that new alternatives, compromises, extensions and re-formulations can emerge and that a broad range of social consequences are addressed (Messick, 1989).

The Controversy over Unintended Consequences

Although Messick's (1989) conception of validity has had an enormous influence on the field of educational measurement, his view that consequences are an integral part of validity has been especially contentious. Messick's (1989) definition of validity as the appropriateness of inferences and actions based on test scores includes an appraisal of unintended social consequences. The reason is that these inferences and actions take place within a social context and have implications and consequences within a broad social context.

According to Moss (1998b), there is a dialectical relationship between testing practices and social realities, and because testing practices transform social realities, the

study of social consequences is essential. Moss (1998b) suggests that evidence of social consequences can be obtained by studying “the actual discourse that surrounds the products and practices of testing” (p. 7). This evidence can provide concrete illustrations of how tests actually work in local contexts, and “about the potential slippage between what we well-meaningly intend and what we in fact effect” (p. 11).

According to Popham (1997), actual and potential consequences are “vitaly important”, but “orthogonal” to validity. Merging consequences with validity only “muddies the waters” and creates confusion. Actual and potential consequences should be identified, but not as a part of a study of validation of score-based inferences. Mehrens (1997) argues that consequences should be moved outside of the discussion of validity because the concept confuses issues in measurement quality with issues in treatment efficacy, which is problematic. The accuracy of an inference about the amount or meaning of any trait should be separable from the treatment. Shepard (1997) agrees that including consequences overburdens the concept of validity and creates confusion. Because it is difficult to establish a cause-effect relationship, consequences cannot easily be separated from confounding variables, but they are still a part of validity and should be examined.

Several authors have examined how consequences can be examined in the validation of large-scale national test scores. Moss (1998b) suggests a study of discourse to assess unintended consequences in the local contexts in which tests are administered. The goal of this research is to develop concrete examples of unintended consequences in specific situations. Cizek (2001) provides a list of 10 “unintended, unrecognised and unarticulated positive consequences of high-stakes testing” (p. 19). This list includes accountability systems, improved student learning and a heightened scrutiny of the content of tests.

Reckase (1998a) suggests that consequential validity could be assessed by 1) an appraisal of the value labels of constructs labels in the test and test manual, 2) an articulation of the ideologies on which the test is based, and 3) an appraisal of actual and potential consequences in schools. Reckase (1998a) then applies this strategy by looking at the ACT Assessment Test Battery for college admissions. With respect to value implications, he examined construct labels in the test manual, such as English usage, and was satisfied with the list knowledge and skills assessed, and the absence of any “overblown” descriptors such as “world class”. He also found that the documentation, theory and ideology supporting the test were reasonable. The theory behind the test was that Grade 12 students who performed well on a test consisting of a sample of items reviewed by college faculty would do better in college. The underlying value implication, that faculty judgements are valued, is also acceptable. The ideology in which the test battery is embedded is that a college education is valued, that students should prepare themselves for it, and that certain fields of study are prerequisites for success. However, Reckase (1998a) raises several questions about how unintended consequences can be evaluated. First, there is the question of when to collect the evidence for a new test? Second, it is difficult to demonstrate that an event is the effect of a test, and not the effect of any number of other contextual variables. Finally, Reckase (1998a) concludes that, given these constraints, an evaluation of unintended consequences may not be possible.

Another source of controversy is whether the test maker or test user is responsible for unintended consequences (Shepard, 1997). If schools offer increased funding to schools with increased test scores, then the learning consequences which follow are an important validity issue. “When are consequences part of the nomological net and when are they the

purview only of policymakers and politicians?” (p. 8) The author recommends that test makers should do at least one study to examine the relationship between the test and the effects and check for regularly occurring side effects such as adverse impact.

Green (1998) argues that test publishers are not in a position to obtain on their own evidence on the consequences and uses to which their tests are put. There is little hard or credible evidence, obtaining cooperation is difficult, and the uses vary widely, thereby making generalization difficult. Publishers are disconnected from the ways in which teachers use test results, but in some sense, both parties are responsible for unintended consequences, and a dialogue between them is recommended.

In conclusion, Messick's (1989) framework on validity has met with more than a decade of intense controversy. According to Messick (1998), such controversy “masks conflicts in values and ideologies” (p. 39). Because validity is a judgement not just about the accuracy of score-based inferences, but also about the appropriateness, meaningfulness and usefulness of score-based inferences, then “intrinsically, judgements of worth need to take into account the consequences of test interpretation and use” (p. 41). Both anticipated and unanticipated consequences, then, form strands in the nomological net by contributing to score meaning and providing evidence for construct validity. One approach to getting past this controversy is to examine more closely how Messick defined the term “unintended consequences”, and to be clear about what he includes in his use of this term, and what he does not include.

Messick's response—Defining the term “unintended consequences”

To bring some clarity to this controversy around Messick's (1989) framework, a clear definition of the term “unintended consequences” is needed. Messick (1998) claims that his critics are “side-tracked by a misplaced concern over test use” (p. 39), and mistakenly believe that he was using the term to refer to the consequences of the misuse or trivial misapplications of tests. “Misuse” refers to using the test in ways in which it was not meant to be used, including procedural errors and unsound interpretations. Messick (1998) states quite clearly that this was not his intention, but that he is concerned with “the unanticipated side-effects of *legitimate* test use” (p. 40). In other words, the concern is with the unanticipated effects of using the test in the way it was intended to be used. Because “validity” refers not only to the accuracy of score-based inferences, but also to the appropriateness, meaningfulness and usefulness of score inferences, a judgement of worth must take into account the unintended consequences, which are intertwined with value implications. The consequences of test misuse, then, are orthogonal to score meaning as construct validity, and are not part of the definition of unintended consequences.

Unintended consequences signal invalidity only if they can be traced to sources of invalidity such as construct under-representation or construct-irrelevant variance (Messick, 1989). If unanticipated side effects arise from legitimate test use, they can be ignored if they are trivial (Messick, 1998). But if they are not trivial, they cannot be ignored; instead, both score meaning and intended uses need to be modified. Unanticipated consequences signal that “we may have been incomplete or off-target in test development and hence in test interpretation and use” (p. 43). However, unintended consequences can sometimes be

positive; this is referred to as “positive washback” or “beneficial by-products” (Messick, 1996).

Summary

Messick (1989) conceptualizes validity as a unified, four-faceted conception of functional worth or value. The validation argument includes an appraisal of the relevance and utility of scores, the value implications of scores as a basis for action, and the unintended consequences of their use. Messick’s (1989) framework of validity rests on a tension between the evidence (hard psychometric evidence) and consequences (values). The term “unintended consequences” is defined as the unanticipated effects of using the test in the ways in which it was intended to be used. Moreover, the four-faceted conception of validity rests on a post-modern approach to values, according to which there could be not just one validity, but many different value-dependent validities. Because this pluralistic approach to values underlies Messick’s (1989) conception of value, the inherent tension between facts and values does not need to be resolved (Moss, 1998a). Consisting of less than 20 words, Messick’s (1989) framework not only makes a major contribution to our understanding of test validity, but is also conceptually economical and elegant. Although the model has generated considerable controversy, the definition of validity in successive versions of the AERA-NCME Standards has slowly evolved to embrace Messick’s (1998) four-faceted conception.

Applying the Adapted Messick's (1989) Framework to the Evaluation of Distance/distributed Courses

My model of evaluation for distance courses is adapted from Messick's (1989) four-faceted conception of validity. My adapted model is a multi-faceted, progressive matrix, based on both facts and values (Messick, 1989). These insights are absent from traditional distance education models, which tend to focus on lists of criteria with unspecified relationships.

Contributions of the Adapted Messick's (1989) Framework

Applying an adapted version of Messick's (1989) approach to the evaluation of distributed courses implies a comprehensive, integrated approach to evaluation in which multiple sources of evidence are assembled to provide a more complete picture than can be provided by traditional distance education models. The overlap of facts and values in Messick's (1989) framework can also inform evaluation models of distance courses, which have tended to ignore values. An investigation of unintended social consequences is another important aspect of the overall value of a distance course which has hardly been mentioned in the literature. Moreover, Messick (1988) believed that educational technology supports a unified conception of validity, and makes the traditional fragmented approach untenable.

I have adapted Messick's (1989) four-faceted conception of validity so that it can be used as a framework for evaluating distance and distributed courses (Figure 4).

	Interpretation	Use
Evidential Basis	Outcomes	Relevance/Utility
Consequential Basis	Value Implications	Social Consequences

Figure 4. An adapted framework for the evaluation of distributed courses

As shown in Figure 4, the value of a distributed course can be conceptualized as a four-faceted construct: 1) *Outcomes*, 2) *Relevance and Utility*, 3) *Value Implications* and 4) *Unintended Consequences*. In Figure 4, *Learner Response* and *Relevance* encompass intended effects, and reflect the notion there are two kinds of intended effects: 1) achieving course objectives and 2) meeting the educational and training needs of society. The label of the upper left box is often used in quantitative evaluation studies. The first component of the *consequential basis* is *value implications*, and the second component is *unintended consequences*, both instructional and social.

The evidential basis of evaluation

In my adapted framework, the *evidential basis* of evaluation includes outcomes and *Relevance/utility*. *Outcomes* includes intended effects such as learner satisfaction ratings, student grades and completion rates. Learner satisfaction can be broken down into satisfaction with course components, such as access, materials, technology, and interaction. The *evidential basis* is in effect a comprehensive assessment of all intended effects as envisioned by the course developers and instructors.

Relevance/utility embraces how the instructional design fits with the educational and training needs of the global information economy, as well as cost-benefit analysis.

Relevance is defined as the link between the course activities and the needs of an information-based society, which requires lifelong learning and a continuous cycle of retraining (Bates, 2000; Rowley, Lujan & Dolence, 1998). The Ministry of Advanced Education, Training and Technology, Government of British Columbia (2000) expressed concern that the post-secondary sector is not changing fast enough for BC to make the transition to a knowledge-based economy. Moreover, the urgency for the public post-secondary sector to adapt to these changes is reflected in the funding strategies of the federal government.

Institutions would be required to adapt to meet consumer demand or diminish in size and importance. This approach, which is favoured by the federal government and based on the conviction that the public system cannot change enough to meet the demands of the current environment, is the rationale upon which the federal government has decided to eliminate direct government-to-government training purchases. (p. 19)

Another related aspect of the value of distance courses is the relevance, authenticity and students' ability to transfer their knowledge and skills the external contexts. Finally, *Utility* refers to a cost/benefit or cost-effectiveness analysis, which is pertinent to an assessment of value because of the high cost of technology, and the opportunity costs involved in funding innovation.

The consequential basis of evaluation

The *consequential basis* of evaluation comprises both value implications and unintended consequences. *Value implications* refers to an appraisal of the value loadings of language used in the course outline and materials, and a clear statement of the theory behind the course and the ideology in which the language of the course is embedded. Whether these values are explicit or lurk unexamined, a course evaluation should include a statement identifying the values, theory and ideology on which the course is based.

As for defining the term “unintended consequences”, Messick (1989) has already pointed the way forward in his discussion of test validity. Borrowing from Messick’s (1989) definition of unintended consequences, I am defining the unintended consequences of a distance course as the unanticipated effects of *legitimate* course implementation. Unintended consequences are the unexpected effects which arise when the course is implemented as it is intended to be implemented. The intended implementation can be identified from the course outline or interviews with the course designers and instructors. The misuse, improper or careless implementation of the course are not part of my definition of unintended consequences.

In my evaluation model for distance courses, there are two types of unintended consequences—instructional and social. Unintended instructional effects are implementation problems with a broad range from the course design to technical glitches. Unintended social consequences include diversion of funds into flashy technology of questionable value (Lookatch, 1997), the isolation of distance learners, lower social status with instructors re-labelled as “human back-up support” (Leslie Buffam, March 23, 1999, personal communication) and the unconscious assimilation of American cultural values in

learners (Fabos & Young, 1999). Electronic monitoring of workers' communications, restricted access to information and other social consequences of technology in the workplace have also been identified (Menzies, 1989; Noble, 1999). In addition, the unintended consequences of distance courses also include positive effects such as an enhanced internationalization for the university. Finally, my use of the term "unintended consequences" does not refer to the unexpected effects of evaluation, but to the unexpected effects on teaching and learning. The former falls under meta-evaluation, which is outside of the scope of this research.

Summary

In Chapter 3, I have presented Messick's (1989) four-faceted conception of validity, and clarified his meaning of the term "unintended consequences". As a framework for assessing the value of the education-related activity of testing, Messick's (1989) framework brings together key concepts that are also applicable to assessing value in a different context—the evaluation of distributed instructional programs. Messick (1989) provides many insights into "unintended consequences" which can respond to the call in the literature of evaluation of distance education to investigate this area. As such, Messick's (1989) framework can provide a bridge between the literature of evaluation in distance education with program evaluation. Moreover, in his discussion of values, he raises another important issue which has not been addressed in the literature of evaluation in distance education, where there are almost no recommendations to investigate underlying theory or values in which these courses are embedded. Yet unintended consequences and values are closely linked, and judgements of worth or value require both scientific evidence and an appraisal of value implications (Messick, 1989).

In this chapter, I have adapted Messick's (1989) framework and given some examples of the kinds of issues which may emerge from applying this framework to the evaluation of distributed courses in general. In recognition of Messick's contribution, this model will be called "the adapted Messick's (1989) model" because it is almost identical to Messick's framework, except for some re-labeling, and the underlying assumptions are also the same. In this chapter, I provided evidence to support the notion of unintended consequences from the literature of educational technology and distance education. With the adapted Messick's (1989) framework, I am introducing program evaluation into distance education, which has hardly been done, and I am shining a light on the bridge which links these two fields, and that bridge is unintended consequences. From here, I will go on to discuss the methodology involved in applying this framework to the data from three post-secondary distributed courses in BC.

CHAPTER 4: The Methodology for an Empirical Application

Overview

In Chapters 1 and 2, I laid out the purpose of this research and the review of the literature in which this research is situated. This research responds to Rumble's (1981) call to use unintended consequences as an evaluation category for distributed courses, a call which has largely gone unheeded. My purpose is to apply the adapted Messick's (1989) framework on validity to the data from three distributed post-secondary courses in BC under the *Learning through New Technologies: The Response of Adult Learners* project (Ruhe & Qayyum, 1999) to see how the model performs on a "test-drive" with real data.

In Chapter 3, I laid out an adapted version of Messick's (1989) four-faceted framework of test validity, and showed how it can be linked to the evaluation literature of distributed courses. I have chosen the adapted Messick's (1989) framework as the basis for a new model of evaluation for distributed courses because it includes criteria in traditional models as well as a new criterion-- unintended consequences. As a progressive matrix, the adapted Messick's (1989) framework shows that facts overlaps with values, which is an important assumption not mentioned in traditional evaluation models of distributed courses.

In Chapter 4, I will describe the process used to gather the data and to apply the adapted Messick's (1989) framework to the data from three post-secondary distributed courses in BC. The reason I am applying the framework to the data from these courses is to see the adapted framework "in action", that is, to learn how it works in evaluation practice, what insights can be obtained from its application which would not be obtained from traditional evaluation models and what kind of hurdles or shortcomings arise when applying this framework to real data.

Background—The Response of Adult Learners Project

This research is based on data collected for the *Learning through New Technologies: The Response of Adult Learners* project, which was undertaken to study learner satisfaction with courses based on new technologies (Ruhe & Qayyum, 1999). This project, which was funded by the Office of Learning Technologies (OLT), Human Resources Development Canada, was conducted by Distance Education and Technology at the University of British Columbia in partnership with local British Columbia Community Skills Centres, Okanagan University College, Open Learning Agency of British Columbia, Simon Fraser University and the University of Victoria. The project took more than two years to complete, and the final reports were disseminated in the spring of 2000. In this project, Bates' (1995) ACTION framework was used to frame questions and code the data from 13 post-secondary courses in BC.

I was one of three researchers who conducted the Response of Adult Learners project, and was given permission to use the data for this dissertation by Dr. Tony Bates, the Director, Distance Education and Technology (Appendix A). I used the ACTION framework to analyse the data for five case studies in this research, including MCSE, Psychology 101 and Modern Languages 400. This project was an "intrinsic" case study, where the focus was on understanding the cases, which were "dominant" (Stake, 1995). The focus was on full documentation of each learning context, and my analysis of these data gave me an intimate familiarity with the cases as stand-alone entities. My ideas about adapting Messick's (1989) framework arose after I had completed all my case reports and the cross-case comparison report for the Response of Adult Learners project.

Design

In this research, I applied an adapted version of Messick's (1989) framework to data collected for the Response of Adult Learners project. This research used the same research design as the Response of Adult Learners project, that is, an equal status mixed methods design in which quantitative and qualitative methods were used as equal and parallel methods (Tashakkori & Teddlie, 1998). The difference was that in the Response of Adult Learners project the cases were "intrinsic" whereas in this research the cases were "instrumental", that is, important for understanding something other than the case (Stake, 1995). The cases, then, were used to "test drive" the framework to see how it performed with real data. The framework was applied across three cases to multiple sources of data, including responses to questionnaire items, interviews of participants and stakeholders and an examination of documents and records. For this research, no more data was collected.

In this research, the focus was not on evaluating the cases, but on investigating the issues which emerged from applying the adapted Messick's (1989) framework to the data. My purpose was to determine relationships, probe issues, aggregate data and discover patterns or themes in the data. I revisited data which were gathered using a traditional evaluation framework, the ACTION model, and applied a new framework to see what issues emerge which did not emerge in the original analysis. Any framework uses categories and asks questions which other frameworks do not. My purpose was to see which new issues emerged from the data when a different framework was used to guide the analysis, to discover unique findings and to see what kind of gains resulted from applying this new framework to data from different subject areas and delivery methods for distributed courses.

Method

Sampling

The Response of Adult Learners project

In the Response of Adult Learners project, purposive and convenience sampling were used to select 13 cases, that is, post-secondary courses delivered through one or a blend of technologies (Table 2).

Table 2
Overview of Case Studies

Case	Institution	Delivery Mode	Learner Location	<u>N</u>	Class Size	Response Rate
AutoCADD 211	Kitimat CSC	Online	Remote	5	14	36%
Computing Science 315	UBC	F2F/ Online	On Campus	51	51	100%
EDST 565	UBC	Online	Remote	20	40	50%
Fine Arts 225	UVic	Online	Remote	5	20	25%
German 430	UBC	CD ROM	On Campus	23	23	100%
Math 235	Kitimat CSC	Online	Remote	10	10	100%
MCSE	Burnaby CSC	CBT	On Campus	9	16	56%
Museum Information Management	UVic	Online	Remote	10	18	56%
Resource & Watershed Management	UBC	CD ROM	Remote Mainly	17	23	74%
Recreational Vehicle Gas	OUC	Audio- Graphics	Remote Mainly	19	19	100%
History 120	OLA	Print	Remote	18	93	19%
PSYC 101	OLA	Print	Remote	24	93	26%
Wood Science 475	UBC	Face-to-face	On Campus	13	95	14%
Totals	N/A	N/A	N/A	219	501	44%

As shown in Table 2, CPSC 315, EDST 565, GERM 430, Psychology 101 had the largest sample sizes. In general, there were higher response rates when instructors gave class time for learners to complete the questionnaires. Lower response rates came from online learners who had no verbal communication with the researchers prior to receiving questionnaires by mail.

Sampling in this research

For this research, I used purposive sampling to select three information-rich cases from the 13 cases in the Response of Adult Learners project (Table 3).

Table 3

Overview of Case Studies Selected for this Research

Case	Institution	Delivery Mode	Learner Location	<u>N</u>	Class Size	Response Rate
Computing Science 315	UBC	F2F/ Online	On Campus	51	51	100%
Educational Studies 565	UBC	Online	Remote	20	40	50%
German 430	UBC	CD ROM	On Campus	23	23	100%
Microsoft Certified Systems Engineer	Burnaby CSC	CBT	On Campus	9	16	56%
Psychology 101	OLA	Print Distance	Remote	24	93	26%
Totals	N/A	N/A	N/A	219	501	44%

To select cases, I used a process of elimination, which I will now explain. First, cases with too little data were eliminated. MIM (Museum Information Management), AutoCAD and Fine Arts 225, for example, were not selected because their case reports are based on ten or fewer responses to the questionnaire. Moreover, there was no interview data for MIM and

AutoCAD. Similarly, Resource & Watershed Management was not selected because this case study was based on two student interviews and no faculty/staff interviews. The second criterion for selection was relevance. Wood Science 475 was not selected because it is a traditional face-to-face course, but this research deals only with distributed courses.

The third criterion was redundancy of information. History 101, for example, was not selected because it is a first-year academic course offered in both print and online versions by the Open Learning Agency. Psychology 101 is also a first-year academic course offered in both print and online versions by the Open Learning Agency. For our purposes, the subject areas and delivery methods are redundant. As a result, Psychology 101 was chosen for this research because it has a higher sample size than History 101 (24 questionnaire responses and two interviews, as opposed to 18 questionnaire responses and one interview for History 101). The fourth criterion was availability. Although it had a high response rate and three interviews, Recreational Vehicle Gas was not chosen because this case was written by another author, and I was unable to locate the data. EDST 565 was not chosen, because after the three cases in this study were analysed, the data was saturated, and further analysis would have added little to the findings.

In sum, I used purposive sampling to select three cases for this research—Modern Languages 400, Microsoft Certified Systems Engineer and Psychology 101. Out of 13 cases, these three cases were information-rich, and generated many different kinds of insights, thereby maximizing our opportunity to learn from applying the framework to the data. Just as important, these three cases represented different subject areas and delivery methods, thereby providing a sufficiently broad range of academic subject areas and

delivery methods, which is important for enhancing generalizability in evaluative case studies (U.S. General Accounting Office, 1982).

Participants

The participants were learners, instructors and course designers who agreed to volunteer to complete a questionnaire and be interviewed. The learners varied considerably in age, employment status, subject area of interest and location in BC. MCSE tended to have mature employment-oriented males, while Psychology 101 had mostly mature female distance learners. In Modern Languages 400, there was an approximately even distribution of young on-campus males and females.

Ethical Issues

The ethical procedures of this study were those of the OLT project, which were approved by the Ethics Review Committee of the University of British Columbia (Appendix B). The freedom to choose to participate was not limited by age, gender or any other factors. Learner anonymity was maintained by the use of learner ID numbers instead of names. Because of concerns about learners' rights to privacy among some of the partner institutions, some learners could not be telephoned after the mail-out to encourage them to complete the questionnaires. Permission to obtain student grades could not be obtained from any of the partner institutions, and information on student performance is therefore absent from this study. The name of one of Ruhe's (1999a) case studies has been changed to provide anonymity for the course developers and instructors.

Consent

In the courses in this study, the researcher asked for learners to volunteer to complete the questionnaire. Those who completed the questionnaires were deemed to have

consented to being contacted by phone or by email for interviews. Before being interviewed, learners were asked to sign a subject consent form (Appendix C). Instructors and course designers were asked to sign a different consent form (Appendix D). All interviewees signed the consent form before they were interviewed. Student interviewees were given \$20.00 for their participation. Faculty members and course designers were not paid, except for OLA instructors, who expected it, and who were paid the OLA meeting rate of \$27.40 for a one-hour interview.

Stakeholder influence

Because this research was an evaluative case study, I adhered to ethical standards of evaluation as articulated by the Joint Committee on Standards for Educational Evaluation (1994). In addition, the multiple perspectives of stakeholders were represented in the case reports, as recommended by Stake (1995). Finally, to avoid undue influence from stakeholders, I chose committee members who had no vested interests in this research.

Instrumentation

This research used a questionnaire and two interview schedules to collect data. The questionnaire, consisting of 90 quantitative variables and six open-ended items, was designed by a core group of BC researchers from the partner institutions of the Response of Adult Learners project. Interview schedules were used to obtain information about the perceptions of the course designers, instructors and learners. The interview findings were used to shed light on the issues by revealing learners' perceptions of how the work had worked for them.

Questionnaire

A questionnaire with 90 quantitative variables and 6 open-ended variables was constructed by a committee of project associates from the OLT partner institutions, pilot tested and revised (Appendix E). The questionnaire items were based on Bates' (1995) ACTION framework, and measured learner response, costs to learners and access to technology. The six open-ended items dealt with problems encountered by learners, benefits and drawbacks of the delivery method, support services, recommendations for additional support services, suggestions for improvements in support services and suggestions for changes to the technology, and course materials and suggestions for improvements.

Interview Schedules

The interview schedules for this research were based on the ACTION framework. One schedule was used for learners (Appendix F), and another for instructors and course designers (Appendix G). Topics included the teaching goals/ learning outcomes and teaching strategies, teaching methods that suit this subject matter, teaching strategy, the role of learner and the role of tutor/instructor. Some questions targeted perceptions of the course components: materials, technology, delivery, interaction, assessment and support. Other questions dealt with institutional factors which facilitated the learners' use of technology and enhanced access. Questions were open-ended, and the interviewers felt free to depart from the questions based on the respondents' previous replies, and to introduce new ideas or topics where appropriate.

Procedures

Data Collection

All of the quantitative and qualitative data for this study was collected for the Response of Adult Learners project in 1998 and 1999; no additional data were collected for this research. Quantitative data was collected from a questionnaire with 90-variable questionnaire that included six open-ended questionnaire items. For courses with face-to-face components, questionnaires were distributed in class to onsite learners who volunteered to complete them. For online courses, questionnaires were mailed out to learners. Interviewees were selected from the names of learners who had completed the questionnaire. When there were too many volunteers for interviews, names of interviewees were drawn from a hat. The selection of interviewees, then, was random, except in one course, where interviewees were pre-selected by the instructor.

Interviewers for this research were trained in interview procedures recommended by Patton (1990). Most interviews were conducted by phone and tape-recorded, with the exception of the interviews with the MCSE administrator, which were done face-to-face and tape-recorded. Before interviews were conducted, questionnaire responses were reviewed for response patterns of interest which could be probed further during the interview.

Although an interview schedule was used, the interviewers sometimes asked questions which departed from the interview schedule. Therefore, although the interview questions provided a common overall direction, the interviewers were allowed to probe into areas of interest which emerged from questionnaire responses or prior interview responses. All interviews were transcribed by a paid professional transcriber who had no vested

interest in the findings. For this research, no more questionnaires were distributed, and no more interview data was collected. Case reports written by me (Ruhe, 1999a, b) and by Qayyum (1999) were also used as data for this study.

Applying the Framework to the Data Using a Mixed Methodology

The analysis for this research was mostly deductive, although the analysis sometimes shifted towards induction in the later stages, pointing to weaknesses or extensions of the coding categories. To apply the adapted framework to the data, I used the labels of the four facets as coding categories. For the evidential basis of the framework, there were three coding categories: 1) Learner Response, 2) Relevance and 3) Cost-benefit. The consequential basis also had three coding categories: 1) Value Implications, 2) Unintended Instructional Consequences and 3) Unintended Social Consequences. Each of these constructs was operationalized with selected questionnaire items, and used to code salient excerpts from the interview data.

Because each of the coding categories was operationalized by one or more selected questionnaire items, the response distributions, means and standard deviations on learner responses to these items were reported for each case. The construct of learner satisfaction was operationalized by learner response to the following questionnaire items: 1) "I like this delivery mode because it gives me flexibility in my studies (e.g. time, place, location)", 2) "If this course was not offered in this delivery mode, I would not be able to complete it.", 3) "How do you rate the course materials?", 4) "In this course, the interaction with the instructor is relevant to my learning", 5) "In this course, the interaction with the other students is relevant to my learning" and 6) "The marking is fair." These items measured satisfaction with the following key course components: 1) flexibility, 2) materials, 3)

interaction with the instructor, 4) interaction with other learners and 5) assessment.

Although other items measured satisfaction with different aspects of the same components, the above questions were chosen because they measured overall satisfaction with essential components, and because there was a need to place reasonable boundaries on the analysis.

The construct of relevance was operationalized by learner response to the following questionnaire items: 1) "The course materials are relevant to my personal or professional needs", 2) "Using technology in this course helps me to learn more relevant information", and 3) "Using technology in this course helps me to learn with greater depth of understanding". To operationalize the construct of cost-benefit, I analyzed responses to the questionnaire item 1) "What are the most important benefits of this delivery method for you?" 2) "What drawbacks, if any, are there?" 3) "This course is not worth the time." and 4) "This course is not worth the money."

To apply the consequential basis of the adapted framework to the data, I began by analyzing responses to the questionnaire items under cost-benefit. Next, to operationalize value implications, I did a qualitative analysis of the content of course outlines and instructor interviews. Unintended consequences were analyzed in three ways: 1) learner responses to the questionnaire item on drawbacks ("What drawbacks, if any, were there?"), 2) a comparison of the normative theories (how instructors and course designers believe their courses should work) with the actual course implementation (how the courses actually work) and 3) observations of a lack of "fit" among the course components. Both Unintended Instructional Consequences and Unintended Social Consequences were used as coding categories.

My approach to coding qualitative data was recommended by Miles and Huberman (1994). Data was coded under themes, the themes being the labels of the four facets in the adapted framework. First, I used the categories of the framework to code open-ended questionnaire responses, interview transcripts, course outlines and web pages. Using electronic transcripts of interviews, tabulated responses to the open-ended items and codes to identify sections of paper documents, I highlighted salient text and code them for the categories of the adapted Messick's (1989) framework. The purpose of the qualitative analysis was to provide a "rich, "thick" description" (Merriam, 1989, p. 11) to "flesh out" the quantitative results. An "aggregation of instances" (Stake, 1995, p. 74) under a common theme was a pattern, and lent supports for that element as a valid and useful evaluation category for distance education programs.

Convergence between the quantitative and qualitative findings was used to validate the findings. If there was convergence between the themes and the data, then these findings validated the framework. As data were compared across cases for recurring patterns and critical differences, the focus of the analysis sometimes shifted towards induction, and the categories of the adapted Messick's (1989) framework were re-examined in response to emergent issues. This coding process continued for each case until the data were saturated.

Summary

In this research, a mixed methodology was used to assess the merit or worth of three BC post-secondary distance/distributed courses. Survey findings obtained in the OLT study were used to assess learner satisfaction and to better understand the implementation systems. In the qualitative component, the categories of the adapted Messick's (1989)

framework were used as coding categories. The coding was mostly deductive; that is, the framework was “held up” against the data and salient excerpts were identified. The categories of the framework were “a priori” themes, but new themes emerged as the coding progressed. In the final stages, the qualitative and quantitative findings were compared for convergence.

Validity Issues in Evaluation Research

Triangulation

Triangulation is very important for enhancing the trustworthiness, or credibility, of case studies (Flick, 1992; Merriam, 1989; U.S. General Accounting Office, 1990). In this research, data was triangulated across sources; for example, open-ended survey responses were compared with interview data. Information from documents, course outlines, web pages, interviews and open-ended survey items were compared for convergence. Data were also compared across respondents and the convergence of multiple perspectives increased the confidence of accurate interpretations. Finally, data were compared across methods and cases. In sum, triangulation across sources, participants, methods and cases provided the required checks and balances which enhanced the validity and generalizability of this research.

The adapted framework was used as a matrix of coding categories to organize the evidence under each of the themes of the four boxes. By performing a similar function to factor analysis, i.e. assembling data according to the construct they measure, matrices produce sharply defined, measurable constructs, thereby bolstering construct validity (Eisenhardt, 1989; Miles & Huberman, 1994). Finally, an independent researcher was asked to confirm the coding of data, and disagreements were negotiated.

Because words can take on different and ambiguous meanings depending on their contexts, my comments and coding categories were written directly beside salient interview quotations. Memo-ing was used to provide a record of my assumptions, reflections and biases which surface during the process of coding the data (Lecompte & Preissle, 1993) (Appendix H). To build credibility, I wrote down marginal notes, "a stream-of-consciousness commentary consisting of hunches, observations, questions and critical self-checking" (Van Maanen, 1988, p. 150). As recommended by Stake (1995), I searched for additional interpretations rather than confirmation of a single meaning. My method was memo-ing, that is, writing notes about what the data meant and how they were linked to other data segments.

Trustworthiness was also built into the research by using an audit trail to create a chain of evidence (U.S. General Accounting Office, 1990). Coded transcripts, documents and memos were filed in a sound organizational system. All filed items were critical components of an audit trail (Merriam, 1989) which allowed an auditor to determine reliability, that is, whether conclusions were warranted by the findings. The research findings was triangulated with the literature, as recommended by Eisenhardt (1989).

In conclusion, the above logic-in-use struck a balance between control and creativity, as recommended by Eisenhardt (1989). That is, attempts to reconcile evidence across cases, types of cases, and different investigators, and between cases and literature increased the likelihood of creative re-framing (Eisenhardt, 1989). Finally, using unintended consequences as a coding category provided controls for "bad news" selectivity, i.e. bias towards programs that work (U.S. General Accounting Office, 1990).

Generalizability

Paradoxically, triangulation provides the controls to build credibility, but also provides mechanisms which lead to the elaboration of a theory with stronger internal validity, and wider generalizability. This research used fuzzy generalization, in other words, generalization were tentative, using language such as *it is likely that*, or *it may be that* (Bassey, 1999).

Utility

Finally, the usefulness, or utilization value, of this study is based on three factors. First, this study brought key concepts from the literature of program evaluation into an evaluation model for distributed courses. This research validated the application of a framework which responds to Rumble's (1981) call to investigate unintended consequences of distance instructional programs. In this way, this study filled a gap in the evaluation literature of distributed learning. By using the adapted Messick's (1989) framework as a set of organizing principles for investigating unintended consequences, I brought his many insights into unintended consequences into the literature of evaluation of distributed courses.

Summary and Conclusion

My purpose was to introduce and apply an adapted version of Messick's (1989) four-faceted conception of validity to data from three distributed post-secondary courses in BC under the Response of Adult Learners project and discuss emerging issues and implications, including how this approach worked, what kinds of themes emerged and whether the framework needed to be adapted or refined. Using a mixed methodology, an adapted version of this model was applied to three case studies which were purposively

chosen to provide information-rich data from different academic subject areas and delivery methods.

The facets of the adapted version of Messick's (1989) framework were used as coding categories. The qualitative findings from this analysis were compared for convergence with the quantitative findings. Finally, the findings from the reports based on the ACTION framework were compared with findings based on the analysis with the adapted Messick's framework. Issues which emerged from the second analysis which did not emerge in the first analysis were used as evidence to show the value of using the adapted Messick's (1989) framework to guide the evaluation of distributed courses.

CHAPTER 5: Results

Overview

In Chapter 3, I discussed the philosophical underpinnings of Messick's (1989) model, that is, consequentialism, and the critics of the model. I then introduced a model of evaluation adapted from Messick's (1989) framework on validity and discussed its four facets in detail. I then "fleshed out" these four facets by looking at concepts which might "fit" within each of them.

In this chapter, I applied the adapted framework to the data from three post-secondary distributed courses. For ease of reading, I am reproducing the adapted framework with a few categories that were implicit in the previous version, and without "grade" because permission to obtain student grades was not obtained in the Response of Adult Learners project (Figure 5). Learner satisfaction is a common measure used in evaluation studies of distance courses (Harrison, Saba, Seeman, Molise, Behm, Saba, Moline & Williams, 1991).

	<u>Interpretation</u>	<u>Use</u>
<u>Evidential</u> <u>Basis</u>	Learner Response/ Satisfaction	Relevance (Constructivist Learning/Workplace) Utility (Cost-benefit)
<u>Consequential</u> <u>Basis</u>	Value Implications	Unintended Instructional and Social Consequences

Figure 5: An adapted framework for the evaluation of distributed courses

To apply this model to the data, I cycled through the four facets of the framework, which was operationalized as questionnaire items and coding categories for qualitative data. As each course was “held up” against these four facets, salient data were identified, and used to demonstrate how each category was relevant and useful, and/or how the category needed to be refined. My purpose was to gather, organize and synthesize all available sources of evidence for the use of the adapted framework as a model of evaluation for distance courses. In doing so, I expected to bring issues to the foreground which would have remained in the background with the use of the ACTION model.

Let me explain what I mean by cycling through the framework. It seems reasonable to begin the cycling process with an overview of the course, and the theory and ideology underlying its development and structure (how instructors and course designer believe the course should work). However, this approach required me to begin the cycling process by starting from the lower left box of the adapted framework, with “Value Implications”. Next, I moved in a clockwise direction to the upper left facet, Learner Response. Next, I coded data for Cost/benefit and Relevance, with relevance being an aspect of benefit, and with the overall focus of this analysis being on how the course “fits” into the particular context and application in which it is situated. For example, “Relevance” can be operationalized as learner perceptions of the relevance of the course to their learning and to their professional and work-related needs, including their need for flexibility. This category can also be understood to refer to an appropriate use of technology, e.g. whether a multimedia-based course is designed along a constructivist approach to teaching and learning.

Next, I moved to the lower right facet, Unintended Consequences, where I compared the data coded under Value Implications (how the course should work) with the actual course implementation (how the course actually works). Observations of a lack of “fit” among the components of various courses constituted evidence for unintended instructional consequences, and the category “Unintended Social Consequences” was examined with qualitative data. Finally, in an overall evaluation of the course, I tied the findings together in a way which described the overlap and inter-dependencies among the various evaluation components. In this wrap-up, the inter-dependencies between values and facts, costs and benefits, and intended and unintended consequences which played themselves out in each course were described and evaluated.

In sum, my purpose was to amass and organize the evidence for the use of the model. My emphasis throughout this process was not an evaluation of the course per se, but rather, an investigation of how the model worked in practice by documenting the kinds of issues which emerged from its application which might not have emerged from the application of traditional evaluation models without Unintended Consequences or Value Implications as evaluation categories.

Modern Languages 400

Modern Languages 400 was a multimedia (CD-Rom) self-study course on reading a foreign language for professional and technical purposes. The delivery mode for Modern Languages 400 is “multimedia”, or CD-Rom. This beginners’ course had no pre-requisites and focused on the teaching of basic reading skills, which prepared students for the follow-up course, Modern Languages 401, a reading course for professional, technical and academic purposes. Modern Languages 400 and 401 were offered as a double section in the

spring of 1998, but in the summer of 1998, this double section was divided into two separate courses and renumbered.

Modern Languages 400 students were required to purchase a print package; this package allowed them to study at home instead of coming to campus. Learners were required to work on the print package at home and attend class once a week where they checked their answers on the CD-ROM and have access to an instructor who plays a facilitator role. The two other weekly on-campus meetings were optional. Learners were responsible for five units per week and the recommended time to spend on the course is six or seven hours per week. There were two tests, worth 50%, a final exam, worth 40% and an online participation activity, worth 10%. All 23 learners in the course filled out the questionnaire for a 100% response rate. There were three participant interviews and one instructor interview.

Modern Languages 400 had run in a face-to-face mode for several years, and the materials were migrated onto the CD-Rom version of the course. There were 26 introductory modules (e.g., basic grammar, vocabulary and reading skills) on CD-ROM and in a print package. All students were required to do the introductory modules, and choose from other modules in different topic area such as Business, Music, Natural Sciences. Modern Languages 400 resembled a print distance (guided home study) course, except that the learners were required to come to campus once a week to check their answers on the computer. Although the CD-ROM was described as "multimedia," the units were text-based. Moreover, the CD-ROM and the print package had the same texts, instructions and tasks. The only difference was that the CD-ROM contained the answers to the exercises in the print package, happy faces, and highlighting features. There were few

pictures and no sound in the CD-Rom, but the CD-Rom had the answers to the exercises in the print package, which had no answers.

Value Implications

The history of development began 7-8 years ago, with a course designer/instructor who was disillusioned with traditional ways of teaching. Modern Languages 400 was based on readings of different content areas such as cultural knowledge, business, music and natural sciences. The course designer's rationale was that "an approach to language teaching based on specialized content can build on a larger number of language cognates, formulas and other universal elements in language, so that the acquisition of the second language can be accelerated" (Roche, 1997). The Modern Languages 400 instructor did not lecture unless there were problems, but circulated and paid attention to individuals; consequently, informal assessment was easier. According to the instructor, "As a teacher, you are freed up and can walk around" and "You can't throw people into foreign language; you need to mediate and provide tutorial assistance."

In the course outline, the instructor stated that the benefits of this course were self-directed pacing, flexibility, acquisition of knowledge in the students' subject matter areas of interest and faster progression (a second year reading knowledge in just one term, according to the UBC calendar) and a higher level of proficiency. The course designer believed the new approach "really improves instruction," and the media allows individual pacing and flexibility. He added that it "frees up time" and allows him to "pay more attention to individuals, do better informal assessment and get to see individuals as they work". The instructor noted that for some students the dependence on computers was a hindrance, and that there were limited opportunities for student-student interaction. The

approach required self-discipline because students work largely on their own. The student's role is more demanding, and there is less "down time." The instructor acknowledged that "overall, the course works well, although it is not for everyone, and there are trade-offs, but the benefits outweigh the problems".

Learner Response

Flexibility

Almost all of the learners in Modern Languages 400 said that they liked the delivery mode because it gave them flexibility (Table 4).

Table 4

Response to "I like this delivery mode because it gives me flexibility in my studies (e.g. time, place, location)." (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	2	4	6	26	9	39	6	26

Response to "If this course was not offered in this delivery mode, I would not be able to complete it." (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
11	48	7	30	4	17	0	0	1	4

When asked why, two student interviewees said that they could work at their own pace without weekly deadlines, which allowed them to “even out” the workload from their other courses during the semester. Finally, they said that they could still have taken the course if it had not been offered in a flexible format. This finding showed that these learners may have had to be on campus for other courses, and that although flexibility was a benefit, it was not an insurmountable barrier to accessing the face-to-face version of the course, especially if they were on campus for other courses.

Materials

As shown in Table 5, 74 % of students rated the materials as average or below average.

Table 5

Response to “How do you rate the course materials?” (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	4	6	26	10	44	5	22	1	4

Source: (Ruhe, 1999a)

The reading skills which form the course objectives included skills such as getting the overall gist of a passage without resorting to the dictionary for every word. “We learned how to pick out from a text in German how to get the..a general meaning of it by picking out the words we sort of recognized”. The content of the course materials included

readings on traveling, foreign culture, e.g. system of government and schooling, and fairy tales.

Learners who responded positively liked the special features of the CD-Rom, which include happy faces, highlighted text, pop-up messages like “congratulations” for right answers, travel units, visuals, vocabulary matching games and sound. For those for whom the course worked well, the content of the readings, combined with the special features of the computer was intrinsically motivating.

I am really excited when I press the computer button and see the answer, it is better it's better to learn a language from this approach. Because I think language is not..[an] interesting course to learn, but if you learn it this way, it helps us to study.

When you type in the answer and then if my answer is correct, it will pop up some message like congratulations or something like that.

Compared with the responsiveness of the CD-Rom, the print materials appeared dull to some of these learners. As one interviewee said, “The book doesn't highlight the important ideas. And it's quite difficult to read and it's quite boring, I just hate reading it.”

Another learner said that the special features of the CD-Rom materials made learning interesting for its own sake.

I really like it's I study because I am interest in because interest in it, but not because the exam...yeah, so everyone time when I click on the computer I feel excited because I'm not..just study for exam, but I'm learning yeah...

I think it's less boring and I'm I really think that I'm learning from this course, like...the culture ...[it's] not exam-oriented. I study because I'm interested, but not because of the exam.

For these learners, the technology enhanced the course materials, but the reader should note, that these learners constituted only 25% of respondents, and there were no interviews with the other 76% of respondents, who responded to the materials with less enthusiasm. One interviewee said that the course would have been easier if she had been allowed to look up the meanings of words in the dictionary before the first reading, a strategy which was discouraged.

There were many recommendations for improved support services (Table 6).

Table 6
Support Services—Recommended Improvements

	N=20	%
Assignments/quizzes	5	25
More technology	3	15
More access	3	15
More interaction	1	4
More lectures	2	8
More structure	2	8
More relevant content (North American)	1	4
More user-friendly	1	4
Weekly goals	1	4
Include answer key in the print package	1	4

Source: (Ruhe, 1999a)

As for the grading criteria, 26% disagreed that it was fair, 35% were neutral and 39% felt that it was fair. Weekly exercises were not graded. “We just go through them on our own, on the computer and it tells you if you are right or wrong.” According to one learner, “the questions are really good. They have a mix of everything. There are some multiple choice, and some translations from picking out grammar, quotes from sentences. Reading the paragraph and picking out words you recognize”. “It gives you three chances I think or you type in the first answer and if it’s wrong, and then you try again, over again”. One interviewee said “There wasn’t any checking to see if you kept up”. One interviewee really liked the use of technology to deliver quizzes and to provide quiz scores immediately after the test.

And its quite competitive and...it...gets more me more excited to do it than do the just the general German exercise. ...And sometimes the teacher or TA will... post the average or some statistics information about the quiz so I can compare my results with other people... And I like that too.

Forty percent of respondents were satisfied with the software, while 40% were neutral and 20% were dissatisfied (Table 7).

Table 7

Response to “I am not satisfied with the software used in this course.” (N=20)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
2	10	6	30	8	40	4	20	0	0

Response to “The technology increases my motivation to work on the course.” (N=20)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
5	25	5	25	6	30	4	20	0	0

Response to “The technology increases my motivation to work on the course.” (N=20)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
5	25	5	25	6	30	4	20	0	0

Response to “I can learn better using print materials than by working on a computer”.

(N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	4	6	26	4	17	5	22	4	17

Source: (Ruhe, 1999a)

In the interviews, the learners talked about the special features of the computer-based set of course materials. These features included highlighting of the important grammar points, instant feedback, pop-up messages like “congratulations” for right answers. As shown above, 20% of respondents agreed that the technology increased their motivation to work on the course, while 30% were neutral and 50% disagreed. There was also a wide spread in their responses to the relative value of print materials—39% agreed they could

learn better with print, 17% were neutral and 30% disagreed. This pattern of responses may have been an endorsement of the print package.

Interaction

In this course design, the instructor circulates and works with the learners on a one-on-one basis whenever they need help. When I asked the students if the instructor addresses the whole group at the beginning or end of class to get them started or wrap up the lesson, I was told that she would start out by pointing out problems previous learners had had in upcoming lessons or clarifying any points of confusion mentioned by several learners in the last class. In response to the questions "I am able to interact with my instructor as much as I want", 74% agreed and 26% disagreed (Table 8).

Table 8

Response to "In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want." (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	4	5	22	0	0	13	57	4	17

Response to "In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want." (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
3	13	5	22	2	8	9	39	3	13

Source: (Ruhe, 1999a)

The spread of responses was even wider for interaction with other students. Thirty-five percent felt they were not able to have as much student interaction as they wanted. As one learner said, "I don't think anyone in the class knows each other, unless they did before. It's...it's...very independent you don't meet anyone and there's no discussion or anything." There was a similar distribution for responses about the relevance of interaction with their peers to their learning.

Cost/Benefit and Relevance

Cost

Next, I looked at the cost to the institution, the instructor and the students. According to the course developer, Modern Languages 400 was "very expensive" to develop, with most of the development costs going to programming in the special features of the CD-Rom materials such as happy faces and pop-up messages. The reason that the materials had no sound and few pictures was that these features would require many more programming dollars than were available. With more funding, the programmer could have added in features which make better use of CD-Rom technology than the happy faces, highlighted text and pop-up messages like "congratulations" in these units. These features distinguish the materials from the print package on which they are based, making them more interesting and intrinsically motivating, which is important for self-study. The instructor felt that the university should be engaged in marketing the materials, and looking to recover costs and lease the materials to other universities and/or sell to publishers.

The instructor commented on the time and energy required to innovate, and the many costs of being a trail-blazer. He said that the institution "needs to clean up its act fast" if they want to encourage professors to do "top-notch stuff." He said that course innovation

was “not worth it in terms of time.” Compared to writing a book or developing a traditional course, developing a CD-Rom course was very time-consuming and the extra effort did not pay off in terms of tenure or promotion decisions.

Learners were divided about the value they felt they had received for their money (Table 9).

Table 9

Response to “This course is not worth the money it costs.” (N=22)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
2	9	6	23	7	32	5	23	2	9

Response to “I would not take another course using this delivery mode.” (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
6	26	3	13	8	35	2	8	4	17

Source: (Ruhe, 1999a)

In their responses to the question of whether they would take another course in this delivery mode, there was a wide spread. More specifically, 39% said they would, 35% were neutral, and 25% said they would not. Responses to an open-ended item on benefits and drawbacks on the questionnaire, however, showed both that enhanced flexibility was a major benefit for learners, and also that the course provided insufficient flexibility (Table 10).

Table 10

Benefits and Drawbacks of the Delivery Mode

Benefits	N=17	%
Flexibility	13	76
User-friendly	1	7
Instant feedback	1	7
Easy, quick to find answers	1	7
Portable materials	1	7
<u>Drawbacks</u>	<u>N=8</u>	<u>%</u>
Insufficient flexibility	3	37
No sound	1	12
No monitoring of progress	1	12
No student interaction	1	12
Hate computers	1	12
Impersonal	1	12
<u>Problems</u>	<u>N= 6</u>	<u>%</u>
Technical glitches/can't take CDRom home	2	33
Insufficient instructor interaction	2	12
Objectives unclear	1	12
Too easy to read answers first	1	12

Source: (Ruhe, 1999a)

As shown in the previous table, 13 out of 17 students mentioned flexibility as a benefit of the course. The most frequently mentioned drawbacks were insufficient flexibility, followed by inability to take the CD home. In fact, these two points are related. The learners liked coming to class once a week instead of three times a week, making it easier for them to schedule in their part-time jobs. But they were aware that the main reason they had to come to class at all was to check their answers to the exercises in the print package, which were not provided in the print materials.

In considering costs, they must be compared to benefits and the main benefit of the course to learners, as indicated in the previous section on learner response, is schedule flexibility. The requirement to come to class only once a week, instead of three times, allowed students to fit in their part-time jobs and even out the workload from their other courses. In turn, this benefit must be compared to the loss of benefit from the inclusion of two face-to-face classes, in particular, the lack of social interaction, which could have been used to practice related skills such as listening and speaking in ways which reinforced and extended the language skills obtained from reading, thereby furthering their acquisition.

Relevance

Forty-eight disagreed about the relevance of the course materials to their personal or professional needs (Table 11).

Table 11

Response to "The course materials are relevant to my personal or professional needs."

(N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
2	4	10	44	5	22	4	17	2	9

Response to "Using technology in this course helps me to learn more relevant information." (N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
5	25	5	5	4	20	6	30	0	0

Response to "Using technology in this course helps me to learn with greater depth of understanding". (N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	4	5	26	10	53	3	16	0	0

Response to "In this course, the interaction with the instructor is relevant to my learning".

(N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
3	13	4	17	10	13	6	26	0	0

Response to “In this course, the interaction with the other students is relevant to my learning.” (N=23)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
6	26	4	17	10	13	3	13	0	0

Source: Ruhe (1999b)

As shown in the previous table, learners were divided on whether the technology helped them learn more relevant information. Thirty percent disagreed that the technology helped them learn with greater understanding. Only 26% saw the interaction with the instructor as relevant to their learning. Forty-two percent disagreed that interaction with other students was relevant to their learning.

The spreads on the three items measuring overall satisfaction (i.e. “This course is not worth the time it takes to complete”, “This course is not worth the money it costs” and “I would not take another course using this delivery mode) raised the question of what kind of learners were in the high-, medium- and low- satisfaction groups. Because there were too few subjects for a correlational analysis, a series of cross-tabulations and chi-squares were performed. First, the proportion of employed learners who would not take another course in this delivery mode was much lower than the hypothesized proportion $\chi^2(2, N=23) = 7.11, p = .029$). There were no significant differences when gender, caregiver status or limited prior experience with the technology were examined. To summarize, employed learners were more willing than unemployed ones to take another course, while learners who are satisfied with the support services felt that the course was worth the time and those who are satisfied with the software were more willing to take another course.

Unintended Consequences

Insufficient flexibility

Learners were not allowed to take the CD-Rom materials home to work on them. Instead, they were required to attend on-campus meetings once a week to access the CD-Rom in the computer lab. The learners were unhappy because the policy requiring weekly in-class meetings made the course less flexible than it could have been. These learners knew that if they had been provided with an answer key in the print package, were allowed to take the CD-Rom home, or submit their answers electronically, they would not have had to come to campus at all.

Sometimes I would do it at home like do the package and then I would bring it school and compare it to the answers the computer. Because we didn't get to take the CD-ROM home, but yeah and then sometimes I would just work in the computer and then if I got the right answer I would put it down in my package. Yeah I could study from there.

Moreover, access to the computer labs was often a problem because of insufficient terminals or too many users.

Yeah we were able to access the computer, but we would go to the health center, but the part..was bad because the lab people go to the help center so sometimes you wouldn't be able to get a computer. Yeah so we would have to wait a longtime or come back later or even even when we came sometimes we just didn't get a computer and sometimes that was difficult yeah.

When asked how the course could have been improved, one learner said,

I think if we got to take it [CD-Rom] home, it would have helped. So we wouldn't always have to go line-up in the multi-media waiting for a computer yeah. I think that would have helped a lot.

As these comments suggest, the flexibility provided by this innovative course undershoots both 1) what the technology is capable of providing and 2) the needs of the market, i.e. the needs of learners (Bennett, 2001). The students were unhappy because the policy requiring weekly in-class meetings in the computer labs made the course relatively inflexible. The most common drawback mentioned in their written comments was that the course was not flexible enough. This finding suggests that learners are very appreciative of increased flexibility and that they will expect as much of it as the technology is capable of providing.

Mismatch between technology and subject matter

As often happens, this course was developed based on existing structures and materials onto which the new course was "grafted". According to the programmer for Modern Languages 400, the CD-ROM materials were very expensive to develop. This was the reason that they were text-based, with only a few animated features such as highlighting and happy faces. There were few pictures and no sound. Moreover, learners can "see through" this situation and are not happy with it. As one interviewee said, "It's exactly the same but it's not a print-out. I think that the computer program was designed from the textbook and not the other way around." These learners realized that the print package and CD-Rom contained material which was "exactly the same," except the CD-Rom had a few animated features and the answers to the exercises.

Interviews with the learners revealed that, although the course was described as multimedia, several of the CD-Rom units do not contain any sound. The CD-Rom units were mostly text-based, which is to be expected in a course on reading skills. The use of CD-Rom in a second or foreign language course was unusual; this medium is typically used to teach listening skills where the particular strengths of the medium can be more fully exploited in a reading skills course. The reason CD-Roms are being installed in language labs across Canada is because they provide benefits which the traditional audio-tape based language lab does not, namely, the ability to combine sound and visual and to provide instant replay of the video.

Conclusion

One of the values underlying the development of Modern Languages 400 was schedule flexibility for on-campus learners. The learners' response to the course was mixed, with bimodal distributions on several items. The quantitative results indicated that this course worked best for learners who were employed than for others. The cost of multimedia development was unknown, and learners' perception of the relevance of the materials was mixed. The unintended consequences were insufficient flexibility for learners and a mismatch between the technology and the subject matter as a computer-based component onto an existing print package to save development costs. Unintended social consequences include a challenge to traditional academic course structures and demoralization in the instructor, who felt the institution could have shown more support.

Because the data was based on the ACTION framework, some information which would be gathered using Messick's adapted framework is not available for analysis. First, the instructor's selection of student interviewees may have resulted in sampling bias.

Interviews with learners who did not respond well to the course about their feelings about the loss of face-to-face lectures and in-class social interaction, and their expectations of multimedia-based learning, would have informed this study. Secondly, I do not have information about the cost of course development, which can be substantial for multimedia.

Psychology 101 (Print Distance Version)

Psychology 101 was an introductory university course offered in both print distance and online delivery by the Open Learning Agency (OLA) in Burnaby, BC. At the time these data were collected, there were 73 students enrolled in the print version and 17 students in the online section. Nineteen print distance students and 5 online students responded to this questionnaire. Because the sample size for the online version was small, this case study will focus on the print distance version. However, the qualitative findings for the online version also provided an interesting perspective on the unintended consequences of revising a print course to include online components.

Course Overview

Psychology 101 (Introductory Psychology I) was a first-year survey introductory course in human behaviour. This course was often one of the first courses taken by someone who is beginning a Bachelor's degree in Arts or Science. Students chose the print or the online version, and were able to begin either version at any time of the year. Learners in the print version worked through the course materials at home and mail their assignments to their tutors. They had telephone access to qualified tutors (many have Master's or PhD degrees) for five set hours each week, with their long distance charges paid for by OLA. The normal completion time for Psychology 101 was four months.

Course Materials

The print version of Psychology 101 course package included a course manual, textbook, tele-course guide, videocassettes and assignment file (Open Learning Agency, 1996a, 1996b). The course manual provided a course overview and information on study strategies, grading criteria, exam procedures, course extensions and withdrawals, transcripts and support services such as financial aid and library services. Wade & Tavris's (1996) textbook entitled Psychology provided a general introduction to the subject. Werthman's (1996) tele-course guide was also required and consists of three videocassettes about famous psychologists and their experiments. The final grade was based on the following tasks: a) multiple choice test—10%, b) critique of a research article —15%, c) a mini-research investigation—15% d) practice exam —10% and e) final exam—50%. The final three-hour exam, which consisted of multiple-choice and short answer questions, is held throughout BC every two months.

Learner Characteristics

Seventeen (89.5%) of the 19 respondents to this survey were women and 2 (10.5%) were men. Their ages ranged from 18 to 53, with the mean age being 30 (SD = 9.29). Forty-seven percent were full-time students and 52% were part-time. Nine learners (39%) were employed and 10 (63%) were unemployed. Those who worked did so from 12 to 44 hours per week (SD = 17.29). Ten (58%) were primary caregivers. Sixty-three percent had limited or no prior experience with print-based distance education. Approximately 50% of the print distance learners in this section of Psychology 101 lived in Victoria or in Vancouver and the surrounding municipalities (Ruhe, 1999b).

Value Implications

Ideology

The Open Learning Agency (OLA) was a public post-secondary institution whose mandate is delivering flexible learning in a variety of non-traditional formats. OLA offers “hundreds of individual distance courses and over twenty degree programs” (BC Open University, 2000, p. 1). The underlying values were open and universal access to students in a “format, place and time frame” which works for students (p. 1). According to Piper (2000), Open University's VP, Education and Provost of BCOU, “education should be easily obtainable, and should not be exclusive. Our commitment is to ensure that all our educational services meet the needs and expectations of our customers” (p. 1). As a result, OLA tailored their courses and delivery methods so that they meet the needs of the unemployed, women with child-care responsibilities, the disabled, prison inmates, rural residents and urban learners who do not wish to commute to campus. Students chose OLA because there was no college or university in the town/area where they live, because they were mature learners struggling with job or child-care commitments or because they wished to blend courses from different institutions.

In line with their mandate to provide open and universal access, the Open Learning Agency also had an “Open Admissions policy”. Prospective students did not need to submit a GPA or transcripts, and there are no prerequisites for first-year courses. The range of services provided to learners included year-round course registration, prior learning assessment, home delivery, and scholarships and bursaries. In recent years, OLA had been migrating many of their print distance courses onto the world-wide web. They had an online shopping cart approach to course registration, and a web page which

provided details on various aspect of distance learning. In sum, the ideology of the- Open Learning Agency was a belief in restructuring learning so that it fit the needs of students, as opposed to expecting students to fit into traditional university structures.

Course Objectives

The main objective of Psychology 101 was to provide a general overview of the science of human behaviour. According to the textbook writers, their objective was to expose students to how research is done. Moreover, the tutor said that her whole focus was on providing students with an introduction to basic research methods and some practice in conducting a small research project, which was one of the course assignments.

Learner Response

Flexibility

Eighty-three percent of respondents liked the delivery method for its flexibility, which allowed them to study while blending work and child-care commitments, blending courses from other institutions and avoiding a daily commute to campus (Table 12).

Table 12

I like this delivery mode because it gives me flexibility in my studies, (e.g. time, place, location). (N=13).

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0	0	2	15	4	31	7	53

Source: Ruhe (1999b)

Responses to "If this course were not offered in this delivery mode, I would be unable to take it." (N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0	0	2	10	8	42	9	47

Source: Ruhe (1999b)

Eighty-nine percent of respondents said they could not have completed the course without this delivery method (Table 13).

Materials

Seventy-five percent rated the materials as good or excellent, 20% were neutral and none rated the materials as fair or poor (Table 13).

Table 13

Response to "How do you rate the course materials?" (N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0	0	4	20	13	65	2	10

Source: Ruhe (1999b)

Response to “I can learn better using print materials than by working on a computer (N=7)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
2	28	1	14	2	28	1	14	1	14

Source: Ruhe (1999b)

As shown in the previous table, 28% of respondents said they could learn better using print materials than by working on a computer, while 28% were neutral and 42% disagreed.

Because there are only 7 respondents, however, these results need to be confirmed with a larger sample size. Finally, because this is a print distance course, responses to questionnaire items on technology are inapplicable and will not be reported.

Interaction

Thirty-five percent of respondents were satisfied with the opportunities for interaction with the instructor, while 30% were neutral and 35% were dissatisfied (Table 14).

Table 14

Response to “In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want.” (N=17)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
0	0	6	35	5	30	6	35	0	0

“Response to “In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want.” (N=12)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
6	50	1	8	4	33	0	0	1	8

Source: Ruhe (1999b)

As shown in the previous table, 58% were dissatisfied with opportunities for interaction with other students, 33% were neutral and only one person was satisfied. Note that with this print-based delivery mode, there were no opportunities for interaction with peers who are not given each other's addresses or phone numbers. One interviewee said he wanted someone to "bounce ideas off of" before talking to their instructors, whom they perceived as using a specialized language which could be "intimidating".

Support Services

Forty-seven percent of respondents were dissatisfied with the support services, while 47% were neutral and 6% were satisfied (Table 15).

Table 15

Response to “Support services for this course are unsatisfactory.” (N=17)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
0	0	1	6	8	47	6	35	2	12

Source: Ruhe (1999b)

As shown in Table 16, recommendations for improved support services include more tutor availability, more student contact, faster delivery of course materials and grades, extended hours for labs, clearer exam questions and a brochure of student comments.

Table 16

Support Services—Recommended Improvements

More tutor availability	7
More student contact	4
Shorter times to receive course materials	4
Shorter times to receive marks	2
Extended hours for computer labs and library	2
Clearer exam questions	2
Brochure of students comments	2
Need counseling services	1
Local staffed center	1
Cellular phone support	1

Source: Ruhe (1999b)

The previous table shows that inconvenient or insufficient number of tutor office hours was the most frequently mentioned drawback of this delivery mode. One problem is that these calls are time-bound, that is, they can only be placed during scheduled tutor office hours. Another problem is that when tutors are busy with other students, the caller is kept on hold; one student said she spent an average of 15 minutes on hold.

Relevance and Cost/Benefit

Relevance

As shown in Table 17, 94% of respondents felt that the subject matter was relevant to their personal or professional needs, while 5% were neutral and none disagreed.

Table 17

Response to "The course materials are relevant to my personal or professional needs."

(N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
0	0	0	0	1	5	12	63	6	31

Response to "In this course, the interaction with the instructor is relevant to my learning."

(N=18)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No	%	No	%	No	%	No	%	No	%
1	5	3	15	7	38	6	33	1	5

Response to "In this course, the interaction with the other students is relevant to my learning". (N=19)

1		2		3		4		5		6	
No	%	No	%	No	%	No	%	No	%	No	%
1	5	1	5	5	26	2	11	1	5	9	48

Source: Ruhe (1999b)

Thirty-eight percent felt that interaction with the instructor was relevant to their learning, while 38% were neutral and 20% disagreed. Finally, fifty-three percent of respondents said that interaction with other students was relevant to their learning, while 26% were neutral, and 10% disagreed.

Costs

The mean total cost of the course was \$438.00 (SD = \$137.00). Seventy-two percent of learners felt the course was worth the money it had cost, while 28% were neutral, and none felt it was not worth the money (Table 18).

Table 18

Response to "This course is not worth the money it costs." (N=18)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0	0	5	28	7	39	6	33

Response to "I would not take another course using this delivery mode." (N=19)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0	0	2	11	8	42	9	47

Source: Ruhe (1999b)

They neither agreed nor disagreed that the course costs less than other modes of delivery ($M = 2.8$, $SD = 1.2$), which suggests that print-based courses are perceived as price-competitive. As shown in Table 13, 89% said they would take another course in this delivery mode, 11% were neutral, and none said they would not take another course this

way (Table 33). Finally, in the OLT study, there is no data on the costs to OLA of delivering PSYC 101.

Benefits

The benefit most frequently mentioned in learners' comments was the flexibility of learning at their own pace, time and location (Table 19).

Table 19

Benefits and Drawbacks of the Delivery Mode

<u>Benefits</u>	<u>N=18</u>
No conflicts with work/childcare/other responsibilities	8
Can complete at my own pace/time/location	6
Variety of teaching methods (text, video)	4
<u>Drawbacks</u>	
Inconvenient/insufficient tutor office hours	8
Delays in receiving mailed course material/marks/transcripts	7
Miss the interaction with instructor	2
Miss the interaction with other learners	2
Limited times for exam writing	1
Lack of motivation	1

Source: Ruhe (1999b)

For these print distance learners, the most important benefit of the course was that they could study at their own pace, time and location without any conflict from

work and childcare obligations. One interviewee said she was taking the course from home because she was disabled, another said there was no university where she lived, and the third wanted to avoid long urban commutes, and blend courses from different universities.

Unintended Consequences

Updating the course materials

One set of issues which the adapted Messick's (1989) framework brings to the foreground deals with relevance of the course materials. Because Psychology is a rapidly evolving field, there are short revision cycles for university textbooks. To be compatible with new editions of the textbook, OLA's course packages also have to be updated. It is a strength of Psychology 101 that the materials are frequently updated, and for this reason, the course can be said to meet the evaluation criterion of relevance. On the other hand, there is one unintended consequence emerging from these frequent revisions. The Psychology 101 course package has to wait in the revision queue along with other courses, and, according to the instructor, this delay sometimes results in delays in shipping materials to learners.

Interruptions in support services

The data for this study reveal that an extensive amount of coordination is needed to deliver print-based education, and that interruptions or breakdowns in support services can cause problems for learners. One "chronic problem", which has already been mentioned, is getting course materials out to learners on time. One interviewee said that delays in receiving his course package meant that he could not study during the summer as he had intended, thereby delaying his completion of the course.

They don't say that they'll ship it to you by any certain date, they just say it will get there when it gets there....when I was first registering I didn't realize that it would take like a few weeks before it got to you so it's pretty important to register early ... I just assumed that I would be getting like then right then. I guess it was inconvenient and I mean I had to make my self busy for three weeks... I had planned on working on the courses during the summer.

A second problem faced by this learner is that he had been sent the wrong cassette, which caused confusion because he could not follow the recommended sequence of activities. He chose to proceed but did so uncertainly because he knew he had missed some of the material.

Oh well I had got the wrong...cassette...and it ...said it was the correct cassette, but when I plugged it in...it was the wrong one so...and I was really stuck there...because the instructions had said that you are supposed to listen to the tape before moving on.. so it kind of messes things up a little a bit because you don't know exactly what you're missing on the tape...But I just pressed on and went ahead.

Another problem mentioned by all three interviewees is that phone contact with tutors and support staff is limited to fixed hours, which can be inconvenient.

I had a terrible time with the tutor thing. I never once spoke to my tutor...I need to be able to contact like days or hours or times that was very frustrating. I would be working Monday evening at 9:00 and I would look at the time and it would be too late, I couldn't call you know so then I had to wait until Thursday at 7:15 and then Thursday, the question was gone or I would forget again and on and on. That was very frustrating.

So I think that would help or if there was person that came to town that could answer your questions in person, because the phone doesn't cut it sometimes. Well you just I mean..if you have to deal with Financial Aid I mean she's only there like Wednesday and Thursday afternoon and Monday mornings, and it's the same as every big place you know, push this button for this option, it's frustrating when you need immediate answers.

All three interviewees said that they missed being able to contact people face-to-face.

"Sometimes I really want to talk to a real person."

One interviewee referred to the deleterious effects of interrupted support services on her motivation. "It's the frustration again of the waiting... I feel like I need to keep kind of going because when there's a break in between then it you kind of lose the ability I guess to study." She also spoke with excitement about a face-to-face computer course she had previously taken and the enjoyment of "bonding" with other learners. However, it didn't seem to occur to her that the problems with the print distance delivery method might not be her fault, but the result of poor course design, or a design that doesn't work for everyone. She was not happy with her grade in Psychology 101, "but that you know [was] probably my fault." "Maybe I didn't try hard enough I don't know".

The second interviewee also mentioned a litany of troubles, but again, this person blamed neither the delivery method nor the institution. Instead, she blamed herself. "I don't believe it's on the other persons..fault". The third interviewee also spoke about the negative effect of interruptions on his motivation. "If you get the wrong materials or whatever it really sets you back a lot. But if you get boggled down on the first little bit you

kind of lose your motivation to keep going and so it makes it more difficult." Yet like the other two interviewees, this person blamed himself, not the course providers. "But that's my own ignorance I had never phoned them and asked or anything so, so it's more my fault if anyone's".

Although any one interruption in support services can be difficult for learners, all three interviewees mentioned that that they had to deal with a series of such interruptions. In effect, the data suggest that they had to navigate a virtual obstacle course to succeed. One interviewee had a list of specific sources of frustration, including the confusing mix of multiple formats, the absence of someone to talk to, pressure to pay back her student loan and isolation from other learners. The second interviewee mentioned lack of clarity about the exam questions, feeling isolated from other learners, delays in receiving feedback, inability to contact tutors during set hours, a looming postal strike and the unexpected extra expense of sending work by Loomis. The third interviewee also referred to delays in receiving the course materials and being sent the wrong materials. In effect, interruptions in support services and various other difficulties reduced the access, flexibility and satisfaction of these print distance learners.

The pathos here is that, according to the instructor, some PSYC 101 learners have been out of school for years, "feel rusty", "hold their cards close to their chests" and need "courage to send in their first assignment". Such learners feel insecure and need support and encouragement, which is especially important because motivation is crucial for persistence. Yet all three Psychology 101 student interviewees mentioned a list of obstacles which they had to overcome to succeed, including delays in receiving materials, receiving the wrong materials, problem with student loans and inability to communicate. When

support services are delayed, interrupted or not well co-ordinated, the self-concept and motivation of these distance learners may suffer.

Unintended Social Consequences

Traditional print courses as ghettos

As discussed previously, OLA's mandate is to deliver print courses to all British Columbians. Although there may be interruptions in support services, the reader should keep in mind that OLA learners can indeed access education. Of their many delivery options, print distance learning meets the needs of low-income students who cannot afford a computer or Internet access. Yet one unintended social consequence is that individuals may become ghetto-ized in print courses because they cannot afford computers and high-speed connections. One learner recalled with distress a previous experience in a face-to-face class which had been deeply humiliating. "The instructor was no less than astonished that...I didn't have a computer, that the 2000 word paper had been typed on a manual typewriter. [She] let loose a loud revelation that informed the entire class." Another student said that she would have preferred to take an online course, but her family's resources were too limited for them to afford a computer, or even a VCR.

My husband works 40 hrs per week at a menial job making \$12.00 /hr. We have a 1 ½ year old son and an 11 year old from a previous marriage who visits every other weekend and on holidays. Most months we sneak past the wolf, other months we don't. It seems that his earnings (my husbands obviously—not the wolf's) make us the richest of the poorest, disqualifying us from most benefit programs (i.e. family bonus etc) and limiting others (i.e. GST, child tax credit, BC Med Premium Assistance etc). Amazingly, we usually end up owing on our Income Tax as well.

Our situation is temporarily acceptable because we have other goals in mind....

Pivotal to [our] plan are the D.E. programs, being that raising and caring for our son is a primary responsibility and not negotiable... [but some] technologies are simply beyond our means. I would love a computer and would like to take the online versions of my courses. Most print courses include video cassettes, and...our machine is on its way out. What then? I realize that these situational factors are my own responsibility ...it's just with limited resources and an unfavorable background (poor), it is extremely difficult to claw your way up to the lower middle class or even, in cases such as the VCR, maintain a level that permits the utilization of the least technological of the DE courses.

Nor is this situation remedied by government student loan policies, which were devised around the needs of traditional face-to-face students, and have not been revised to meet the needs of distance students. For this reason, these policies may constitute a barrier to access and once again, may leave low-income learners "ghetto-ized" in traditional print courses. One learner expressed her frustration with BC government policies and unsympathetic financial aid officials who will not provide loans to distance students to purchase computers, even when the courses she needs to complete her program are offered only online.

Yeah...its very frustrating because there's no communication, I can't just go to my Financial Aide office and you know talk to the person and say you know I'm frustrated could you tell me what's happening...Eventually I will have to get a computer because some courses that I will be taking in the future they are only offered on-line so...And then there's there's always these the brakes...I've just sent in all my applications for students for financial assistance...for my next two quarters...so at least then I can I will know if it's approved well in advance.

OLA's web page on financial aid mentioned a second problem with student loan policies, which is a loss of flexibility. "Much of the flexibility of open learning is lost when applying for full-time government student loans" (Open Learning Agency, 2000c, p. 1). The reason is that students must enrol at four fixed times during the year to qualify for funding, thereby losing the benefit of continuous enrolment. In sum, one unintended consequence of the current student loans policies is that low-income learners may be unintentionally excluded from online and web-based courses because they cannot receive funding for them.

Mismatch between the Online Forum and Enrolment Policies

According to the instructor, it was the intention of the OLA course developers that the online version of Psychology 101 would relieve the isolation of print learners by giving them access to other learners through an online discussion group. It is important to note that the online version of Psychology 101 was designed from the print package, and contains the same materials and assignments. In addition, online learners are expected to read the online lectures and participate in online discussions. However, this design caused several problems for learners.

First, there is the added complexity involved in navigating through multiple modes, including print, tele-course, telephone, mail and online components.

Yeah they could get their act together. I mean with the written stuff and the on-line stuff...I should either be in one format or it should be in the other. It shouldn't be a mixed media thing. It should be on-line or it should be written...So you had to go through all these realms you know reams and reams of material and then there was also stuff on-line...

This learner found that the use of three different methods was very confusing and became discouraged.

The confusion between the three different formats...not all lining up and I think that ...it's not fair to us and...there's so many other things going on and when you want to concentrate I mean you're you're sitting at the table doing the writing...you're at the TV watching the videos, you're on the computer you're just all over the house you know. So I would say that was the worst of it

She also referred to the aggravation of technical glitches “which set my teeth on edge”, and of not being able to open files from her instructor.

Building an online course from an existing print package is a common practice which may not work because the end product may have the appearance of an assembly of disconnected pieces (Tony Bates, personal communication, October 23, 1999). With Psychology 101, the online version was “bolted onto” the print version, and the unintended result was mismatched course components (Tony Bates, personal communication, October 23, 1999). One student interviewee perceived that different components written by

different people had been combined without sufficient thought given to how they fit together.

I also felt that the Psych... material that accompanied all of my course material constantly contradicted what she was telling us on-line. Like it had been written by two completely different people that had never met. Yeah and I lost marks because of that.....And it was like they added the on-line thing as kind of an after thought.

If the online course had been designed from scratch, the course would have had a more consistent “look and feel”, and it is likely that learners would have responded better.

But the most unexpected effect of this course design was that the online version was not successful in relieving the isolation of the distance learners. The reason has to do with the continuous enrolment model. With each online learner beginning the course at a different point in time, there was no cohort. This made group discussions almost impossible; instead, the tutor wrote a response to each posting, which generated a series of two-way dialogues. Although this activity may have relieved student isolation to some extent, this was not how the online discussion group was intended to work. In sum, these findings suggest that migrating courses like Psychology 101 onto an online format or onto the web can cause problems in the “look and feel” of the final product, resulting in a course design which some learners may find frustrating.

Conclusion

The underlying value implications which drive Psychology 101 is universal access for on-campus learners. The learners’ response to the course is mostly positive. The cost of the course to learners is inexpensive, and learners’ perception of the relevance of the

materials is mostly positive, even though frequent updating of the course materials may lead to shipping delays. The unintended consequences are interruptions in support services, which may be compounded into a battery of obstacles learners must negotiate to succeed. Another is the failure of the online discussion forum to “take off” because of a mismatch with the continuous enrolment policy. Unintended social consequences include learners not having the opportunity to take online courses because of inadequate student loan policies, and demoralization caused by a battery of obstacles, which may even lead to drop-out.

Microsoft Certified Systems Engineer (MCSE)

In this case study, Qayyum’s (1998) case report on the Microsoft Certified Systems Engineer (MCSE) course, the seventh of thirteen cases studies in the Response of Adult Learners project, will be used as a source document. Tables were reproduced with permission from the copyright holder, Adnan Qayyum (Appendix I). Tables without references to Qayyum (1999) were constructed by me and do not in appear Qayyum’s (1999) case report. Using categories from the adapted version of the adapted Messick’s (1989) framework (Figure 5), I re-coded and re-analyzed the same documents and interview transcripts used by Qayyum, who coded them with Bates’ (1995) ACTION framework. The reader should note that value implications, relevance and unintended consequence were not used as evaluation categories by Qayyum (1999). Using these categories yielded new findings which presented a different picture of the merit and worth of a course offered by a private sector institution.

Course Overview

MCSE is computer-based training (CBT) program which prepared learners to earn Microsoft Windows NT Certification; this credential qualified them for employment as Microsoft systems engineers. There was no pre-requisite, but there was a rigorous screening program for admission and a long waiting list. The course, which consisted of 6 modules, was held at the Burnaby Community Skills Centre, and had been running for two years when these data were collected. The course materials were Shalinsky's (1998) orientation manual, materials on the computer hard drive, required books for each of the 6 modules, CD-ROMs, instructional video, and online resources such as Microsoft's Technic-Library. Students were not permitted to copy this material, or take CD ROM's or instructional videos home because of licensing restrictions, and there was no remote access. Consequently, all of the technology-based materials could be accessed only from the Burnaby CSC's on-site computer labs.

There were two versions of the course: 1) CNS, which began in June, 1998 and ended in November, 1998, and 2) continuous enrolment, which students can begin at any time, and work on at their convenience. Both sections had the same curriculum and flexible pacing. The CNS students had their course fees paid by various government agencies such as Employment Insurance (EI) and the Worker's Compensation Board (WCB), and were required to be in the computer lab every afternoon on weekdays. The continuous enrolment students were paying themselves and set their own hours for coming into the lab. The eleven male and three female students in this research were mature learners, aged 24 to 55. All were employed, and all but one were CNS students.

Value Implications

Microsoft's ideology

Microsoft is a multinational corporation dedicated to maximizing profit, which by definition includes minimizing costs, and global market expansion. Microsoft's ideology emerged in the language of Shalinsky's (1998) course orientation manual, which refers to "Microsoft's marketing muscle" (p. 6) and "directories... weaving their tentacles into everything from e-mail systems to net management tools" (p. 6). "The Burnaby Skills Centre has formed a partnership with Microsoft to provide our students with high level training in a cost-effective manner" (p. 5), which "means significant savings" (p. 6). As a private sector institution, Microsoft has a philosophy and set of values quite different than those public sector post-secondary institutions.

Course objectives

In the orientation manual, there emerged an ideology of the benefits of using a variety of technologies in a self-study format. The primary objective of the MCSE course was to provide the "prerequisite skills" in computer networking and Windows NT for students to become certified as Microsoft systems engineers (Shalinsky, 1998, p. 2). In addition to developing "skills that are high 'in demand' for the computer industry", the course "provides training in job search and professional interpersonal skills" so that learners can "find employment" (p. 2). Additional course objectives include applying "learning to hands-on/real-life situations" and increasing your self-understanding as an adult learner".

A perusal of Shalinsky's (1998) orientation manual showed that Microsoft's statement of their teaching philosophy and theory of learning includes exaggerated and unsubstantiated promises. The Burnaby CSC provided an "enhanced learning environment" in which "learning involves a lot of self-study" which "comprises reading from books and computers", computer-based testing (CBTs), CD-Roms and "watching educational videos" (p. 4). The program administrator, who was taking a Master's program in instruction and performance technology at a distance from Boise State University, believes that this mix of methods customizes learning to the needs of the individual.

I think that what one of the big advantages of the way our program works is the mix of the different types of methods and combined together is what creates a rich learning environment... and some people it might be spending three-quarters of their time reading and only a quarter of the time on a CBT. And other people it might be 3/4s on CBTs and a quarter of the time, you know, reading. And that's where it's really customized to each learner. But the resources are there at their fingertips.

Shalinsky's (1998) manual contained some unsubstantiated and questionable claims such as "computer-based training ensures a consistently high level curriculum" (p. 6) and "computer-based training, otherwise referred to as CBT, is considered one of the best methods of training because of its flexibility" (p. 5). According to Shalinsky (1998), "By covering the material ahead of time, you will be able to maximize your learning potential" (p. 9). Other recommendations for "maximizing learning potential", which were a little more difficult to understand, were to "keep the noise level at a minimum, "dress appropriately" and "not eat or drink in the lab" (p. 4).

According to the program administrator, students worked through interactive computer-based materials at their own pace with two instructors available for support. The MCSE instructor's role was to facilitate, that is, to provide "guidance and support for student questions" (Qayyum, 1999). Instructors had an "open door" policy, where students ask "the instructors for help whenever they need it, which we've always tried to encourage", and to make the program administrators "accessible so that if they want to talk about the instructors that they can come to us". Students were also encouraged to collaborate with their peers to clarify the course material. Yet the "core" of Microsoft's teaching philosophy was learner self-reliance.

We tend not to give too much instructor-lead-training. We want to put the onus on the learner, because we are trying to prepare them for the real world scenario. In a real world scenario these are basically all vocational skills, they are not 100% academic, something to enhance their English speaking or social skills or anything like, these are skills that they would have to use in the work-place.

If you have the learning style and you're very self-disciplined and you're the type of person that can learn strictly from books and you don't need to collaborate with other people to learn, then it [self-study] will be very effective.

Certification was obtained by passing a battery of computer-based exams. After each module, students wrote practice exams, called Transcenders. The PEP exams went into more detail and the Assessment exams prepared students for the final exams, which were written at Sylvan testing centres and ExecuTrain. According to Shalinsky (1998), Microsoft was "an industry leader in certification on the forefront of testing methodology"

(p. 1). In sum, Microsoft’s vision of education was a self-study model with a broad range of technologies, frequent tests and a facilitating role for instructors. Interaction with instructors took place, but interaction with the technology and passing the tests, not human interaction, occupied centre stage.

Learner Response

Flexibility

Respondents stated they appreciated the flexibility of the pacing and scheduling. Sixty percent of respondents said they liked the delivery mode because it gave them flexibility, 33% were neutral and 7% disagreed (Table 20).

Table 20

Response to “I like this delivery mode because it gives me flexibility (e.g. time, place, location).” (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	7	0	0	5	33	5	33	4	27

Response to “If this course was not offered in this delivery mode, I would not be able to complete it.” (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
4	27	1	7	6	40	2	13	2	13

Source: Qayyum (1999)

As shown in the previous table, 34% disagreed, 40% were neutral and 26% agreed that they did not need this particular delivery mode to complete the program. Respondents agreed that this course requires taking more personal responsibility for completion than does a face-to-face course.

Materials

In the MCSE program, there were six modules which cover the basics of Microsoft systems, including workstation, servers, TCP/IP and networking. The course materials were in various formats, including CBT, CD-ROM and print. As shown in Table 21, 67% of students rated the materials as good, 7% as average, 20% as fair and 7% as poor.

Table 21

Response to "How do you rate the course materials?" (N=15)

Poor		Fair		Average		Good		Excellent	
No	%	No	%	No	%	No	%	No	%
1	7	3	20	1	7	10	67	0	0

Response to "The technology increases my motivation to work on the course." (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	4	27	3	20	5	33	3	20

Source: Qayyum (1999)

As shown in the previous table, 53% of respondents agreed that the technology increased their motivation to work on the course, 20% were neutral and 27% disagreed. Unlike the other cases in this research, technology was used because the subject matter required it (Qayyum, 1999). Even so, 27% said that the technology did not increase their motivation to learn. As for learner satisfaction with the software, 31% were satisfied, 31% were dissatisfied, and 38% were neutral (Table 22).

Table 22

Response to "I am not satisfied with the software used in this course." (N=13)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	4	31	5	38	4	31	0	0

Response to "I can learn better using print materials than by working on a computer." (N=14)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
2	13	2	13	8	53	1	7	1	7

Source: Qayyum (1999)

Finally, in their ratings of the relative value of print materials, 14% agreed they could learn better with print, 53% were neutral and 26% disagreed.

Interaction

In response to the question "I am able to interact with my instructor as much as I want", 47% agreed, 27% were neutral and 27% disagreed (Table 23).

Table 23

Response to "In this course, I am able to interact (communicate and exchange ideas) with the instructor as much as I want" (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	4	27	4	27	7	47	0	0

Response to "In this course, I am able to interact (communicate and exchange ideas) with the other students as much as I want." (N=15)

1		2		3		4		5	
No	%	No	%	No	%	No	%	No	%
1	7	0	0	0	0	8	53	6	40

Source: Qayyum (1999)

As shown in the previous table, ninety-three percent were satisfied with the opportunities for student interaction, and only one person was dissatisfied.

Support services

The distribution of responses for satisfaction with the support services was bimodal. Forty percent felt they were satisfactory, 13% were neutral, and 40% felt they were unsatisfactory (Table 24).

Table 24

Response to "Support services for this course are unsatisfactory." (N=14)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
2	13	4	27	2	13	6	40	0	0

Source: Qayyum (1999)

When asked to make recommendations for improving support services, access was the most frequently mentioned issue (Table 25).

Table 25

Support Services—Recommended Improvements

	<u>N=17</u>
<u>Access:</u>	
More Lab Access	4
Loan out materials	1
Remote Access (i.e. from home)	2
More Computers in Labs	1
<u>Resources:</u>	
Library	1
Wider use of existing resources (e.g. job search)	1

Instructors:

More Instructor Contact	1
Certified Trainers	2
Regular review of student progress	1
Schedule lab assignments with instructors	1

Content:

More hands on/ practical training	1
-----------------------------------	---

Technical:

Address Technical Problems	1
----------------------------	---

Source: Qayyum (1999)

The second most frequently mentioned issue was the instructor, and respondents recommended more instructor contact, scheduled lab assignments with instructors, regular review of student progress and better certification of trainers.

Relevance and Cost/Benefit**Relevance**

The MCSE course is relevant to these learners, who are unemployed, because it certifies them for future employment. Sixty-seven percent felt the materials were relevant to their personal or professional needs, 20% were neutral and 13% disagreed (Table 26).

Table 26

Response to "The course materials are relevant to my personal or professional needs."

(N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	2	13	3	20	6	40	4	27

Response to "Using technology in this course helps me to learn more relevant information." (N=13)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	2	15	3	23	6	46	2	15

Response to "Using technology in this course helps me to learn with greater depth of understanding." (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
0	0	3	20	2	13	8	53	2	13

Response to "In this course, the interaction with the instructor is relevant to my learning." (N=14)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	7	4	28	2	14	6	43	1	7

Response to “In this course, the interaction with the other students is relevant to my learning”. (N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
1	7	0	0	2	13	9	60	20	3

As shown in the previous table, sixty-seven percent of respondents felt that the technology helped them learn more relevant information, 20% were neutral and 13% disagreed. Sixty-eight percent agreed that the technology helped them learn with greater understanding, 13% were neutral and 20% disagreed. Fifty percent saw the interaction with the instructor as relevant to their learning, 14% were neutral and 34% disagreed. Finally, 80% of respondents disagreed that interaction with other students was relevant to their learning. In sum, these findings show that learner response to the MCSE course was mixed.

Cost

According to the respondents, the average amount spent on tuition was \$10,200 (SD=\$1476, N=10). Human Resources Canada paid the tuition fees for CNS students, while the continuous enrolment students paid for their own tuition, books and tests. The average amount spent on books was \$354 (SD=19.15, N=6). Each of the four operating system exams and two elective exams required for certification costs \$100 US.

Thirty-four percent of learners felt that the course was not worth the money, 41% were neutral and 25% felt it was worth the money (Table 27).

Table 27

Response to "This course is not worth the money it costs." (N=12)

1		2		3		4		5	
No	%	No	%	No	%	No	%	No	%
1	8	2	17	5	41	2	17	2	17

Response to "I would not take another course using this delivery mode."

(N=15)

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
No.	%	No.	%	No.	%	No.	%	No.	%
4	27	1	7	6	37	2	13	2	13

Source: Qayyum (1999)

As shown in the previous table, 34% said they would, 40% were neutral, and 26% said they would not.

Benefits

One benefit of the course, mentioned by 10 out of 13 learners, is self-pacing (Table 28).

Table 28

Benefits and Drawbacks of the Delivery Mode

<u>Benefits</u>	<u>N=10</u>
Self-pacing	10
<u>Drawbacks</u>	<u>N=9</u>
Lack of instruction	3
Student motivation	1
No hands-on experience	1
Lacks structure	1
Routine	1
Lack of support	1
Technical problems	1

Source: Qayyum (1999)

For CNS learners, the course has a quick, flexible time frame, with a maximum of 26 weeks. This is an important benefit to learners because it minimizes the transition time between their former jobs and their new careers as systems engineers. There is no time frame given for continuous enrolment learners, although one interviewee had been in the course for a year and a half. The most frequently mentioned drawbacks were that it was not instructor led, that there was insufficient structure, computer glitches and errors in the materials. Even with a weekly lecture, student interviewees said they wanted more

instructor-led sessions, including “an overview of a module from an instructor before starting it”, or “a instructor that lead the class in some exercises with what we are learning i.e. administration, security, networking” (Qayyum, 1999). In response to the question about problems, there were 17 comments covering a wide range of issues (Table 29).

Table 29

<u>Student Problems with the Course</u>	<u>N= 17</u>
Computer glitches, down time etc.	3
Errors in materials	3
Lack of formal instruction	2
Instructors not qualified or knowledgeable	2
Long waiting period/paperwork processing	2
Lack of hardware	1
Lack of hands-on training	1
Access to materials is place-bound	1
Independent study format	1
Preferential treatment to fee paying students	1

As shown in Table 30, the most frequently mentioned problems were computer glitches and down time, errors in materials, lack of formal instruction and instructors not being qualified or knowledgeable and long waiting period and paperwork processing.

Microsoft's Costs

In the OLT study, data on the costs to Microsoft of delivering the MCSE course were not collected. The only data on costs is incidental information about money moving back and forth between Microsoft, the Burnaby CSC, and the test developers and test delivery agencies which emerges from the interview data. This data is insufficient to do any meaningful analysis or draw any conclusions.

Unintended Consequences

Relevance--Mismatch with the needs of the market

The orientation manual promises that "computer professionals who become Microsoft Certified are recognized as experts and are sought after industry wide" (Shalinsky, 1998, p. 1). Yet, according to the program coordinator, it is more difficult to get a job as a Microsoft systems engineer than it used to be. Although the course is geared to training for jobs in the computer industry, the market is shifting in favor of "people skills", which does not fit with technology-based self-study model. New graduates who tend to get jobs tend to have "soft skills", that is interpersonal and communications skills. But the ideology underlying the course is that students learn best in a technology-based self-study format, and MCSE students are not provided with any systematic program of instruction, other than an orientation and interview activity, to teach them these "soft skills".

The hypothesis that there is a mismatch between the needs of the job market and the skills provided by the MCSE course needs to be confirmed by an investigation of employment rates of new graduates.

I want them to be clear, and to be honest the starting salary has gone down and it doesn't surprise me because there's a lot of MCSE [engineers] out there and we are getting into a bit of a recession. So the income has gone down a little bit.

If it is true that the market for Microsoft Systems Engineers is becoming saturated, then there is a mismatch between the course design, a technology-based self-study model and the needs of the market. To meet employers' needs for "soft" skills will require more than a few "face-to-face add-on's" provided by the program administrator. Moreover, using the "boot camp" approach to pedagogy favored by the American military is hardly an appropriate way of modeling these "soft skills"! A better method is to provide in a systematic modular face-to-face instructional program of activities to teach effective communications or teamwork skills. One such program which has been receiving a lot of recognition in BC is the BCIT model, which comes complete with a field-tested guide for faculty (Hartley & Robson, 1998a, b).

Lack of access and flexibility

The kinds of technologies used in this course are often used elsewhere to provide learners with the means for studying from home. Some MCSE students buy a computer, set up their own local area networks at home and experiment, which is considered "value-added" by the program administrator. In fact, however, Human Resources Canada and Microsoft policies and regulations intervene, making the MCSE course "place-bound". First, Human Resources Canada requires CNS learners to be at the Burnaby Community Skills Centre during business hours five days a week. Secondly, learners are not permitted to take course materials home because of Microsoft's licensing agreements. For the same reason, there is no remote access. As a result, the potential of the technologies to provide

remote access is not realized. Policies and regulations intervene in the course implementation, and are required to travel to the Burnaby Community Skills Centre and to the testing centres, where they work through the materials.

Adding in face-to-face components

Instructor lectures were added on to the course after learners in a previous section demanded it. The orientation manual stresses a model of instruction in which a range of technologies in a self-study format is optimum. In fact, instructors were “added on” when learners complained that the self-study format was not working. According to Qayyum (1999) “students in a previous offering of the program had asked for more formal instruction from the instructor”. So in this offering the instructors gave weekly lectures which CNS and MCSE (continuous enrolment) students had the option of attending.

So in those courses ...should be learner-centered, not instructor-centered or you know institution-centered. And so towards that end...we offer here the computer based training coupled or complemented with the instructor-lead-training which actually...work very well together.

Over the two years that the MCSE course has been running, the program administrator has added several face-to-face components to meet the needs of learners. He visited industry people at their work sites, assessed their employment needs and reported back to the group at face-to-face orientation sessions. Based on his findings, he redesigned the lectures and activities to include a new focus on “soft skills”, which are in demand by the industry.

One is it helps me keep in touch with what industry is saying... Industry is now saying that MCSE on its own is no longer that marketable and that they want people with experience. So now I go back to our current students and I can tell them that, so it's not just us preaching. They are hearing it from the people who are doing the hiring.

I discuss a lot about soft skills as well as hard skills. Soft skills meaning communication skills, interpersonal skills, attitude. The soft skills are extremely important if you want to go anywhere in this career. So I...explain what industry is demanding. OK, experience is becoming more and more in demand.

To provide learners with these "soft skills", the instructors give face-to-face motivational lectures, and require learners to do two face-to-face job interviews with managers as a practice job search at the beginning of the program. "They have to go out into industry, speak with and find out some key questions." Students are also required to do a practicum during the course so they can apply what they've learned.

Whereas what we were finding when we tried to get them to do job search in the middle of the program they would just keep blowing it off and then at the end of the program they were stuck without a job and they're blaming everybody else except for realizing that there is a lot of initiative that they need to put in.

In addition, the instructor meets one-on-one with learners to discuss their interviews and the quality of their research. This face-to-face monitoring allows the instructors to identify and confront learners with "negative attitudes".

And I might send them back out again. And it depends on their attitude, you know, if I ask them how their drive in was in the day and they start complaining about a lot of things, then, you know, then I'm already, my antenna's up for certain type of negative attitude. We can screen our students, and we only want the good students in this program... [We've] become much more selective.

These negative attitudes are shaped using strategies based on the methods used by American military academies.

Basic training...military academies where like in in the Armed Forces in the United States...they really give them a rough time and try to even burn up their sheds, or underpants or give them a kick and just yell and scream at them right. So it's like a boot-camp initially, so that's what we do to these people come in here for project-based-training. So we're trying to in a sense change their attitude right from the beginning and shape their behaviour so that's a lot of what I've been doing in the front end.

Short product cycles and updating of the curriculum and tests

Because specifications in the industry are constantly changing, there is a need for almost constant updating of the course materials. Instructors struggle to keep up with these changes by updating the CBTs and studying the material themselves along with the students.

I have been in the industry for three years. I was a student here...I took the CNE, and the CNE was very popular...and suddenly the market took a down turn and Microsoft...took the industry by storm..it's clear...and Novell just went down and...the industry actually tilted so much...I had to do Microsoft...so I had to re-train myself. Tomorrow it could be different,... there is a lot of uncertainty in the industry. Tomorrow if you ELINX comes up we will have to go and do ELINX. Microsoft is ... always developing proto-type stuff and they send you updates...Everything is a beta, so the amount of information they [students] get is constantly changing and evolving. ... We have to constantly adapt to the various changes that are being made by the vendors, Microsoft Windows, as well as Microsoft itself. So we have to be very, very open-minded and I sometimes I hate to use the word nebulous... The curriculum is very, very dynamic.

Presumably, materials designers and test developers are also struggling to respond to short product cycles. The orientation manual admits to the presence of errors in the tests, and gives students advice on how to deal with them. Both a student and an instructor interviewee mentioned that there were sometimes wrong answers in the answer keys, and the instructor described the tests as "nebulous".

The exams are open. The principles and all the things are quite the same, but the exams are very open, because **the exams are not tailored...according to the material.** It's how Microsoft views it and [what] Microsoft wants (emphasis mine).

The short product cycles may explain the presence of errors in the testing materials, and the “nebulous” nature of the course materials. In sum, because short product cycles leave the developers and instructors struggling to stay current, the curriculum is “nebulous” and the tests sometimes contain incorrect answers.

Eliminating qualified instructors

One would think that the “nebulous” and “dynamic” nature of the curriculum, and the presence of testing errors would make it important to have an on-site instructor who is an expert in the material, and can assist the students with these problems. One imagines that resolving these problems is important for sustaining the motivation of learners, especially those in continuous enrolment. The program administrator acknowledges that students will need to ask someone, such as an instructor, for clarification when they are struggling, yet the core of the course is self-study.

We [instructors] are only pointers in many sense and we are only here to disseminate certain different, certain concepts that the adults or students might find difficult to understand and learn and thereafter they are pretty much on their own for most part of the course.

As demonstrated in Tables 4, 80% who agreed that discussion with their peers was relevant to their learning, compared to 50% of respondents who agreed that discussion with the students was relevant to their learning (Table 15).

I'd like to say the only thing that worked for me was interaction with the other students. And other than that interaction with the teacher didn't help at all... Like if I needed to turn to somebody, that's [the students were] the only people I could turn to. I couldn't turn to the teacher because I kept getting the wrong answers, if that helps.

One learner said the instructor was working on the lessons just ahead of the students, while another said that instructors lacked certification.

There's basically, there is no instruction, the teacher isn't certified, he's not a Microsoft certified trainer. But I was really frustrated with the whole program.... It doesn't help when someone doesn't know more than you, and they're the instructor...

According to another student,

[You] didn't find out it was wrong until [you] write your test, fail it ... One of the things that really upset me, actually, was on two of the questions transenders were wrong and every time I went to the instructor the instructor said they were right.

Yeah, going through TechNet and all that material I found out it was wrong to begin with... so that was quite frustrating.

This may explain why learners depend on each other for support with their learning. This also has an unintended consequence, which has to do with exam preparation. For those who engage in discussion groups on the web, answers to the exam questions are sometimes shared, a practice not encouraged by Microsoft.

OK, well, there's the Microsoft websites and there's what they call brain dumps which we don't like to encourage but the students do use them, and that's where students who've written exams, people who've written exams and stuff they share concepts and they discuss them, and this is all through the Internet. The students often access that on their own.

There is no information on how widespread this practice of "student collaboration" is, but it raises questions about the validity of the MCSE tests, and the validity of the MCSE certification.

In a previous course offering, Microsoft had attempted to eliminate instructors, but student complaints forced them to bring the instructors back as "facilitators" (Qayyum, 1999). The role of the instructor, then, is downgraded to "expert learners" who are themselves working through the course, according to one of our interviewees. Despite learners' demand for qualified instructors, Microsoft's plan for the future is to eliminate instructors by moving to web- and video-based courses.

We are testing in testing beta stage right now and we will be going distance learning because **the cost of education is getting cheaper** and it will not be possible...to have a lot of instructors at individual sites and...so that the delivery could be could be effective to a larger number of or group of people, students. So we are going to go on-line very shortly (emphasis mine).

This change will allow Microsoft to achieve a key objective, which is to cut costs, because students will take over the instructors' work without being paid for this work.

Continuous Enrolment

The needs of the MCSE (continuous enrolment) learners are especially hard to meet because each one is at a different place in the curriculum at any given time. To make the lectures relevant for these learners, the instructor has to design each lecture so that it includes concepts from all six modules.

We have to make sure that we always teach all six modules and we do it on a continuous basis and that there's enough during the week of each one so that everybody feels that their needs are met. Like I said, to make sure within one week's time we have enough of the different topics being taught.

For continuous enrolment learners, obtaining help from other students who are in the classroom at the same time is not always a viable option.

Because everybody's on different schedules so...if one guy's doing Essentials and you're doing Server you can't really go to him for any help...Self pacing, you tend to get a little bit lazy. Especially when the material is as dry as it is, you don't have a lot of drive. With the classroom, everybody has a set time, a set goal. Like for me, I sort of need someone with a gun to my head, so (laughter) No, but you know what I mean. ...basically, motivation (Qayyum, 1999, p. 5)

This learner, who was off work because of a work-related disability, took one year and a half to work through five of six modules. He was not happy with continuous enrollment, and wanted more in-class training and set goals instead of self-pacing.

The only student interviewee says he was not motivated “because of my laziness...I should have put more into it.” When asked if he had tried to set deadlines to pace his learning, he replied, “Well, you’d try, but if you pushed too hard, you’d fail (laughter) Simple as that.I mean, you’re lucky, you get to rewrite again, but it’s quite expensive...\$150 per test” (Qayyum, 1999, p. 7)

Although an interview with one learner does not allow for generalizations, it does show that the continuous enrollment model does not work for everyone.

Conclusion

In conclusion, Microsoft’s belief in technology and mandate to cut costs leads to a course implementation to which the response of learners to the MCSE course is lukewarm. In an industry with short product cycles, materials require such frequent updating that it is difficult to keep up with change. There are several unintended consequences in this course. First, licensing restrictions prevent learners from using the technology to work from home. Face-to-face components were reluctantly added in to satisfy learners, who needed help with the errors in the course materials. Instructors are perceived as poorly qualified, so students rely on each other for assistance, and sometimes even for the answers to exams. Learners take on the instructors’ role, even though they are not being paid for this work. This situation suits Microsoft, who tried unsuccessfully to eliminate instructors before, and will try again.

Because a traditional evaluation framework, the ACTION model, was used to structure Qayyum’s evaluation report, data was not collected on several topics which should be collected to provide a comprehensive assessment of merit or worth. First, the time span for learners is important background information. How long can learners take to

finish the course? Questions are also raised about the time taken by continuous enrollment learners to complete the course and their completion rates, which will require more data to address. The data in this study suggests that the continuous enrollment model does not work for everyone.

Second, information should be provided on the frequency with which the course materials and tests are updated, and what kind of quality control is done on the student tests. Thirdly, data on salaries and employment rates of new graduates would address the criteria of relevance. A better estimate of costs to learners and the institution is needed. Only the cost items for tuition, books, exams parking, travel, and Internet rates were analyzed because of low response rates and missing data on the other items about cost (e.g. software, internet connections). Does Microsoft provide any scholarships or other types of funding to help learners cover the \$10,000 enrollment fee? How does this figure compare to tuition rates at traditional post-secondary institutions? Is Microsoft implying that face-to-face instruction is one of the worst methods of training? As the Liberal government cuts back on funding for public post-secondary education in BC drastically to further the interests of private sector educators (Killian, 2002), it is sobering to reflect on a future in which there are fewer public-sector alternatives to programs like MCSE.

CHAPTER 6: Discussion

The Contribution of this Research

Let me summarize so far. In Chapter I, I argued that the increasing complexity of globally delivered distance/distributed courses has created a need for a comprehensive model of evaluation in distance education which includes unintended consequences as an explicit evaluation criterion. In Chapter 2, I supported this argument with the literature of program evaluation, quality assurance, educational technology and evaluation in distance education. I also showed that 20 years of evaluation models in distance education have generally not included unintended consequences as a criterion. To provide a comprehensive picture of the implementation systems of distance courses, an evaluation framework based on the rich literature of assessment and program evaluation is needed.

In Chapter 3, I introduced Messick's (1989) framework on validity, which is a comprehensive approach to assessing merit and worth in the area of assessment. I then adapted this framework so that it could be applied to authentic evaluation data from three BC post-secondary courses. Not only does the framework include elements common to evaluation models for distance education, such as outcomes, relevance and costs, but it also provides an effective response to Rumble's (1981) call to investigate unintended consequences. My purpose was to demonstrate how the adapted framework performs in a "test-drive" situation. In Chapter 4, I discussed the methodology used to apply the adapted framework to authentic data in a distance education context. In Chapter 5, I presented the findings which emerged from this application. As these findings demonstrate, the adapted Messick's (1989) framework provides us with a comprehensive picture of merit and worth because it is based on both facts and values. My findings demonstrate the benefits for evaluators of distance programs of

articulating the values underlying distance courses and of analyzing unintended instructional and social consequences. Finally, in this chapter, I will review some key theoretical ideas, such as the consequential basis and the tension between facts and values, and recap the specific ways in which these concepts illuminate the workings of the implementation systems in the novel context of three distance/distributed instructional programs.

What Have We Learned from Applying the Adapted Messick's Framework?

The Novel Contribution of this Research

This research provides a novel contribution in three important respects. First, the adapted Messick's (1989) framework responds to Rumble's (1981) call to investigate unintended consequences. In 20 years of evaluation literature in distance education, there has been implicit support for the notion of unintended consequences, but the term "unintended consequences" is avoided. Tenner (1996) shows that unintended consequences are a common feature of technology in general, and others, such as Fabos and Young (1999) and Fahy (1998), have provided specific examples of the unintended consequences of educational technology. In addition, the adapted Messick's (1989) framework extends our understanding of unintended consequences by defining it as "the unintended side effects of *legitimate* use" (Messick, 1998, p. 40) and showing the overlap with value implications.

Secondly, the framework reminds evaluators of distance courses that "values" are central to the work of evaluation and offers an approach to values which reflects the diversity of contemporary society—value pluralism. Thirdly, the underlying assumption of the tension between facts and values draws our attention as evaluators to the dynamics, both intended and unintended, which play themselves out in the implementation system.

Because they do not include these three elements, traditional evaluation frameworks in distance education are incomplete, and applying them to evaluation data is likely to provide an incomplete assessment of merit and worth.

Borrowing from Assessment and Evaluation

In this research, I've adapted a framework from the field of testing and assessment, and applied it to a new area—program evaluation. In Chapter 3, I pointed out that these fields used to be the same and that theorists such as Cronbach have worked in both areas. It was only later on that the two fields emerged as distinct. Unintended consequences and value implications are precisely the two “corners” of assessment which I wish to borrow for this research. Because “Messick has been down this road before” (Zumbo, personal communication, June 14, 2002), I have borrowed his ideas, as well as ideas from the debate among Moss (1998), Reckase (1998b), Markus (1998) and Messick (1998).

Bringing Issues from the Background to the Foreground

Messick's (1989) framework is founded on Singer's (1959) view of rationality, where two different systems of inquiry confront one another in order to bring forward and make visible their underlying epistemological and value assumptions (Messick, 1989). The adapted Messick's (1989) framework illuminates “taken-for-granted assumptions, knowledge and practices” (Moss, 1998a, p. 65), which would otherwise be “disqualified” “against the claims of a unitary body of theory” (Foucault, 1980, p. 82). According to Moss (1998a),

The issue is not about what's *possible* within different perspectives (as Bernstein (1979) notes); it's about what's emphasized, illuminated or made more likely; what's relegated to the background as trivial or impractical; and what impact this

prevailing emphasis has on the actual practices of social scientists and the communities they study and serve (Moss, 1998a, p. 56)... This emphasis on the importance of an outside perspective to illuminate what is taken for granted (as natural, normal, the 'way things are done') and thereby to provoke critical self-reflection is a theme that resonates across multiple philosophies of social science... This insight is one of the most profoundly important insights that Messick has brought to the tradition of educational and psychological measurement (p. 62).

By highlighting findings which emerged from the use of the adapted Messick's (1989) framework but which were in the background with the ACTION model, I do not intend to criticize the ACTION model. Moreover, the reader should note that the use of traditional distance education evaluation models does not mean that unintended consequences will never surface. (In fact, they did surface in my first analysis of the data using the ACTION model.) My purpose is not to run a horse race between two models, but to bring forward findings which remained in the background in the ACTION model, but which came to the foreground with the adapted Messick's (1989) framework. "It is precisely such mutual confrontation of theoretical systems, *especially in attempting to account for the same data*, that opens their underlying scientific and value assumptions to public scrutiny and critique" (Messick, 1989, p. 61-62) [italics mine].

Because their underlying epistemological and value assumptions are different, the two frameworks will likely produce different findings when applied to the same datasets. When I applied the adapted Messick's (1989) framework, in all three cases new findings corresponding to costs, relevance, value implications and unintended consequences emerged, were highlighted or "came together" in ways which provided a comprehensive

assessment of merit and worth. I will now summarize the findings which the adapted Messick's (1989) framework brought to the foreground in this research.

Summary of the Findings

Unintended consequences

The adapted Messick's (1989) framework brought forward several important issues around unintended consequences which had remained hidden, dispersed or in the background in my previous analysis. In all three cases, these effects were not trivial, and did not result from misapplications of the courses, but from implementing them as they were supposed to be implemented, as described in the course objectives. This analysis shows that Messick's framework brings forward additional aspects of merit and worth, specifically unintended consequences, which remained in the background with the ACTION model. Again, I am discussing the contributions of Messick's model, not running a horse race which would deflect attention from the contributions made by Messick's model.

With Modern Languages 400, they include students' demand for more flexibility than the course provided, bimodal satisfaction ratings from students, a mismatch between the technology and the subject matter and the loss of benefits from face-to-face classes. Unintended social consequences include the challenge to traditional academic structures like classroom credit-hours and the changing role of the instructor. With Psychology 101, they include interruptions in support services, negative effects on the motivation of distance learners, difficulties in contacting the instructor, and receiving the wrong package of materials. One unintended social consequence in this course is that low-income learners are unable to obtain student loans or afford to buy a computer to take online courses. With

MCSE data, the examples of unintended consequences include lack of flexibility because of licensing arrangements, the reluctant addition of face-to-face components in response to learner complaints, learner dissatisfaction with the errors in the tests and the elimination of qualified instructors. These unintended consequences appear to have had compounding effects, which I will return to later in my discussion of the tension between facts and values.

Value implications

When the adapted Messick's (1989) framework was applied to the evaluation data of the three case studies in this research, specific instances of value implications were brought to the foreground where they could be carefully scrutinized. As this research has demonstrated, values permeate meaning and consequences in "subtle and insidious ways" (Messick, 1989, p. 59). To understand how this applies to distance education, I will give examples from the three cases in this research based on Messick's (1989) three areas of analysis: construct labels, ideology and theory.

Construct labels

Evaluators of distance courses need to be aware that the discourse in distance and distributed learning is often characterized by bias, defined by Messick (1989) as "the intrusion of ordinarily tacit extrascientific motives or beliefs into the fulfilment of scientific purposes" (p. 59). The labels used to talk about distance courses often carry what Messick (1989) refers to as the "evaluative overtones" (p. 59) of broader schema, theories and ideologies, which then become assimilated into the language used to discuss technology. In *Modern Languages 400*, for example, the claim that learners can make a "faster progression" was unsubstantiated. Although the term "traditional knowledge transmission model" is often used in the literature to

describe face-to-face delivery (and the connotation is negative), the bimodal survey results in this course show that several students prefer face-to-face instruction. The OLA's web pages were characterized by the rhetoric of equity, access and service, which contrasts with the experiences of Psychology 101 learners. In MCSE, the theory and terminology of a "learner-centeredness" appeared frequently in the course manual, but learners who did not like the course were labelled with "a bad attitude". This kind of discourse, which constitutes a "rhetoric of intent" (Scriven, 1972), tends to emerge in the course objectives and interviews with course developers. Against this rhetoric, an investigation of unintended consequences can provide a considerable contrast. Bringing this contrast between "the image" and "the reality" to the foreground is the reason that it is important to include value implications in the evaluation of distance and distributed courses.

Ideology

Messick (1989) believes that many important values "are likely to remain relative to their community of stakeholders or believers" (Messick, 1998, p. 38). This is especially relevant for distance and distributed learning, where different stakeholders have different ideologies about the use of technology. One ideology surrounding technology is service to learners, and different technologies provide different kinds of service. Psychology 101, for example, reflects the underlying value of universal access to education. The ideology or mandate of the Open Learning Agency is to bring education to all British Columbians, including those who may not fit into traditional university structures because of low income, work or child-care commitments, prison incarceration or disabilities. This analysis of values sets the context, making the emergence of an obstacle course for learners all the more surprising.

In Modern Languages 400, the face-to-face component was reduced to provide schedule flexibility for on-campus learners, so that they could fit in their other courses and part-time jobs. According to Oppenheimer (1997), the beginning of the shift away from teacher-fronted classrooms was Apple's realization that computers could be most effective for learning only if they coupled with new pedagogical values such as inquiry-based learning. Yet Modern Languages 400 was a reading skills course and there were mismatches between the values underlying the technology, the value of flexibility for learners and the values underlying a traditional reading skills course.

MCSE is based on an ideology of service to learners, but also on an unrealistic expectations of the pedagogical role of technology. Underlying this course is the value of cost relative to benefit, with the balance tilted in favour of the former. Although Ungerleider and Burns (2002, April-May) maintain that the high cost of computer technology can only be justified if it is associated with improved teaching and learning, this value did not seem to be shared by Microsoft. Yet on the value of relevance to the needs of the global economy in the information age (Reigeluth, 1999), MCSE scores highly because its vocational job training program is frequently updated to remain relevant to a changing job market, their approach to teaching "soft skills" notwithstanding.

Theory

Distance courses reflect both learning theories and theories about the use of technology. Learning theories such as distributed cognition (Salomon, 1993), constructivism (Duffy and Jonassen, 1991) and discovery learning (De Jong & van Joolingen, 1998) underlay the design of distance/distributed courses. For example, using multimedia for a reading skills course creates a tension between the constructivist values

underlying multimedia and the traditional pedagogy of a reading skills approach. Learning a foreign language can be made easier by the addition of pictures or key visuals which give clues to the meaning of the language. Modern Languages 400 could have been designed upon learning across the curriculum principles, e.g. sheltered and adjunct models, so that pictures provide clues to meaning, and listening and speaking skills reinforce reading skills and language elements. Instead, using CD-Rom to teaching text-based reading skills creates conflicts between the pedagogical values of sound and image for which the technology is noted, the traditional values underlying a reading skills approach and the cost of course development.

Markus (1998) maintains that “the tension internal to the theory becomes a tension internal to the process of validation” (p. 77). This means that evaluators need to keep a critical eye open to conflicts among opposing ideologies and theories is needed in their evaluations of distance courses. In a later version of Psychology 100, e-mail was adopted to relieve the isolation of distance learners, but the continuous enrolment policy interfered with the effectiveness of the proposed “learning communities”. Multimedia, though “place-bound”, was used in Modern Languages 400 because of a belief in the notion that self-directed learning, immediate checking of answers and happy faces can improve learning and motivation. The lesson is that evaluators need to be open to the presence of these theories, but also to reflect upon and patiently wait for the “emergence” of other elements which may or may not be effective applications.

The multiple value bases of distance/distributed education

The findings of this research show that distance education courses do not rest on a singular value basis, but on multiple value bases. These findings support Moss's (1998a)

view that Messick's (1989) conception of validity is an "ongoing critical reflection about our interpretations and theories in light of challenges from alternative perspectives" (p. 55). In distance/distributed education, multiple and diverse sets of values underlay the theories and use of technology, learning theories, course objectives, choices made by course designers and motives for developing post-secondary courses. Bringing OLA's values to the foreground provides a poignant contrast with Microsoft's values, and this contrast provides us with a better understanding of distance courses than if we had never looked at values.

We have already seen that different underlying values may come into conflict within the same course. With Modern Languages 400, for example, we observed a conflict between the traditional value of maintaining a traditional academic structure of weekly class meetings and the new values of flexibility and learner autonomy. We also observed instances where some values do not "mesh" very well with other values underlying the course design, pedagogical approach and subject matter. Modern Languages 400 data brought to the foreground the trade-off between 1) the richness and depth of multimedia and 2) the high cost of multimedia development. This finding supports Markus's (1998) point that "the tension can be analysed within bases as well as between" (p. 79).

The tension between facts and values

This research shows that the tensions underlying distance and distributed courses take the form of both facts (e.g. choice and use of technology), and values (that is, construct labels, theory and ideology). These tensions are both internal to each of the bases and also cut across the bases, the clearest evidence being "that consequences are facts" (Markus,

1998, p. 79). As Messick (1989) recommends, the tension between facts and values needs to be carefully negotiated in practice.

In distance education courses, the “facts” include the technologies, subject matter, and use of technologies, while the “values” include organizational goals, learning theories, and ideologies. As demonstrated by these findings, the tension between facts and values is a “metaphor” for the workings of the implementation system, which need to be analyzed on a case-by-case basis. Taking this perspective enables us to see how conflicts among facts and values play themselves out, how these tensions pull the course in different directions, and whether there are any adjustments being made, or which need to be made, in the implementation system.

In Modern Languages 400, for example, the underlying value implications work at cross-purposes to “pull” the course implementation in unexpected directions. Learner interaction is traded off for flexibility and the students’ need for more flexibility comes into conflict with the instructors’ need to maintain a traditional classroom structure. The need to manage “costs” conflicts with the need to create pedagogical richness and depth. The findings demonstrate that a comprehensive assessment of merit and worth of a multimedia course must include information on the cost of course development. These findings show the dynamic nature of value, the trade-offs among multiple values and unexpected interactions across the four facets of value.

With Psychology 101, the underlying assumption of the tension between facts and values lead to the emergence of a chain of unintended consequences which cuts across the four facets. Specifically, various interruptions in support services had negative effects on learners’ motivation and self-confidence, which is crucial for persistence. These learners

came into a distance course with an uncertain self-concept, and one important question is the extent to which this battery of obstacles generated feelings of frustration and self-blame. Another question is to what extent these feelings impacted on completion rates, for which unfortunately, no information was available.

With MCSE, Microsoft's goal to cut costs conflicts with their goal to meet the needs of learners. The unsubstantiated claims in the course orientation manual reflect an uncritical belief in technology, and a superficial, confused and contradictory treatment of the dynamics of human learning. The course objective of increasing the self-understanding of adult learners is understood to mean turning out students who respond well to a technology-based self-study model. One senses that traditional educational values are being overridden by a blind faith in technology and an imperative to cut costs, which is behind recurring moves to replace the instructors (whom learners have said they want) with technology.

Microsoft's mandate to increase profits and cut costs lead to several unintended consequences in terms of diminished access and flexibility, quality of the curriculum and the tests, student complaints over the absence of qualified instructors, and student collaboration on tests. Because of Microsoft's short product cycles, test materials require such frequent updating that they contain errors. Because students are not completely satisfied with instructors, they rely on each other for assistance and take on the instructors' role without being paid. This situation suits Microsoft, who tried unsuccessfully to eliminate instructors before, and will try again. In this course, there is a mismatch between the potential and actual benefits of technology, between the values and ideology and the reluctant addition of instructors to meet learner needs, and between the needs of learners

and the recurring moves to eliminate instructors. The underlying values of the course come into conflict with the rapid speed of updating of materials, the errors in the tests and the needs of learners.

How do Evaluators Deal with Multiple Values?

In contrast to Markus (1998), Moss (1998a) argues Messick (1989) stops short of calling for a singular validity. Because validity rests on the tension between facts and values, validation practice does not require a singular, completed synthesis and the tension between facts and values is not, and does not need to be, resolved (Moss, 1998a). The perspective that these tensions, including tensions within the bases, do not need to be resolved also applies to distance programs. There is no need for evaluators to camouflage, dismiss or “force” a kind of internal consistency on programs which reflect these kinds of underlying tensions. Instead, evaluators can be sensitive enough to recognize them, allow them to emerge and document their emergence. At a minimum, evaluators should acknowledge the merits of conflicting value perspectives, be ready to justify the value position inherent in their evaluation and be willing to admit that other value positions can also be justified (Markus, 1998).

Summary

These findings support Messick’s (1998) notion that “if consequences are not part of the validation process, many sources of validity will remain unexposed” (p. 43). Because traditional evaluation models in distance education are based only on the evidential, and not on the consequential basis, evaluation studies based on traditional models will likely result in incomplete results, and in results which emphasize positive findings. Having demonstrated how Messick’s (1989) framework can be used as a model of program

evaluation, I would now like to go to the program evaluation literature to flesh out the use of this framework.

Insights from Social Advocacy Approaches and Responsive Evaluation

Social Advocacy and Program Evaluation

Messick's (1989) framework shares an emphasis on multiple perspectives and unintended consequences with the social advocacy approach, which Stufflebeam (2001) ranks as the most promising for the 21st century. This approach aims at providing a full understanding of both intended and unintended consequences for the betterment of the program, institution or society. Unintended consequences signal "that we have been incomplete or off-target" in "development", "interpretation" and "use" (Messick, 1998, p. 43), and areas in need of program improvement. By including both value implications and unintended consequences in their evaluation models, distance educators can be advocates for the "moral good", and their evaluation results used for the betterment of society. Finally, within the social advocacy approach, Stufflebeam (2001) includes client-centred or responsive evaluations, which can provide insights into applying the adapted Messick's (1989) framework to distance and distributed courses.

A responsive approach to evaluation of distance programs

The adapted Messick's (1989) framework shares common elements with Stake's (1995) responsive approach, including the assumption of complexity, emergence and multiple values. Stake's (1995) approach aims for a comprehensive assessment of merit and worth through a comparison of intended and unintended or emergent effects, an understanding the full complexity of a program and a post-modernist value system where

the multiple perceptions, expectations and values of different stakeholders are brought forward (Abma & Stake, 2001). According to Stufflebeam (2001),

[The evaluation approach] seeks no final authoritative conclusion, interpreting results against stakeholders' often diverse and values. The approach seeks to examine a program's full countenance, and prizes the collection of multiple and often conflicting perspectives on the value of a program's format, operations and achievement. Side effects and incidental gains as well as intended outcomes are to be identified and examined" (p. 63).

The adapted Messick's (1989) framework can provide direction for evaluators in their sensitivity to the emerging particularities of the case in all its complexity. According to Stronach (2001), the "unspoken" Stake (1995) contains "a largely unaddressed tension between *direction* (in the form of theory, principles, procedures, tips, checklists, and the like) and what is called here for want of an appropriate term *indirection*—those references to the ineffable nature of education or the research task" (p. 67). Stake's (1995) narrative approach would also be appropriate as a methodology for evaluating distance courses using the adapted Messick's (1989) framework.

The findings that distance/distributed courses can serve multiple purposes and reflect multiple and diverse agendas calls for a post-modern approach to values. According to Moss (1998), Messick (1989) is a postmodernist, while Stronach (2001) implies that Stake can be "recruited" to the postmodern. Being explicit about their own values, and about the values implied by the course objectives, design and activities can lead evaluators to be more cautious in their judgements of merit and worth and to clarify the boundaries, differences and sources of conflict (Messick, 1989). However, in the end, Stake (2001)

maintains that a post-modern approach to values does not preclude the final outcome—that it is the evaluator who must make the final judgment of merit and worth.

Where Do We Go From Here?

One limitation of Messick's (1989) framework is that it has seldom been applied in practice despite its considerable theoretical influence (Zumbo, personal communication, August 10, 2002). One reason is the resistance of standardized test developers to the notion of unintended consequences, which often takes the form of arguments about the practical difficulties of using the framework in evaluation practice (e.g. Green, 1998), despite Cizek's (2001) "evidence for 10 unintended, unrecognized and unarticulated positive consequences of high-stakes testing" (p. 19).

I believe that Messick's (1989) framework is suited more to program evaluation than to the context from which it emerged, that is, assessment. This research shows that it can be applied much more easily to program evaluation than to assessment data. With a qualitative methodology, the categories of the framework can be used as categories to code the data. In survey research, they can be used as key constructs to guide the writing of survey items. For this reason, the adapted Messick's (1989) framework is better suited to program evaluation than to assessment, where it can be unclear how to apply the framework to the data from standardized testing.

The Need for Further Research

Because case studies are unique, the adapted Messick's (1989) framework needs to be applied to other distance education courses and programs to determine how well the categories of the consequential basis are supported by the evidence. In this research, unintended social consequences emerged because interviewees made conceptual associations based on the

interview questions, none of which dealt with unintended consequences. Because it was not included in the ACTION framework, there were no questionnaire items on this variable. If our survey and interview questions had included this variable, it is likely that we would have obtained more data to support this facet of value. Despite this situation, however, some unintended social consequences did emerge for each of the three courses in this study, which supports its inclusion as a facet of value for distance courses.

In addition, these new applications will help to clarify the assumptions underlying the framework. The findings of this research, for example, show that the tension between facts and values can “pull” a course in different directions. The implication is that this “tension” is symptomatic of a “problem” with the course. The implication is that “good” courses reflect a singular value basis or “complementary” value bases. This may not necessarily be the case, and future evaluation studies of distance/distributed courses may demonstrate that there are considerable benefits of courses designed along multiple value systems.

The ease with which the adapted Messick’s (1989) framework can be used in evaluation practice increases the likelihood of its adoption and use by evaluators. In the field of distance education, the problem is that a discussion of unintended consequences in the literature has been avoided for 20 years. This resistance makes it more likely that the framework will be adopted by evaluators in fields other than distance education, especially because it broadens the inquiry by including positive unintended consequences. Yet in another sense, it hardly matters if distance educators reject the adapted framework, which I predict they will, because university faculty and administrators will have a powerful new tool with which to question the findings of distance education evaluation studies. By broadening the boundaries of the inquiry

and raising the “right to know”, these professors and administrators will be contributing to improved evaluation practice and to social justice.

Is the adapted Messick’s framework the last word?

In response to the question, “Is the adapted Messick’s framework the last word?”, the answer is “no”. Applying this framework to the data also raises the question of whether evaluation evidence belongs under only one category or whether it could belong under more than one category. Does the category or categories in which the same evidence is “placed” matter? Should evidence be placed in only one category, or in more than one category? What are the strengths and limitations of using one box instead of another? Cycling through the framework for Modern Languages 430, for example, demonstrates that survey items on learner satisfaction with the relevance of the software to their learning could be placed under both Relevance and Cost/benefit. Which box should one choose, and does it matter? Further research in various contexts is needed to address these kinds of questions.

Dissemination of the findings

Because more and broader applications will bring out additional strengths and weaknesses, the adapted framework also needs to be applied to other educational programs, as well as to programs in the fields of health or social welfare. For this to occur, these findings need to be disseminated. Hence, the next step is to present the framework at conferences, publish these findings and talk to people about this research. So far, the adapted Messick’s framework has been well received. At the invitation of Rutgers University, I presented some of these findings at the NCME annual conference in Seattle in 2001. My paper on the adapted Messick’s framework has already been published in an international journal (Ruhe, 2002), and I also plan to have these findings published. Moreover, in my new position as Research

Associate/Project Coordinator at the Center for Research and Evaluation at the University of Maine, I expect to use the adapted Messick's (1989) framework, or some version of it approved by the research team, to evaluate a state-wide literacy program for Grade 1 children. Another likely application is the evaluation of a new state-wide program, which takes effect in 2002-2003, to give every Grade 7 child in selected schools in the State of Maine a laptop.

Conclusion

In the new century, as distance education continues to expand and becomes more central to the work of the academy, there is a need for an evaluation framework which will provide stakeholders with a comprehensive assessment of the merit and worth of distance programs. According to Gooler (1979), "it makes sense to apply the notions of evaluation to distance educational programs" (p. 43). The adapted Messick's (1989) framework is a framework from the field of assessment, which includes categories commonly found in distance education evaluation models such as learner response, cost-benefit and relevance. My findings have demonstrated the benefits of using the framework as a model to guide the evaluation of distance programs. The adapted Messick's (1989) framework makes several unique contributions, including an insightful understanding of the term "unintended consequences", value implications, the tension between facts and values and a pluralistic approach to values. It is truly remarkable that Messick's (1989) framework contributes so many insights to the evaluation of distance instructional programs in fewer than 20 words.

In this research, I took Messick's (1989) framework for a "test-drive" by applying it to the data from three BC post-secondary courses. My purpose was to demonstrate how the framework "works" in actual evaluation practice. The framework brought issues around learner response, costs, relevance, unintended consequences, value implications and the

interaction of these aspects of value from the background to the foreground. Secondly, the underlying assumption of the tension between facts and values provides a “full countenance” of the workings of the sub-systems of the implementation system. Finally, the plurality of values resonates with the approach to values in the social advocacy approaches to program evaluation, in particular Stake’s (1983) client-centered approach.

By including the consequential basis of value, the adapted Messick’s (1989) framework provides a comprehensive assessment of the merit and worth of distance instructional programs. In contrast, traditional evaluation models in distance education have been based only on the evidential basis, and consequently, are likely to provide an incomplete picture of merit and worth. Because it raises the “right-to-know”, the adapted Messick’s (1989) framework can be used to enhance the quality of evaluation of distance courses in a way which advocates for social justice. In his ranking of 22 approaches to evaluation, Stufflebeam (2001) ranks client-centred evaluation as one of the top five most promising approaches to program evaluation for the twenty-first century. By situating Messick’s (1989) framework within this approach, my research has fulfilled its promise—to merge the fields of distance education and program evaluation, and to bring the insights of the latter into the former, with a view to improving the quality of evaluation practice in distance education.

References

Abma, T. A., & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. In Greene, J. C., & Abma, T.A. (Eds.), Responsive Evaluation, New Directions in Evaluation, No. 92. (pp. 7-21). San Francisco, CA: Jossey-Bass.

Academic Committee for the Creative Use of Learning Technologies (2000). The creative use of learning technologies. Unpublished manuscript. University of British Columbia.

American Council on Education (March 8, 2002). Guiding principles: An overview. Retrieved May 24, 2002 from American Council on Education web site:
http://www.acenet.edu/calec/dist_learning/

American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. Psychological Bulletin, 51(2, Pt. 2).

American Psychological Association (1966). Standards for educational and psychological tests and manuals. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974). Standards for educational and psychological tests. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Andrusyszyn, M., & Davie, L. (1997). Facilitating reflection through interactive journal writing in an online graduate course: A qualitative study. Journal of Distance Education, 12(1/2), 103-126.

Angoff, W.H. (1988). Validity: An evolving concept. In H. Wainer & H. I. Braun (Eds.), Test Validity (pp. 19-31). Hillsdale, NS: Lawrence Erlbaum.

BC Open University (2000a). A message from Terry Piper, Open University's VP, Education and Provost of BCOU. Retrieved May, 2001 from BC Open University website: <http://www.ola.bc.ca/bcou/piper-msg.html>

BC Open University (2000b) BC Open University. Retrieved May, 2001 from BC Open University web site: <http://www.ola.bc.ca/bcou/home.html>

BC Open University (2000c). Financial Services. Retrieved May, 2001 from BC Open University web site: <http://www.ola.bc.ca/bcou/services/financial.html#bridging>.

Banathy, B. H. (1999). Systems design of education: Concepts and principles for effective practice. Englewood Cliffs, NJ: Educational Technology Publications.

Bartolic-Zlomislic, S. & Bates, A. W. (1999). Assessing the costs and benefits of tele-learning. Ottawa, ON: Office of Learning Technologies. Retrieved February 10, 2001 from University of British Columbia, Distance Education and Technology web site: <http://det.cstudies.ubc.ca/detsite/framewhat-index.html>

Bassey, M. (1999). Case study research in educational settings. Philadelphia, PA: Open University Press.

Bates, A. W. (1995). Technology, open learning and distance education. London: Routledge.

Bates, A. W. (2000). Managing technological change: Strategies for college and university leaders. Windsor, ON: Jossey-Bass.

Belanger, F., & Jordan, D. H. (2000). Evaluation and implementation of distance learning: Technologies, tools and techniques. Hershey, PA: Idea Group.

- Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. Education Policy Analysis Archives, 9(5). Retrieved April, 2001 from web site:
<http://epaa.asu.edu/epaa/v9n5.html>
- Bernstein, R. J. (1976). The restructuring of social and political theory. Pennsylvania, PA: The University of Pennsylvania Press.
- Bickman, L. (1987). The functions of program theory. In L. Bickman (Ed.), Using Program Theory in Evaluation: New Directions for Program Evaluation, No. 33 (pp. 5-18). San Francisco, CA: Jossey-Bass.
- Bourdeau, J., & Bates, A. (1997). Instructional design for distance learning. In S. N. Dijkstra, N. M. Seel, F. Shott, & R. D. Tennyson (Eds.), Instructional design: International perspectives: Volume 2. Solving instructional design problems (pp. 369-397). Mahwah, NJ: Lawrence Erlbaum.
- Braun, H. (1994). Assessing technology in assessment. In E. L. Baker & M. F. O'Neil, Jr. (Eds.), Technology Assessment in Education and Training (pp. 231-246). Hillsdale, NJ: Lawrence Erlbaum.
- Caracelli, V. J. (2000). Evaluation use at the threshold of the 21st century. In Caracelli, V. J. & Preskill, H. (Eds.), The Expanding Scope of Evaluation Use. New Directions for Evaluation, 88, (pp. 99-111). San Francisco, CA: Jossey-Bass.
- Caulder, J. (1994a). Course evaluation: Improving academic quality and teaching effectiveness. In F. Lockwood (Ed.), Materials Production in Open and Distance Learning (pp. 361-370). London: Paul Chapman Publishing.
- Caulder, J. (1994b). Programme evaluation and quality: A comprehensive guide to setting up an evaluation system. London, UK: Kogan Page.

- Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. American Journal of Evaluation, 19(1), 35-55.
- Chen, H. T. (1990). Theory-driven evaluation. Newbury Park, CA: Sage Publications.
- Cheung, D. (1998). Developing a student evaluation instrument for distance teaching. Distance Education, 9(1), 23-42.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. Educational Measurement, Issues and Practice, 20(4), 19-27.
- Clark, R. E. (1994a). Assessment of distance learning technology. In E. L. Baker & O'Neil, Jr. (Eds.), Technology assessment in education and training (pp. 63-78). Hillsdale, NJ: Lawrence Erlbaum.
- Clark, R. E. (1994b). Media will never influence learning. Educational Technology Research and Development, 42, 21-29.
- Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. Educational Technology Research and Development, 45 (4), 21-35.
- Coldeway, D.O. (1988). Methodological issues in distance education research. The American Journal of Distance Education, 2(3), p. 45-54.
- Collis, B. A. (1993). Evaluating instructional applications of telecommunications in distance education. Educational and Training Technology International, 30(3), 266-74.
- Cook, T.D. (1985). Post-positivist critical multiplism. In R. L. Shotland & M. M. Mark (Eds.), Social Science and Social Policy, (pp. 21-62). Thousand Oaks, CA: Sage.
- Creeve, J. C., & Caracelli, V. J. (Eds.) (1997). Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. San Francisco, CA: Jossey-Bass.

- Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. Toronto: Holt, Rinehart & Winston.
- Cronbach, L. J. (1982). Designing evaluations of educational and social programs. San Francisco, CA: Jossey-Bass.
- Cronbach, L. J. (1989). Five perspectives on validity argument. In H. Wainer & H. I. Braun, (Eds.) Test Validity (p. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L.J. (1990). Essentials of psychological testing (5th ed.). New York : Harper & Row.
- Cronbach, L. J. & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-32.
- Cureton, E.E. (1951). Validity. In E.R. Lindquist (Ed.), Educational Measurement (pp. 621-694). Washington, DC: American Council on Education.
- Davidson, E. J. (2000). Ascertaining causality in theory-based evaluations. In Rogers, P. J., Hacsı, T. A., Petrosino, A., & Huebner, T. A. (Eds.), Program Theory in Evaluation: Challenges and Opportunities. New Directions for Evaluation, (pp. 17-26). San Francisco, CA: Jossey-Bass.
- Davis, B. (2002). Complexity science and schooling. Presentation at Faculty of Education, UBC, February 26, 2002.
- De Jong, T., & van Joolingen, W. (1998). Scientific discovery learning with computer simulations and conceptual domains. Review of Educational Research, 68(2) 179-201.
- Duffy, T.M. & Jonassen, D.H. (Eds.) (1992). Constructivism and the technology of instruction: A conversation. Lawrence Erlbaum Associates.

- Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532-550.
- Fabos, B., & Young, M. D. (1999). Telecommunication in the classroom: Rhetoric versus reality. Review of Educational Research, 69 (3), 217-259.
- Fahy, P. J. (1998). Reflections on the productivity paradox and distance education technology. Journal of Distance Education, 13(2), 66-73.
- Farhad, S. (1999). Towards a systems theory of distance education. American Journal of Distance Education, 13(2), 24-31.
- Flagg, B. (1990). Formative evaluation for educational technologies. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Flick, U. (1992). Triangulation revisited: Strategy of validation or alternative? Journal for Theory of Social Behaviour, 22(2), 175-198.
- Forsyth, I., Jolliffe, A. & Stevens, D. (1995). Evaluating a course: Practical strategies for teachers, lecturers and trainers. London, UK: Kogan Page.
- Foucault, M. (1977). Discipline and punish. New York, NY: Vintage Books.
- Foucault, M. (1980). Two lectures. In C. Gordon (Ed.), Power/Knowledge: Selected Interviews and Other Writings by Michel Foucault (pp. 78-108). New York, NY: Pantheon Books.
- Frank, T. (2000). Universities compete for a presence online. University Affairs, 41 (8), 10-14. Ottawa: Association of Universities and Colleges of Canada.
- Gadamer, H. G. (1981). Reason in the age of science. Cambridge, MA: MIT Press.

- Garrison, D. R. Anderson, T. D. (1999). Avoiding the industrialization of research universities: Big and little distance education. American Journal of Distance Education, 13(2), 48-63.
- Galston, W.A. (1999). Value pluralism and liberal political theory. American Political Science Review, 93(4), 769-778.
- Gibson, C. (Ed.). (1998). Distance learners in higher education: Institutional responses for quality outcomes. Madison, WI: Atwood.
- Gooler, D. (1979). Evaluating distance education programmes. Canadian Journal of University Continuing Education, 6(1), 43-55.
- Gray, A., & O'Grady, G. (1993). Interactive television: Expanding established skills or new experiences for old? In M. Ryan (Ed.), Proceedings of APITITE 94 (pp. 665-670). Murray Hill, NSW, Australia: Australian Computer Society.
- Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. Educational Measurement: Issues and Practice, 17(2), 16-19.
- Guion, R. M. (1980). On trinitarian doctrines of validity. Professional Psychology, 11, 385-398.
- Hammond, R. L. (1973). Evaluation at the local level. In B. R. Worthen & J. R. Sanders, (Eds.), Educational Evaluation: Theory and Practice. Worthington, OH: Charles A. Jones (pp. 157-169).
- Hannafin, M. J., Hannafin, K. M.; Land, S. M.; Oliver, K. (1997). Grounded practice and the design of constructivist learning environments. Educational Technology Research and Development, 45(3), 101-117

- Harasim, L., Hiltz, S. R., Teles, L., & Turoff, M. (1996). Learning networks: A field guide to teaching and learning online. London, UK: MIT.
- Harrison, P. J., Saba, F., Seeman, B. J., Molise, G., Behm, R., & Williams, M. D. (1991). Development of a distance education assessment instrument. Educational Technology Research and Development, 39(4), 65-77.
- Hartley, K. & Robson, L. (1998a). Teams that work: A team skills handbook for students. Burnaby, BC: BCIT.
- Hartley, K. & Robson, L. (1998b). Teaching teams that work: A faculty guide. Victoria, BC: Province of British Columbia: Ministry of Advanced Education, Training and Technology.
- Heinzen, T. E., & Alberico, S. M. (1990). Using a creativity paradigm to evaluate teleconferencing. American Journal of Distance Education, 4(3), 3-12.
- Henderson, L., & Putt, I. (1999). Evaluating audioconferencing as an effective learning tool in cross-cultural contexts. Open Learning, 14(1), 25-37.
- Henry, G.T. (2000). Why not use? In V. J. Caracelli & H. Preskill (Eds), The Expanding Scope of Evaluation Use, New Directions for Evaluation, No. 88 (pp. 85-98). San Francisco, CA: Jossey-Bass.
- Henry, G.T. & Julnes, G. (1998). Values and realist evaluation. In G.T. Henry, G. Julnes & M. M. Mark (Eds.), Realist Evaluation: An Emerging Theory in Support of Practice. New Directions for Evaluation, No. 78 (pp. 53-71). San Francisco, CA: Jossey-Bass.
- Herrmann, A., Fox, R., & Boyd, A. (1999). Benign educational technology? Open Learning, 14(1), 3-7.

House, E. R. & Howe, K. R. (2000). Deliberative democratic evaluation. In Ryan, K. E. & deStefano (Eds.), Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Evaluation, 85, (pp. 3-12). San Francisco, CA: Jossey-Bass.

Hubley, A.M. & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. The Journal of General Psychology, 123(3), 207-215.

Institute for Higher Education Policy. (1999). What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education.

Washington, DC: The Institute for Higher Education Policy. Retrieved December 22 from Institute for Higher Education Policy web site: <http://www.ihep.com/difference.pdf>

Johnstone, S. M., & Krauth, B. (1996). Balancing quality and access: Some principles of good practice for the virtual university. Change, 28(2), 38-41.

Joint Committee on Standards for Educational Evaluation. (1994). Program evaluation standards: How to assess evaluation of educational programs (2nd ed.). Thousand Oaks, CA: Sage.

Jones, A., & Petre, M. (1994). Computer-based practical work at a distance: A case study. Computers in Education, 22(1-2), 27-37.

Jones, A., Scanlon, E., Butcher, P., Greenberg, J., Ross, S., Murphy, P., & Tosunoglu, C. (1998). Learning with computers: Experiences of evaluation. Computers and Education, 30(1/2), 1-9.

Jones, A., Scanlon, E., Tosunoglu, C., Ross, S., Butcher, P., Murphy, P., & Greenberg, J. (1996). Evaluation of computer assisted learning at the Open University—15 years on. Computers and Education, 26(1-3), 5-15.

- Kanuka, H., & Anderson, T. (1998). Online social interchange, discord and knowledge construction. Journal of Distance Education, 13(1), 57-74.
- Keegan, D. (1993). Reintegration of the teaching acts. In D. Keegan (Ed.), The theoretical principles of distance education (2nd ed.) (pp. 113-134). New York, NY: Routledge.
- Klinger, S. (2002). "Are they talking yet?" Online discourse as political action in an education policy forum. Vancouver: UBC: Unpublished doctoral dissertation.
- Knapper, C. K. (1980). Evaluating instructional technology. New York, NY: John Wiley & Sons.
- Lane, C. (1989). A selection model and pre-adoption evaluation instrument for video programs. American Journal of Distance Education, 3(3), 46-57.
- Lecompte, M. D., & Preissle, J. (1993). Ethnography and qualitative design in educational research (2nd ed.). Toronto: Academic Press.
- Levin, H. M. (1983). Cost-effectiveness: A primer. New perspectives in evaluation, v. 4. Beverly Hills, CA: Sage.
- Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In Light, R. J. (Ed.). Evaluation Findings that Surprise. New Directions for Evaluation, No. 90. (pp. 55-68). San Francisco, CA: Jossey-Bass.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1997). Complex performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 5-21.
- Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of science. In D. S. Cordray (Ed.), Utilizing Prior Research in Evaluation Planning. New Directions for Program Evaluation, No. 27 (pp. 51-87).

- Lookatch, R. (1997). Multimedia improves learning—Apples, oranges and the Type I error. Contemporary Education, 68(2), 110-113.
- Mabry, L. (2001). Responsive evaluation is to personalized assessment as... In Greene, J. C., & Abma, T.A. (Eds.), Responsive Evaluation, New Directions in Evaluation, No. 92. (pp. 89-102). San Francisco, CA: Jossey-Bass.
- MacPhail, F. (1998). Moving beyond statistical validity in economics. Social Indicators Research, 45(1), 119-149.
- Madaus, G. F. & Kellaghan, T. (2000). Models, metaphors and definitions in evaluation. In Stufflebeam, D. L., Madaus, G. F. & Kellaghan, T. (Eds.). Evaluation models: Viewpoints on educational and human services evaluation (2nd ed., p. 19-32). Boston, MA: Kluwer Academic Publishers.
- Mann, C.C. (1998). Quality assurance in distance education: The Surrey MA (TESOL) experience. Distance Education, 19(1), 7-22.
- Markus, K. (1998). Science, measurement and validity: Is completion of Messick's synthesis possible? Social Indicators Research, 45(1), 7-34.
- McAlister, S. (1998). Credible or tentative? A model of Open University students with "low" educational qualifications. Open Learning, 13(3), 33-42.
- McCulloch, K. H. (1997). Participatory evaluation in distance learning. Open Learning, 12(1), 24-30.
- Mehrens, W.A. (1997). The consequences of consequential validity. Educational Measurement: Issues and Practice, 16(2), 16-18.
- Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. Journal of Counseling Psychology, 37, 113-115.

- Melton, R. F. (1995). Developing a formative evaluation system for distance teaching. Open Learning, 10(2), 53-57.
- Menzies, H. (1996). Whose brave new world? The information highway and the new economy. Toronto: Between the Lines.
- Merriam, S. B. (1989). Case study research in education: A qualitative approach. San Francisco, CA: Jossey-Bass.
- Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), Test Validity. Hillsdale, NJ: Laurence Erlbaum.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (3rd ed., pp.13-103). New York, NY: MacMillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13-23.
- Messick, S. (1995a). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741-749.
- Messick, S. (1995b). Standards of validity and the validity of standards in performance assessment. Educational Measurement: Issues and Practices, 14(4), 5-8.
- Messick, S. (1996). Validity and washback in language testing. Language Testing, 13(3), 241-256.

- Messick, S. (1998). Test validity: A matter of consequence. Social Indicators Research, 45(1), 35-44.
- Michalos, A.C. (1992). Ethical considerations in evaluation. Canadian Journal of Program Evaluation, 7(2), 61-75.
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage.
- Ministry of Advanced Education, Training and Technology, Government of British Columbia. (2000). Charting a new course. Part 1: The context. Retrieved November 22, 2000 from Ministry of Advanced Education, Training and Technology, Government of British Columbia web site: <http://www.aett.gov.bc.ca/strategic/newcourse/part1.htm#driving>
- Moore, M., & Kearsley, G. (1996). Distance education: A systems view. Toronto: Wadsworth.
- Moss, P. (1998a). Recovering a dialectical view of rationality. Social Indicators Research, 45(1), 55-67.
- Moss, P. (1998b). The role of consequences in validity theory, Educational Measurement: Issues and Practice, 17(2), 6-12.
- Nielsen, H. D. (1997). Quality assessment and quality assurance in distance teacher education. Distance Education, 18 (2), 284-317.
- Noble, D. F. (1999). Digital diploma mills: The automation of higher education. Toronto: Between the Lines.
- Oblinger, D., & Maruyama, M. (1996). Distributed learning. CAUSE: Professional Paper Series No. 14. Boulder, CO: CAUSE.

Open Learning Agency (1996a). Introductory Psychology I: Assignment file. Form A. (Print Version). Burnaby, BC: Open Learning Agency.

Open Learning Agency (1996b). Introductory Psychology I: Course manual. Basic information about the course. (Print Version) Burnaby, BC: Open Learning Agency.

Oppenheimer, T. (1997). The computer delusion. Atlantic Monthly, 280(1), 45-62.

Organization for Economic Co-operation and Development (1999). No significant difference. Paris: Organization for Economic Co-operation and Development.

Parlett, M., & Hamilton, D. (1977). Evaluation as illumination: A new approach to the study of innovatory programmes. In D. Hamilton, D. Jenkins, C. King, B. McDonald, & M. Parlett (Eds.), Beyond the Numbers Game: A Reader in Educational Evaluation (pp. 6-22). Basingstoke, UK: Macmillan.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Newbury Park, CA: Sage.

Patton, M. Q. (1997). Utilization-focussed evaluation: The new century text. (3rd ed.) Newbury Park, CA: Sage.

Popham, W. J. (1993). Educational evaluation. Toronto: Allyn & Bacon.

Popham, W. J. (1997). Consequential validity: Right concern—wrong concept. Educational Measurement: Issues and Practice, 16(2), 9-13.

Preskill, H. & Torres, R. T. (2000). The learning dimension of evaluation use. In Caracelli, V. J. & Preskill, H. (Eds.), The Expanding Scope of Evaluation Use. New Directions for Evaluation, 88, (pp. 25-37). San Francisco, CA: Jossey-Bass.

Qayyum, A. (1999). Case #7. Learning through new technologies: The response of adult learners to Microsoft Certified Engineer Program at Burnaby Community Skills Centre.

Retrieved February 10, 2001 from University of British Columbia, Distance Education and Technology web site: <http://det.cstudies.ubc.ca/detsite/frameworkat-index.html>

Reckase, M.D. (1998a). Consequential validity from the test developer's perspective. Educational Measurement: Issues and Practice, 17(2), 13-16.

Reckase, M.D. (1998b). The interaction of values and validity assessment: Does a test's level of validity depend on a researcher's values? Social Indicators Research, 45(1), 45-54.

Reigeluth, C. M. (Ed.) (1999). Instructional design theories and models: A new paradigm of instructional theory. Volume 2. Mahwah, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & Frick, T. (1999). Formative research: A methodology for creating and improving design theories. In C. M. Reigeluth (Ed.), Instructional-design theories and models: A new paradigm of instructional theory (Vol. 2, pp. 633-651). Mahwah, NJ: Lawrence Erlbaum Associates.

Robson, J. (2000). Evaluating on-line teaching. Open Learning, 15(2), 151-171.

Roche, J. (1997). Reading German: A Multimedia Self-Study Course on Reading German Language for Professional and Technical Purposes. Introduction. Vol. 1. Vancouver, B.C.: UBC Access, 1997.

Rogers, P. J., Petrosino, A., Huebner, T. A., Hacs, T. A. (2000). Program theory evaluation: Practice, promise and problems. In D. J. Rog & D. Fournier (Eds.), Progress and Future Directions in Evaluation: Perspectives on Theory, Practice and Methods. New Directions in Evaluation, no. 76 (pp. 5-14). San Francisco, CA: Jossey-Bass.

Romiszwowski, A. (1981). Designing instructional systems. London: Kogan Page.

Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counselling psychology. Journal of Counselling Psychology, 35, 203-208.

Ross, S. M., & Morrison, G. R. (1997). Measurement and evaluation approaches in instructional design: Historical roots and current perspectives. In R. D. Tennyson, F. Schott, N. Seel, & S. Dijkstra (Eds.), Instructional design: International perspectives. Volume 1: Theory, research and models (pp. 327-351). Mahwah, NJ: Lawrence Erlbaum.

Rowley, D., Lujan, H., & Dolence, M. (1998). Strategic choices for the academy: How demand for lifelong learning will Re-create Higher Education. San Francisco, CA: Jossey-Bass.

Ruhe, V. (1999a). Case #5. Learning through new technologies: The response of adult learners to German 430. Retrieved February 10, 2001 from University of British Columbia, Distance Education and Technology web site: <http://det.cstudies.ubc.ca/detsite/framewhat-index.html>

Ruhe, V. (1999b). Case # 12. Learning through new technologies: The response of adult learners to Psychology 101. Retrieved February 10, 2001 from University of British Columbia, Distance Education and Technology web site: <http://det.cstudies.ubc.ca/detsite/framewhat-index.html>

Ruhe, V. (1998). E-mail exchanges: Teaching language, culture and technology for the 21st century. TESL Canada Journal, 16 (1), 88-95.

Ruhe, V. & Qayyum, A. (1999). A cross-case comparison of the "Learning through new technologies: The response of adult learners" project. Vancouver: University of British Columbia, Distance Education and Technology. Retrieved February 10, 2001 from University of British Columbia, Distance Education and Technology web site:
<http://det.cstudies.ubc.ca/detsite/framewhat-index.html>

Rumble, G. (1981). Evaluating autonomous multimedia distance education systems: A practical approach. Distance Education, 21, 64-90.

Russell, T. (1999). The no significant difference phenomenon. Retrieved January 21, 2000 from web site: <http://tenb.mta.ca/phenom/phenom.html>

Salomon, G. (Ed.). (1993). Distributed cognitions: psychological and educational considerations. New York, NY: Cambridge University Press.

Scanlon, E., Jones, A., Barnard, J., Thompson, J. & Calder, J. (2000). Evaluating information and communication technologies for learning, Educational Technology and Society, 3(4), 1-10. Retrieved February 14 from: http://ifets.ieee.org/periodical/vol_4_2000/scanlon.html

Schofield, J. (1999, September 6). Back to School Online. Macleans, 112, 22-29.

Schwandt, T. A. (1989). Recapturing moral discourse in evaluation. Educational Researcher, 18(8), 11-16.

Schwandt, T. A. (2000). The landscape of values in evaluation. In D. J. Rog & D. Fournier (Eds.), Progress and Future Directions in Evaluation: Perspectives on Theory, Practice and Methods. New Directions in Evaluation, No. 76 (pp. 25-40). San Francisco, CA: Jossey-Bass.

- Scriven, M. (1972). Pros and cons about goal-free evaluation. Evaluation Comment, 3(4), 1-8.
- Seely Brown, J. S., & Duguid, P. (2000). The social life of information. Boston, MA: Harvard Business School Press.
- Selwyn, E. (1997). The continuing weakness of educational computing research. British Journal of Educational Technology, 28 (4), 305-307.
- Shalinsky, Andrea (1998). Microsoft certified systems engineer orientation guide for Burnaby Community Skills Centre. Burnaby, BC: Open Learning Agency, Workplace Training Systems.
- Shepard, L.A. (1997). The centrality of test use and consequences for test validity. Educational Measurement: Issues and Practice, 16(2), 5-8.
- Sidani, S., & Sechrest, L. (1999). Putting program theory into operation. American Journal of Evaluation, 20 (2), 227-238.
- Simonson, M., Schlosser, C. & Hanson, D. (1999). Theory and distance education: A new discussion. American Journal of Distance Education, 13(1), 60-75.
- Simpson, W. B. (1991). Cost containment for higher education: Strategies for public policy and institutional administration. New York, NY: Praeger.
- Singer, E.A., Jr. (1959). In C.W. Churchman (Ed.), Experience and Reflection Philadelphia, PA: University of Pennsylvania Press.
- Smith, P. J., & Smith, S. N. (1999). Differences between Chinese and American learners: Some implications for distance educators. Distance Education, 20(1), 64-80.
- Stake, R. E. (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540.

- Stake, R. E. (1975). Evaluating the arts in education: A responsive approach. Columbus, OH: Merrill.
- Stake, R. E. (1995). The art of case study research. London: Sage.
- Stenhouse, L. (1988). Case study methods. In J. P. Reeves (Ed.), Educational research, methodology, and measurement: An international handbook (pp. 41-53). Oxford: Pergamon.
- Stoll, C. (1999). High-tech heretic: Why computers don't belong in the classroom and other reflections of a computer contrarian. New York, NY: Random House.
- Stronach, I. (2001). The changing face of responsive evaluation: A postmodern rejoinder. In Greene, J. C., & Abma, T.A. (Eds.), Responsive Evaluation, New Directions in Evaluation, No. 92. (pp. 59-72). San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. L. (2001). Evaluation models. New Directions for Evaluation, No. 89. San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. I., Merriam, H. O., & Provos, M. N. (1971). Educational evaluation and decision-making. Bloomington, IN: Phi Delta Kappa.
- Suen, H., & Stevens, R. (1993). Analytic considerations in distance education research. American Journal of Distance Education, 7(3), 61-69.
- Tashakkori, A., & Teddlie, C. (1998). Mixed methodology: Combining qualitative and quantitative approaches. Thousand Oaks, CA: Sage.
- Tenner, A. R. (1996). Why things bite back: Technology and the revenge of unintended consequences. New York, NY: Knopf.

Tennyson, R. D. (1997a). A system dynamics approach to instructional systems development. In R. D. Tennyson, F. Shott, N. Seel, & S. Dijkstra (Eds.), Instructional Design: International Perspectives. Volume 1: Theory, Research and Models (pp. 413-426). Mahwah, NJ: Lawrence Erlbaum.

Tennyson, R. D. (1997b). Evaluation techniques in instructional development. In S. Dijkstra, N. Seel, F. Shott, & R. D. Tennyson (Eds.), Instructional Design: International Perspectives. Volume 1: Theory, Research and Models (pp. 19-26). Mahwah, NJ: Lawrence Erlbaum.

Tessmer, M. (1993). Planning and conducting formative evaluations: Improving the quality of education and training. Philadelphia, PA: Kogan Page.

Thorpe, M. (1988). Evaluating open and distance learning. Harlow, UK: Longman.

TLT Group (2000). The flashlight project. Retrieved November 4, 1999 from <http://www.tltgroup.org/programs/flashcsi.html>

Tyler, R. W. (1942). General statement on evaluation. Journal of Educational Research, 35, 492-501.

Ungerleider, C. S. & Burns, T. C. (2002, April-May). Information and communication technologies in elementary and secondary education: A state of the art review. Paper presented at the 2002 Pan-Canadian Education Research Agenda Symposium. April 30-May 2, 2002, Montreal, PQ. Retrieved June 27, 2002 from: <http://www.cmec.ca/stats/pcera/rsevents02/>

Universitas 21 (2002). Introduction. Retrieved March 15, 2002 from Universitas 21 web site: <http://www.universitas.edu.au/introduction.html>

U.S. General Accounting Office, Program Evaluation and Methodology Division. (1990). Case study evaluations. Washington, DC: Government Printing Office.

- Van Maanen, J. (1988). Tales of the field: On writing ethnography. Chicago: University of Chicago Press.
- Van Slyke, C., Kittner, M., & Belanger, F. (1998). Identifying candidates for distance education: A telecommuting perspective. Proceedings of the America's Conference on Education Systems (pp. 666-668).
- Wade, C. & Tavris, C. (1997). Psychology. (4th ed.). Toronto: Addison-Wesley.
- Weiss, C. H. (1986). The stakeholder approach to evaluation. In R. E. House (Ed.), New Directions in Program Evaluation (pp. 145-157). Philadelphia, PA: Falmer Press.
- Weiss, C. H. (1998). Evaluation: methods for studying programs and policies. 2nd ed. Upper Saddle River, N.J.: Prentice Hall.
- Werthman, M.S. (1996). Psychology: The study of human behavior. Tele-course guide. (3rd ed.). Costa Mesa, CA: Harper Collins College Publishers.
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. In L. Bickman (Ed.), Using Program Theory in Evaluation: New Directions for Program Evaluation, No. 33, (p77-92).
- Willis, J. (2000). The maturing of constructivist instructional design: Some basic principles that can guide practice. Educational Technology, 40(1), 5-16.
- Wilson, M., Qayyum, A., & Boshier, R. (1998). Worldwide America: Think globally, click locally. Distance Education, 19(1), 109-123.
- Wolf, R. M. (1987). The nature of education evaluation. International Journal of Educational Research, 1, 7-20.
- Woodley, A., & McIntosh, N. (1980). The door stood open: An evaluation of the Open University younger learners pilot scheme. Barcome, Sussex, UK: Falmer Press.

Worthen, B., Sanders, J., & Fitzpatrick, J. (1997). Program evaluation: Alternative approaches and practical guidelines. New York, NY: Longman.

Results: The final report will be available on a Webpage housed at UBC Continuing Studies (<http://research.cstudies.ubc.ca/RALP/>) after March 31, 1999.

Contact: If you have any questions or want further information concerning the study, you may contact either Valerie Ruhe or Adnan Qayyum at any time. If you have any concerns with your rights or treatment as a research participant, you may contact Dr. Richard Spratley, Director of Research Services, UBC at 822-8598.

Consent: I understand that my participation in this study is entirely voluntary and that I may refuse to participate or withdraw from the study at any time without jeopardy to my class standing.

I consent to participate in this study and understand that I will receive \$20.00 compensation.

Signature of study participant

Date

Compensation:

Please indicate your name and address to which your \$20.00 will be sent.

Name (to whom cheque will be issued)

Street

City

Postal Code

Contact:

If you have any questions or want further information concerning the study, you may contact either Valerie Ruhe or Adnan Qayyum at any time. If you have any concerns with your rights or treatment as a research subject, you may contact Dr. Richard Spratley, Director of Research Services, UBC at 822-8598.

Consent: I consent to participate in this interview.

Signature of faculty member

Date

Signature of witness

Date

Appendix E: Questionnaire

LEARNER QUESTIONNAIRE

Please *fill in the blanks* or *circle* the appropriate choice(s). For questions involving a *scale of responses*, please read each statement and then *circle* the response which best shows what you think.

Not all questions will apply to your situation, depending on what class you are taking. If a question does **not** apply, please enter N/A (not applicable) as your response.

If there are any questions you feel uncomfortable with, just skip them and move on to the next item.

We estimate this questionnaire will take approximately 15 minutes to complete (This is based on pilot tests).

By completing this questionnaire, you will influence the quality of future courses you may take and how technology is used in those courses. Your co-operation is important and greatly appreciated.

COURSE: _____

INSTITUTION: _____

RESEARCH ID: _____

I. COURSE DELIVERY

By "delivery" we mean the method by which the course is given to the learners. Common delivery methods include:

- (a) face-to-face,
- (b) print-based distance (may include video/audiocassettes),
- (c) print-based distance with the addition of online, CD-ROM, teleconferencing, or video conferencing,
- (d) online, CD-ROM, teleconferencing, or video conferencing as the main delivery method and
- (e) a mix of technologies (i.e., online, CD-ROM, teleconferencing, or video conferencing).

Please use the following scale:

1 = Strongly Disagree

2 = Disagree

3 = Neither Agree nor Disagree

4 = Agree

5 = Strongly Agree

N/A = Not Applicable

1. A) I like this delivery method because it gives me flexibility in my studies (e.g., time, place, location).

1	2	3	4	5	N/A
---	---	---	---	---	-----

b) In this course, I am able to interact (communicate and exchange ideas):

i. With the instructor as much as I want.

1	2	3	4	5	N/A
---	---	---	---	---	-----

ii. With other learners as much as I want.

1	2	3	4	5	N/A
---	---	---	---	---	-----

c) In this course, the interaction

i. With the instructor is relevant to my learning.

1	2	3	4	5	N/A
---	---	---	---	---	-----

ii. With other learners is relevant to my learning.

1	2	3	4	5	N/A
---	---	---	---	---	-----

d) If this course was not offered in this delivery method, I would be unable to complete it.

1	2	3	4	5	N/A
---	---	---	---	---	-----

e) I would not take another course using this delivery method.

1	2	3	4	5	N/A
---	---	---	---	---	-----

2. I have limited experience with the various technologies.

YES

NO

If yes, please skip questions 3 and 4.

3. The delivery method(s) I prefer to use are (circle as many as apply):

a. Face-to-face

b. Print-based distance (may include video/audiocassettes)

c. Print-based distance with online, CD-ROM, teleconferencing, or video conferencing

d. Online, CD-ROM, teleconferencing, or video conferencing as the main delivery method

e. A mix of technologies

4. The delivery method(s) I prefer not to use are: (circle as many as apply.)

a. Face-to-face

b. Print-based distance (may include video/audiocassettes)

c. Print-based distance with online, CD-ROM, teleconferencing, or video conferencing,

d. Online, CD-ROM, teleconferencing, or video conferencing as the main delivery method

e. A mix of technologies

5. Have you had any problems taking this course in this delivery method (e.g., complications with admissions, inconvenient location, technical troubles, delay in receiving mailed materials)? If yes, please be specific about the problem and its impact.

6. What are the most important benefits of this delivery method for you? What drawbacks, if any, are there?

II. SUPPORT SERVICES

By "support services" we mean services the institution provides to learners to help them complete their education. Support services include but are not limited to: technical assistance, library facilities (including extension library resources), counselling services, and computer labs.

Please use the following scale:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither Agree nor Disagree
- 4 = Agree
- 5 = Strongly Agree
- N/A = Not Applicable

7. Support services for this course are unsatisfactory.

1 2 3 4 5 N/A

8. How can the existing support services be improved? In your response, please include the type of service you are describing.

9. What other support services should be available?

III. FOR COURSES USING TECHNOLOGY-BASED DELIVERY (i.e., online, CD-ROM, teleconferencing, or video conferencing) If you are in a course that does not use technology-based delivery, please circle NA.

Please use the following scale:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither Agree nor Disagree
- 4 = Agree
- 5 = Strongly Agree
- N/A = Not Applicable

10. a) When I began this course, I was worried about the delivery method.

1 2 3 4 5 N/A

b) At this point in the course I am comfortable with the delivery method.

1 2 3 4 5 N/A

c) Using technology in this course helps me learn:

i. With greater depth of understanding.

1 2 3 4 5 N/A

ii. More relevant information.

1 2 3 4 5 N/A

d) The technology increases my motivation to work on the course.

1 2 3 4 5 N/A

e) This course requires taking more personal responsibility for completion than does a face-to-face course.

1 2 3 4 5 N/A

f. I was not provided with enough training in the use of the technology at the start of the course.

1 2 3 4 5 N/A

g. I come to campus less often because of the technology used in the course.

1 2 3 4 5 N/A

h. I can learn better using print materials than by working on a computer.

1 2 3 4 5 N/A
 11. What changes to the technology, if any, do you think are needed? Please give specific examples.

12. For courses with a computer component. (If the course you are in does **not** have a computer component, please circle N/A).

a. Using the computer software (e.g., Virtual-U, WebCT, WebCSILE, Lotus Notes, TLM) for this course is boring.

1 2 3 4 5 N/A

b. Using the computer software for this course is easy.

1 2 3 4 5 N/A

c. I am not satisfied with the software used for this course.

1 2 3 4 5 N/A

IV. RESPONSE TO COURSE

Please use the following scale:

1 = Strongly Disagree

2 = Disagree

3 = Neither Agree nor Disagree

4 = Agree

5 = Strongly Agree

N/A = Not Applicable

13. a.) The tutor/instructor provides useful feedback.

1 2 3 4 5 N/A

b) The feedback I receive is individualized.

1 2 3 4 5 N/A

c) I do not receive feedback in a timely manner.

1 2 3 4 5 N/A

d) The course objectives are specific and meaningful.

1 2 3 4 5 N/A

e) The grading criteria are clear.

1 2 3 4 5 N/A

f) The course materials are well-organized.

1 2 3 4 5 N/A

g) The course materials are relevant to my personal or professional needs.

1 2 3 4 5 N/A

h.) The course objectives, content, and assessments are consistent

1 2 3 4 5 N/A

i) The marking is fair.

1 2 3 4 5 N/A

j) The course content is at about the right level of difficulty.

1 2 3 4 5 N/A

14. How do you rate the course materials? (Please circle).

Poor Fair Average Good Excellent

15. When you consider the course and the course materials, what works well? What needs to be improved? Why?

V. TIME DEMANDS

16. On average, how many hours per week do you spend working on this course? (If applicable, include time in class).

_____ hours

17. Is this more or less time than the average amount of time you spend working on courses in a traditional classroom setting?

More Less Same N/A Don't know

18. Is this more or less time than you expected to spend?

More Less Same N/A Don't know

19. If you have to travel to take this course, how much time do you spend travelling?

_____ hours per/week

Please use the following scale:

1 = Strongly Disagree

2 = Disagree

3 = Neither Agree nor Disagree

4 = Agree

5 = Strongly Agree

N/A = Not Applicable

20. This course is not worth the time it takes to complete.

1 2 3 4 5 N/A

VI. COSTS

21. Please estimate the expenses that are associated with your taking this course. Please fill in all that apply.

Course/registration fee	\$ _____
Travel	\$ _____
Accommodation	\$ _____
Per diem	\$ _____
Long distance telephone charges	\$ _____
Postage/courier	\$ _____
Textbooks	\$ _____
Software	\$ _____
Internet/Online costs	\$ _____
Parking	\$ _____
Other (please specify)	\$ _____

_____ \$ _____
N/A

22. Who pays for the above costs? Please estimate the amount that is paid by the following:

Myself (or a family member) \$ _____
 Employer \$ _____
 Institution offering the course \$ _____
 Other (please specify) _____ \$ _____

Please use the following scale:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither Agree nor Disagree
- 4 = Agree
- 5 = Strongly Agree
- N/A = Not Applicable

23. Taking this course in this delivery method costs less than other methods of delivery.

1	2	3	4	5	N/A
This course is not worth the money it costs.					
1	2	3	4	5	N/A

VII. INFORMATION ABOUT YOURSELF

24. Male Female

25. Year of birth: 19 _____

26. Please indicate your highest level of education.

- Some high school
- High school completed
- Some post secondary credit
- Certificate
- Diploma
- Bachelor's degree
- Master's degree
- Doctorate

27. How important are the following goals to you?

Please rate all that apply, using the following scale:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither Agree nor Disagree
- 4 = Agree
- 5 = Strongly Agree
- N/A = Not Applicable

• To obtain the qualification or credit	1	2	3	4	5	N/A
• Interest in the subject/content for its own sake	1	2	3	4	5	N/A
• Contact with distinguished instructors	1	2	3	4	5	N/A

- Content is relevant to the work I do/will do
1 2 3 4 5 N/A
- Socialize with others
1 2 3 4 5 N/A
- Personal growth/broaden perspective
1 2 3 4 5 N/A
- To show myself I can do it
1 2 3 4 5 N/A
- To get high grades
1 2 3 4 5 N/A
- Other (please specify) _____

28. What was your grade point average for last term?

If you are not sure, please indicate your best guess. _____ %

If you did not take courses last term, please check here. _____

29. How many courses are you currently enrolled in?

_____ courses

How many courses have you taken in the past twelve months, including those in which you are currently enrolled?

_____ courses

30. What is your student status?

Part-time

Full-time

Co-op

Other, please specify _____

31. Are you currently employed (paid work)?

YES

NO

32. If yes, on average, how many hours a week do you work for pay?

_____ hours per week

N/A

33. Are you the primary caregiver in your family?

YES

NO

Please use the following scale:

1 = Strongly Disagree

2 = Disagree

3 = Neither Agree nor Disagree

4 = Agree

5 = Strongly Agree

N/A = Not Applicable

34. If you are taking an online course, please circle the location(s) where you use a computer for this course. (Please circle all that apply).

• Home

1 2 3 4 5 N/A

• Workplace/Work Office

1 2 3 4 5 N/A

• On-campus

1	2	3	4	5	N/A
• Community					
1	2	3	4	5	N/A
• Other (please specify) _____					

Please use the following scale:

- 1 = Strongly Disagree
 2 = Disagree
 3 = Neither Agree nor Disagree
 4 = Agree
 5 = Strongly Agree
 N/A = Not Applicable

35. At home, I can use the following for study purposes: (Please circle all that apply).

• A computer					
1	2	3	4	5	N/A
• E-mail					
1	2	3	4	5	N/A
• The World Wide Web					
1	2	3	4	5	N/A
• A VCR (videocassette player)					
1	2	3	4	5	N/A
• An audio (tape) cassette player					
1	2	3	4	5	N/A

Please use the following scale:

- 1 = Strongly Disagree
 2 = Disagree
 3 = Neither Agree nor Disagree
 4 = Agree
 5 = Strongly Agree
 N/A = Not Applicable

36. There is somewhere in my community where I can go to use the following for study purposes: (Please circle all that apply).

• A computer					
1	2	3	4	5	N/A
• E-mail					
1	2	3	4	5	N/A
• The World Wide Web					
1	2	3	4	5	N/A
• A VCR (video cassette player)					
1	2	3	4	5	N/A
• An audio (tape) cassette player					
1	2	3	4	5	N/A

Thank you for your assistance!

Appendix F: Semi-structured Student Interview: Sample Questions

The project associate will ask about

-reason(s) for taking the course and whether any change since beginning the course.

-student's overall assessment of course, of the use of technology in this course, of learning in the course, and of one's self interacting with technology. The emphasis is on instructional and institutional factors that help/hinder learning.

e.g., What do you think of course? What should be changed? What did you learn? Did you learn what you wanted to, what you expected, what you think the course intended? What strengths/weaknesses does the delivery mode have? Would you use it again? Would you recommend it to others? What problems, if any, with using technology? How comfortable were you with the technology to begin with? How comfortable now? What (technical, personal, financial) support was provided? What support was missing? What costs/benefits did you experience? Were they expected or not? Is it important to be a self-directed person with good time management/good study habits to use this delivery mode effectively? Did you have those skills to start with, learn them, feel hindered without them?

-student's overall assessment of learning in the course, how his/her learning was assessed, and how assessments affected approach to study. Ask about the course requirements and how learning was assessed. This discussion may include discussion of specific tests, assignments etc., with a view to understanding how the student understood what was required.

What's important in successfully completing the course? In getting good marks? What are the assessments/tests looking for? Where do you focus your efforts? Why? What activities/preparation seems most important? What changes in assessment would improve what you learn from course?

Appendix G: Semi-structured Faculty/Staff Interview Schedule

The project associate will ask about

(a) the models and methods of course design and delivery for this particular course; e.g., Describe the development model: Is it individual faculty member(s), group of faculty members (no design/ media specialists), team that includes design/ media specialists, modified team? If modified team, describe.

What is the instructor's role, if any, in the design? Is it the sole designer, one of group of faculty members, member of a team that includes design/ media specialists, member of a modified team, or no role in design?

What is the source of funding for development, source of funding for delivery?

(b) institutional and instructional factors, including resources and support services, that influence the adult learner's response to the course and its delivery; e.g., Do the resources available to learners meet their needs? What is missing? What would improve their experience? Are there policies or procedures that are barriers? What can be done to improve the course, the delivery of the course?

Appendix H: Sample Memo - February 24, 2001

Reread GERM report and transcripts today. Noticed new things in Germ 430, i.e. Student line-ups outside the university computer labs. One student said that she "didn't really see the point of the highlighting," and she believed the CD-Rom had sound, but my other interviewees said it had no sound. The same student, whose views represented the views of the satisfied group, also said she missed the lecture format. Why had I not picked up on this in my first analysis in November, 1999?

The RMES learners were very positive about their course. They were older, employed professionals who accessed the course from home. Like the Germ 430 learners, what they appreciated the most was the flexibility. They also wanted more instructor interaction. These findings converge with the findings from Fine Arts 225.

Based on my analysis up to now, it seems that key success factors for distributed courses are:

Client groups: mature, employed adults with their own equipment which is sophisticated and at their home.

Motivation: influenced by lots of instructor interaction (instructor interaction is more highly valued than peer interaction).

Flexibility/Access: the *raison d'être* (Garrison & Anderson reference).

Organizational policies which do not impede any of the functions of the course, and which support and reward the instructor, who tend to suffer from burnout.

Instructor: valued; enhances motivation.

Materials: good, clear materials, well-designed, sequenced and visually effective.

Job relevance: enhances motivation.

Interaction: someone to bounce ideas off of (Psyc, RMES, Fine Arts)

February 24, 8:44 p.m.: After going through the data today, I drew a few different models in quick succession. Realized that there is no “true” model, only current favorites. I’ve also caught myself looking for proof of my current favorite model and ignoring other evidence when going through the data. Must be careful to be aware of this kind of bias and keep checking myself for it!