Objectives

After reading Chapter Two, you should be able to:

• describe four major inquiry paradigms that influence contemporary evaluation practice;
• describe several evaluation models that are relevant to evaluating interactive learning systems; and
• identify strengths and weaknesses of competing paradigms and models of evaluation.

Evaluation: The state-of-the-art

Contemporary conceptions of evaluation range from an absolute “measurement” perspective to a completely relativistic “constructivist” perspective (Scriven, 1993). The former is characterized by the motto: “If anything exists, it can be measured.” Advocates of the measurement perspective believe that there is an objective reality existing apart from the beliefs of those who seek to reveal its nature. Measurement proponents seek better and better instruments (e.g., tests) to calibrate the nature of that ultimate reality (Thorndike, 1971). Their search for finer and finer measures is based on the belief that it is only human weakness that blocks perception of the unchanging natural laws that govern all phenomena, from the motions of planets to the processes of human learning. Isaac Newton’s precise, mechanistic universe is the preferred model of how the world works for evaluators who emphasize the measurement perspective.

In sharp contrast, the constructivist perspective is characterized by the words of Egon Guba and Yvonna Lincoln (1989), who wrote: “it [constructivist evaluation] takes the position that evaluation outcomes are
not descriptions of the ‘way things really are’ or ‘really work,’ or some ‘true state of affairs,’ but instead represent meaningful constructions that individual actors or groups of actors form to ‘make sense’ of the situations in which they find themselves” (p. 8). Constructivists generally view measurement as a futile act because they contend that the act of measuring always affects what is being measured. Constructivists seek better and better ways of sharing understandings of the world, but ultimately view as irrelevant the prediction and control so integral to the measurement perspective. Guba and Lincoln (1989) go so far as to maintain that: “To accept the basic premises undergirding responsive constructivist evaluation is virtually to abandon hope that solutions to social problems can ever be found” (p. 47).

Fortunately, in between these two radically different perspectives, one overly optimistic about prediction and control, the other viewing these goals as irrelevant, exist other conceptions of evaluation (Patton, 1997; Shadish, Cook, & Leviton, 1991). For example, one influential evaluation authority, Michael Scriven (1993), argues that evaluations must ultimately yield judgments of what has value and worth. Scriven writes, “Evaluation is what separates astronomy from astrology, good explanations from bad ones, good experimental designs – or bridge designs – from inferior ones, good scientists, engineers and technologists from run-of-the-mill ones” (p. 30). Another evaluation “guru,” Elliot Eisner (1985), calls for the training of evaluators who will serve as connoisseurs of what is “good” in education and training.

We believe that most developers and managers of interactive learning systems fail to conduct evaluations at all, or focus on evaluating delivery
systems per se rather than the actual pedagogical or instructional design aspects of the systems being evaluated (Clark, 1994, 2001). They can hardly be blamed for this because evaluation as a form of professional practice itself has evolved (some might say devolved) into a contentious field with many competing models and several conflicting paradigms. The various ways that different evaluators use similar terms to mean different things can be especially frustrating to the uninitiated (Hopson, 2000). The proliferation of evaluation paradigms and models is confusing for people new to the field, but we believe that this complexity also enriches the options available for disciplined evaluation. It is important to review the breadth and nature of these paradigms and models, and that is the purpose of this chapter.

**Conflicting paradigms**

Although popularly used in recent years to describe everything from a change of philosophy to a change of clothing styles, “paradigm,” in the scientific sense, represents a well-established, clearly-delineated approach to conducting inquiry in a field. The status of evaluation within the context of interactive learning is hardly “well-established” or “clearly-delineated,” and yet evaluation practitioners are influenced by inquiry paradigms, whether consciously or not (Schubert & Schubert, 1990; Soltis, 1992).

According to the *International Encyclopedia of Educational Evaluation* (Walberg & Haertel, 1990), the predominant usage of “paradigm” refers to the explicit and tacit assumptions that guide inquiry within a field. In *The Structure of Scientific Revolutions*, Thomas Kuhn (1962) described how paradigms change within fields; for example, how the shift from Newtonian to quantum physics required fundamental changes in the philosophy, assumptions, theories, and methodology of research in the “purest” of sciences. According to Kuhn, the dominant mode of inquiry within a field at any given time depends upon complex interrelationships among theories, measurement assumptions, research methods, and analytical procedures. Kuhn’s “revolution” thesis holds that as a mode of inquiry is practiced over time, anomalies accumulate concerning the assumptions within a field and thus reconceptualization is required. Eventually, a new paradigm emerges.

The prospect of a new paradigm for inquiry emerging from time to time in a field is very powerful. However, within the context of interactive learning, this process is not likely to occur. Evaluation has been so infrequently and inconsistently applied that anomalies have had few opportunities to be revealed, and reconceptualization has rarely been
required. Much more effort is put into creating new methods of evaluation (usually called models) than into applying them, and thus there is no real resolution of the effectiveness of one paradigm in comparison to another. Not surprisingly, the field of evaluation is replete with alternative, some would say incompatible, paradigms and diverse models (Rogers, Hacsi, Petrosino, & Huebner, 2000; Shadish, Cook, & Leviton, 1991). In the following pages, we attempt to delineate four major paradigms related to the evaluation of interactive learning.

Analytic-Empirical-Positivist-Quantitative Paradigm

The “Analytic-Empirical-Positivist-Quantitative Paradigm” represents the most established of the paradigms that guide evaluation in education and the social sciences. The “analytic” aspect of this paradigm reflects a belief in a mechanistic, deterministic reality whereby parts can be separated from wholes, and cause and effect relationships among parts can be revealed. The “empirical” aspect refers to the goals of inquiry as being the definition, prediction, control, and explanation of physical phenomena as revealed through experience (induction) and experiments (deduction). The “positivist” aspect represents a faith in scientific progress and the perfectibility of humanity. Finally, the “quantitative” aspect stems from a reliance on measuring variables and analyzing the relationships among them with descriptive and inferential statistics.

Subscribers to the “Analytic-Empirical-Positivist-Quantitative Paradigm” believe in a separate, material reality that exists apart from the beliefs of individuals, groups, or societies. Following E. L. Thorndike and other early measurement theorists, they believe that for anything to be said to exist, it must be able to be measured. They seek to explain changes in variable aspects of reality (e.g., learning or performance) through controlled experimentation. Detachment from the phenomena under study is preferred to maintain objectivity. Mathematical analysis and statistical significance are held in the highest regard.

The “Analytic-Empirical-Positivist-Quantitative Paradigm” is the most widely used evaluation paradigm within both education and training (Kirkpatrick, 1998), and not surprisingly, it also is most often employed when evaluations of interactive learning systems are conducted. The reliance upon this paradigm stems partly from the desire to determine the comparative effectiveness of one program or product over another. Often, the first question asked about an innovative interactive learning program is: “Is this interactive learning program better than competing programs, be they other innovations or the ‘traditional’ programs or products?” Classical experimental or quasi-experimental methods are frequently the preferred method of determining the comparative superiority of one program over another (Campbell & Stanley, 1966).
The dominance of this paradigm also derives strength from the degree to which it is entrenched in the universities where many novice evaluators learn their craft. Most members of the tenured generation of educational psychology and instructional technology faculty, who dominate the teaching of evaluation as well as instructional design methods, were schooled in positivist, quantitative methods, and as a result, they tend to teach these same approaches. In addition, they may be unfamiliar with or uncomfortable with alternative paradigms. This situation may be changing as more graduate students enroll in courses focused on alternative paradigms such as qualitative methodologies and critical theory taught by faculty outside evaluation and instructional design departments.

It is common within the instructional design literature to divide evaluation into two types, formative (conducted to “improve” whatever is evaluated) and summative (conducted to “prove” the effectiveness or worth of whatever is evaluated). The application of experimental methods within the context of designing interactive learning systems is limited by several factors, especially by the time and resources usually required by this approach. Carey and Dick (1991) sum up the limited utility of experimental methods as follows:

Such experimental studies are rarely employed in summative evaluation field trials for lengthy instruction because it is almost impossible to hold constant all factors in the design, much less those in the instruction. The list of study limitations often exceeds the list of study questions. (p. 300)

Carey and Dick recommend alternatives to the experimental evaluation approach, including expert judgments and detailed analyses of the relationships among implementation and outcome variables. Heller (1996) also maintains that “it is difficult to design experiments for summative evaluation because experiments require that all aspects of the application, save one or two, are held constant” (p. 76). However, Heller recommends small-scale quasi-experimental methods as a formative evaluation strategy, especially with respect to investigating specific questions such as whether to label icons with text in a multimedia kiosk. Pearrow (2000) even suggests that experimental methods can be used within the context of testing the usability of Web sites.

**Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm**

If the “Analytic-Empirical-Positivist-Quantitative Paradigm” represents the most established of the paradigms that can guide evaluation, the “Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm” can be described as first among the up-and-coming alternatives. Interest in qualitative evaluation methodologies in education has grown steadily
during the past twenty years. One small but telling indicator is that the 1993-94 president of the American Educational Research Association (AERA), a group associated with experimental approaches to investigating education since 1915, was Elliot W. Eisner, a long-time proponent of qualitative evaluation and educational criticism (Eisner, 1977, 1985).

The “constructivist” aspect of the “Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm” reflects the belief that humans individually and collectively construct reality. The “hermeneutic” aspect of this paradigm has its original meaning in the analysis of religious texts, especially in reference to different scholarly interpretations of the Bible. More recently, hermeneutics has come to mean the analysis of curriculum in the broadest sense and instructional programs and products in a more focused sense, including attempts to expose the values underlying these phenomena (Schubert & Schubert, 1990). The “interpretivist” aspect stresses the need to put analyses in context, presenting the interpretations of many, sometimes competing, groups (often called stakeholders) interested in an evaluation’s outcomes. The “qualitative” aspect represents the emphasis on the human being as the primary evaluation instrument, rejecting the mathematical modeling of phenomena upon which the quantitative paradigm depends so heavily.

Proponents of the “Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm” have sharply divergent views about the nature of reality from those of proponents of the quantitative paradigm. They believe that “Truth is a matter of consensus among informed and sophisticated constructors, not correspondence with an objective reality” (Guba & Lincoln, 1989, p. 44). Further, they are concerned with understanding the nature of this constructed reality from multiple perspectives, emphasizing the roles of culture, gender, context, and other factors in the construction of “reality.” With regard to evaluation methodology, they have adopted many anthropological methods of inquiry, especially human observation. Not surprisingly, immersion in the context of an evaluation is highly preferred over the detachment of the classical laboratory scientist.

The emphasis on observation as an evaluation strategy is especially relevant to evaluations of interactive learning. Observational methods are among the most effective strategies for collecting formative evaluation information during the development of an interactive learning program, including highly formalized observations such as usability tests (Nielsen, 2000; Pearrow, 2000). As noted above, formative evaluation is focused on finding ways of improving a program under development. According to Flagg (1990), “Formative evaluators employ a wide range of data collection measures, drawn mainly from four categories of methods: self-report, observation, tests, and records or documents” (p. 149). Masters of observational methods for evaluating instructional programs include
the creators of “Sesame Street” at the Children’s Television Workshop (Schauble, 1990) and operators of software usability testing labs at companies such as Bell Laboratories and the IBM Corporation (Hix & Hartson, 1993; Nielsen, 1993, 2000).

Critical Theory-Neomarxist-Postmodern-Praxis Paradigm

Whereas proponents of the quantitative paradigm proclaim their objectivity and believers in the qualitative paradigm revel in their subjectivity, advocates of the “Critical Theory-Neomarxist-Postmodern-Praxis Paradigm” wear the label “social activists” with pride. Critical theorists condemn the “Analytic-Empirical-Positivist-Quantitative Paradigm” as reactionary and criticize the “Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm” as unengaged. In fact, members of the Neomarxist faction of this paradigm view themselves as the forces of liberation engaged in an on-going conflict with the powers of oppression.

The “critical theory” aspect of the “Critical Theory-Neomarxist-Postmodern-Praxis Paradigm” relates to a concern “with questions of power, control, and epistemology as social constructions with benefits to some and not to others” (Muffoletto, 1993, p. 4). The “neomarxist” aspect of this paradigm is derived from the liberation education movement led by Paulo Freire (1970) and Ivan Illich (1970). Neomarxist evaluators seek to expose the “hidden curriculum” underlying instructional technology and other educational reforms. The “postmodern” perspective questions the conception of instructional technology “as neutral or as leading inevitably to progress” (Hlynka & Yeaman, 1992, p. 2). Postmodernists deconstruct the “texts” inherent in the products and programs developed by instructional designers, seeking to reveal contradictions and the exclusion of minority interests. Deconstruction, which has its roots in literary analysis as a process of revealing the hidden meanings of texts, is a primary strategy within this paradigm. The “praxis” aspect represents a desire to abandon the search for “truth” as sought by empiricists or “understanding” as desired by interpretivists in favor of seeking “little truths which are situationally appropriate” (Anderson, 1993, p. 1).

From the perspective of the evaluation of interactive learning systems, the “Critical Theory-Neomarxist-Postmodern-Praxis Paradigm” may be difficult to conceive, much less apply. After all, why would proponents of this anti-authoritarian paradigm seek to serve an instructional design process that has its roots in systems models developed for large-scale weapons production by the military-industrial complex (Noble, 1991)? Nonetheless, evaluators should be aware of the “critical theory” perspective if only because it encourages evaluators and instructional designers to question again and again the cultural, political, and gender
assumptions underlying an instructional product or program. Skepticism and questioning are the basic tenets of this paradigm. In fact, Michel Foucault, a major figure in this movement, demands “that we question everything, including law, science, religion, and Western philosophy” (Anderson, 1993, p. 1). At the risk of stating the obvious, it is clear that not everyone will be comfortable with this paradigm.

The primary modes of inquiry for followers of this paradigm are criticism and deconstruction. For postmodernists, the basic notion of evaluation as a quasi-scientific process is secondary to the practice of criticism. Hlynka and Yeaman (1992) regard evaluation as a strategy for improving the effectiveness and efficiency of instructional design and the innovations it develops, whereas they promote criticism as having the more important role of “constantly rethinking and deconstructing our beliefs, tools, and technology” (p. 2).

Eclectic-Mixed Methods-Pragmatic Paradigm

Our last “paradigm” for evaluation cannot be found in any traditional analysis of paradigms for inquiry in education (Schubert & Schubert, 1990; Soltis, 1992). Nonetheless, it is perhaps the most important of the four paradigmatic orientations because it is the one approach capable of handling the complexity (some would say chaos) that is the hallmark of contemporary society and technology (Casti, 1994; Pascale, Millerman, & Gioja, 2000; Sedgwick, 1993). It is also the paradigm we espouse in the rest of this volume.

The concept of complexity is exceedingly relevant to evaluation within the context of instructional design. Sedgwick (1993), in an article titled “The Complexity Problem,” maintains that: “It is becoming increasingly clear that the comfort of a good fit between man and machine is largely absent from the technology of the information age” (p. 96). A similarly pessimistic observation can be made about the fit among instructors, learners, and the interactive learning programs resulting from systematic instructional design (ID). According to Leshin, Pollock, and Reigeluth (1992), the major problem with ID models is that “they have totally ignored guidance as to what makes good instruction” (p. 1). Leshin et al. (1992) also maintain that traditional ID models can be followed rigorously and still yield poor instruction.

At least part of the poor success record of systematic instructional design stems from the inadequate, simplistic evaluation methods incorporated within most ID models (Reeves, 1989). Evaluation activities are often carried out without clearly specifying the many decisions that must be made during the ID process. Ultimately, the failure of instructional innovations to transform practice in education and training can be
partially attributed to a failure to anticipate critical decisions about the design, implementation, and maintenance of interactive learning systems. The failure to incorporate evaluation focused on critical decisions is one of the reasons instructional design is under attack in some sectors (Gordon & Zemke, 2000).

The “eclectic” aspect of the “Eclectic-Mixed Methods-Pragmatic Paradigm” refers to its openness to borrowing the methods of the other three paradigms to collect information and solve a problem. The “mixed methods” aspect relates to the recognition that multiple perspectives are necessary to “triangulate” or “bracket” information and conclusions regarding complex phenomena. The “pragmatic” aspect reflects the practical orientation that, although ultimate prediction and control may never be achieved in the design and implementation of interactive learning systems for education and training, things can get better.

Proponents of the “Eclectic-Mixed Methods-Pragmatic Paradigm” rarely concern themselves with ultimate conceptions of reality, preferring to deal with the practical problems that confront them as educators and trainers. They view modes of inquiry as tools to better understanding and more effective problem solving, and they do not value one tool over another any more than a carpenter would value a hammer over a saw. They recognize that a tool is only meaningful within the context in which it is to be used. Mark, Henry, and Julnes (2000) exemplify this perspective in their conceptualization of evaluation as “sense-making.” Pragmatists accept their interconnectivity with the phenomena they seek to understand and change (Bruce & Rubin, 1992). They also recognize the weaknesses of their tools, and struggle against great odds to ensure that evaluation will affect decision making more than will politics, ignorance, intuition, habit, and prejudice. Finally, they are honest with themselves and their audiences about the tentative and probabilistic nature of the recommendations they make.

There are few indications that instructional design theorists or evaluators in general recognize the importance of an eclectic approach to evaluation within instructional design (Mark & Shotland, 1987). An exception to this trend is Johnson (1989), who describes the complex problems that instructional designers face amid a world of “chaos, connectivity, and computers” (p. 182), and argues for applying the science of chaos within instructional design. The more common approach in the instructional design literature is to fall back on traditional notions of formative and summative evaluation (Briggs, Gustafson, & Tillman, 1991; Leshin, Pollock, & Reigeluth, 1992). Some experts seek to build new tools that can support instructional designers or completely automate the instructional design process (Merrill, Li, & Jones, 1990), but how these systems will encompass evaluation is unclear.
Fortunately, a few contemporary instructional design theorists recognize the need for enhanced approaches to instructional design (Seels, 1993). For example, Gustafson (1993) calls for a major examination and renewal of the fundamentals of instructional design. Instructional systems design has been severely criticized in some sectors (Gordon & Zemke, 2000), and reform may be overdue. Regardless of how the field of instructional design evolves, better evaluation methods, especially enhanced formative designs and strategies, must play a major role in the development of interactive learning systems (Maslowski & Visscher, 1999). Indeed, we view decision-oriented evaluation as the most important aspect of the entire ID process.

From the perspective of the evaluation of interactive learning systems, we will argue throughout this volume that the “Eclectic-Mixed Methods-Pragmatic Paradigm” has the most potential for enhancing interactive learning systems and providing evidence of their effectiveness and worth. We do not wish to convey that the other paradigms lack merit, but simply that they sometimes place blinders on their proponents that a more receptive and open stance may reduce or eliminate.

**Competing models**

Whereas paradigms refer to a general orientation toward inquiry, models are much more specific and prescriptive in nature. Although some authorities trace the roots of contemporary evaluation practice back thousands of years to ancient China (Popham, 1974), most experts acknowledge the development of scores of evaluation models since the 1930s. We do not intend to provide a comprehensive review of all the evaluation models that have been developed; instead, we refer you to other sources for this information (Owen & Rogers, 1999; Shadish, Cook, & Leviton, 1991; Worthen, Sanders, & Fitzpatrick, 1996). Nevertheless, it is useful to review a few of the models that have been employed for evaluating interactive learning systems.

Instructional designers and evaluators usually concern themselves with discrete models rather than paradigms. Alkin and Ellett (1990) maintain that “Evaluation models either describe what evaluators do or prescribe what they should do” (p. 15). Prescriptive models are more specific than descriptive models with respect to procedures for planning, conducting, analyzing, and reporting evaluations. According to Alkin and Ellett, descriptive models are more general in that they describe the theories that undergird prescriptive models. In this regard, descriptive, theory-based models are akin to inquiry paradigms. There are dozens of prescriptive evaluation models, but descriptive models are relatively rare.

---

**WEB LINK**

Tylerian Objectives-Based Evaluation Model

No description of evaluation models would be complete without including the seminal contributions of Ralph W. Tyler and his many followers (Smith & Tyler, 1942; Tyler, 1949). According to Alkin and Ellett (1990), the essence of the Tylerian model is that “evaluation should judge that a program is good if, and only if, its objectives are achieved” (p. 16). Tyler changed the focus of evaluation from assessment of the inputs into an educational program to a comparison between a program’s objectives as stated and as achieved.

Tyler’s emphasis upon the importance of objectives has influenced many aspects of education, including the design of interactive learning systems. The specification of objectives is a major factor in virtually all instructional design (ID) models (Gustafson, 1991). A major benefit touted for specifying objectives during ID is that they provide the basis for the development of measurement procedures and instruments that can be used to evaluate the effectiveness of instruction. Indeed, it is difficult to imagine how ID might proceed very far without some attempt to delineate the objectives of the development effort.

Tyler’s approach is deceptively simple: Establish goals, derive specific behavioral objectives from the goals, tailor instruction and performance measures to the objectives, collect performance data on the objectives, and finally judge the effectiveness of a program based upon the learners’ achievement of the objectives. If only it were that simple, there would be no need for this book.

The first problem with Tyler’s approach is that many goals cannot be specified in terms of easily measured behavioral objectives, e.g., the goal of preparing children to be good citizens. Second, measurement of even the most behavioral of objectives is often a difficult process involving complex issues of feasibility, reliability, and validity. Third, programs often have unintended outcomes that may be more important than the stated goals of a program, either positively or negatively. Moreover, focusing on the specific objectives of a program may obscure what critical theorists call the “hidden curriculum.” Fourth, focusing on whether objectives have been attained does not address the worth of the objectives themselves (Scriven, 1974). For these and other reasons, the Tylerian evaluation model has been supplanted by other models.

Experimental Evaluation Model

The experimental (or more often, quasi-experimental) model is a widely accepted and frequently employed approach to evaluation and research within education and training. Not surprisingly, it has had, and continues to have, many proponents (Campbell & Stanley, 1966; Rossi & Free-
man, 1989; Tate, 1990). The most familiar version of this approach is represented in Figure 2.1.

![Figure 2.1](image)

**Figure 2.1.** The true control group, pretest-posttest evaluation design.

House (1991) identifies Suchman as one of the originators and strongest advocates of the experimental approach to evaluation:

>The logic of this design is foolproof. Ideally, there is no element of fallibility. Whatever differences are observed between the experimental and control groups, once the above conditions are satisfied, must be attributed to the program being evaluated. (Suchman, 1967, pp. 95-96; cited in House, 1991, p. 7)

It is difficult to argue with authorities who speak in terms of their approach being foolproof or infallible. It is also difficult to oppose an approach that is deeply entrenched in the minds and actions of so many educators. However, there is a distinct contrast between the advocacy of experimental methods by many evaluators and academics (Fitz-Gibbon & Morris, 1987; Gage, 1989; Schrag, 1992) and the rejection of these methods by contemporary philosophers of science (Casti, 1989; House, 1991; Phillips, 1987). Phillips remarks on this irony:

> New approaches to the design of evaluations of educational and social programs are being formulated that make the “true experiment” seem like a lumbering dinosaur, yet some folk persist in thinking that dinosaurs are wonderful creatures. (p. viii)

What are some of the problems with experimental and quasi-experimental approaches to evaluation? First, the control of treatment variables, as required by experimental methodologies, is impractical in most contexts where interactive learning systems are implemented. Decades of implementation studies clearly indicate that there are substantive differences between the designs of innovative programs and
products and their actual implementation (Berman & McLaughlin, 1978; Cuban, 1986). Second, the emphasis on measuring educational outcomes via tests in this approach is rarely matched by a concomitant effort to establish the reliability and validity of those measures. In addition, the tests themselves are often poor indicators of the important outcomes of instructional programs (Stake, 1990). Third, the experimental approach can only support or fail to support pre-stated hypotheses; it cannot discover unexpected effects of a product within an instructional context. Fourth, randomized experiments are extremely difficult to conduct and can be unethical in some situations. The distinguishing feature between true experimental and quasi-experimental studies is the lack of randomization in the assignment of subjects to treatments in the latter.

Not surprisingly, the vast majority of evaluations conducted with this model are “quasi-experimental,” a compromise that introduces many difficulties with respect to the analysis and interpretation of findings. As a result, evaluators operating within the experimental model frequently fall back upon designs that can be most easily managed, focus on variables that are easiest to measure, apply statistical methods without meeting the assumptions underlying their use, and draw conclusions that have little or no practical application (Schwab, 1970). Readers interested in a more extensive critique of experimental and quasi-experimental approaches to evaluation are referred to Cronbach (1980), Guba and Lincoln (1989), House (1991), and Lagemann (2000).

**Patton’s Qualitative Evaluation Model**

Although the earliest efforts to define qualitative methods for educational evaluation originated in Great Britain (Hamilton, 1976), a major contributor to this movement has been an American scholar, Michael Quinn Patton (1980, 1987, 1990, 1997). Patton (1987) stresses that qualitative methods are “particularly useful for studying variations in program implementation” (p. 13). He also emphasizes the value of qualitative methods, especially observations, if the intent of an evaluation is formative, i.e., focused on improving a program or product.

Patton (1980, 1987) describes several different methods of qualitative evaluation, including observations, case studies, interviews, and document analysis. He also stresses rigorous training for qualitative evaluators. Quantitative evaluators and researchers are often subjected to a regimen of statistics and research methodology courses before they can undertake inquiry. Patton maintains that disciplined training and careful preparation are required for qualitative evaluators as well.

Critics of the qualitative approaches advocated by Patton (1987) and others claim that they are too subjective and that the results of qualitative
studies lack generalizability (Cizek, 1995). Opponents also maintain that qualitative evaluators have a bias against the precision and accuracy of numerical data. Patton, however, claims that the numbers derived from quantitative methods are seductive because they “convey a sense of precision and accuracy even if the measurements which yielded the numbers are relatively unreliable, invalid, and meaningless” (pp. 165-66). Patton says that he is not anti-numbers, but pro-meaning.

Instructional designers conversant with the formative evaluation methods described by Flagg (1990) will find support and guidance in the writings of Patton (1980, 1987, 1990, 1997). Dick and Carey (1991) also promote the value of early one-on-one observations and field trial observations within ID projects. Reeves (1992) suggests that qualitative methods are particularly appropriate in instructional design projects within school settings where more formal, controlled methods are unfeasible. Much of the six-phase evaluation model we describe later in this book is based upon the application of a variety of qualitative methods.

Fourth Generation Evaluation Model
The constructivist aspect of the “Constructivist-Hermeneutic-Interpretivist-Qualitative Paradigm” described above is perhaps best characterized by the model that Guba and Lincoln (1989) call “Fourth Generation Evaluation.” Guba and Lincoln (1989) present seven principles that underlie their model:

- evaluation is a sociopolitical process;
- evaluation is a joint, collaborative process;
- evaluation is a teaching/learning process;
- evaluation is a continuous, recursive, and highly divergent process;
- evaluation is an emergent process;
- evaluation is a process with unpredictable outcomes; and
- evaluation is a process that creates reality.

A primary outcome of “Fourth Generation Evaluation” is “rich, thick” description based upon extended observation and careful reflection. Constructivists seek more effective ways of sharing understandings of the world, but ultimately regard the prediction and control so integral to the quantitative measurement perspective as irrelevant. As noted at the beginning of this chapter, Guba and Lincoln (1989) express a pessimistic perspective about the possibilities of evaluations actually leading to the ultimate resolution of social problems. They recommend substituting “local understanding for generalized explanation” (p. 48).
Guba and Lincoln’s (1989) gloomy (they would say realistic) outlook plays right into the hands of the critics of the Fourth Generation Evaluation model. Advocates of the quantitative, positivist paradigm who believe in the ultimate perfectibility of humanity criticize the lack of emphasis on generalizability within this model. Critical theorists who believe in social action and political commitment criticize the hopelessness of Guba and Lincoln’s perspective. Constructivists would respond to their critics that the burden of generalizability and action should be placed on the consumers of an evaluation rather than the evaluators themselves.

A common element in both Fourth Generation Evaluation as an evaluation model and instructional design as a process is the importance of negotiation. Guba and Lincoln (1989) recommend negotiation strategies for reaching consensus about the purposes, methods, and outcomes of evaluation. Similar negotiation strategies can also be used by instructional designers attempting to develop instructional innovations such as e-learning solutions and electronic performance support systems within the constraints of limited time, money, and personnel. Although hard-core positivists will almost certainly find the relativism espoused by Guba and Lincoln disconcerting, this constructivist approach has the potential to help instructional designers deal realistically with the complexities they will inevitably face in developing interactive learning systems for education and training.

**Eisner’s Art Criticism Evaluation Model**

Elliot W. Eisner, a former art teacher, coined the term “educational connoisseurship” to describe the first of two primary facets of his art-based evaluation model (1977, 1985). Just as there are connoisseurs in the worlds of literature, drama, and art, Eisner argues that education should have its own connoisseurs, individuals with refined tastes and sensitivity to educational phenomena. The second primary facet of Eisner’s approach is educational criticism, i.e., the public expression of the connoisseur’s observations and appraisals of education.

Educational criticism has three major dimensions: description, interpretation, and assessment. It is the last dimension, assessment, that distinguishes Eisner’s approach from traditional qualitative inquiry. As Barone (1990) describes, “...it is also the task of the educational critic to explicitly appraise features of the critical object according to personally held educational criteria” (p. 198, emphasis in original). The critic has the responsibility to convey his or her values to the public so that they may judge the worth of the critic’s assessment.

Traditional empirical methods are concerned with internal and external validity. Eisner (1985) has described two related processes that he calls
“structural collaboration” and “referential adequacy.” The former is concerned with the coherence, cohesiveness, and persuasiveness of the criticism. The latter is related to the insights or “psychological generalizability” of the critique. These two processes have the goals of ensuring that the reader finds the critical piece credible and that it moves the reader to some decision or action.

Educational connoisseurship and criticism, as proposed by Eisner (1977, 1985), have not been implemented widely. Opponents of educational criticism claim that it is too subjective. Proponents of this model respond that positivists are also subjective when they formulate research questions, choose instruments, and interpret data, but that their subjectivity is hidden. Educational critics, on the other hand, have an explicit responsibility to disclose the values underlying their subjectivity; they do not suffer from the illusion of objectivity (Barone, 1990).

Another reason for the limited application of Eisner’s ideas may be the difficulty in training people to be connoisseurs and a lack of venues for expression of educational criticism, especially in established research journals. Eisner (1985) suggests that anyone involved in education (teacher, student, parent, or administrator) has the right and responsibility to be a critic, but that certain people can be trained in the approach and eventually establish a reputation for the refined tastes and appreciation that are indications of authentic connoisseurship.

Can designers and users of interactive learning systems apply Eisner’s ideas to the challenges they face? Certainly, the frequent use of expert reviewers as an evaluation strategy within the instructional design process has something in common with educational connoisseurship and criticism. When employing expert reviewers during formative or summative evaluation, developers of interactive learning systems should seek to clarify the underlying values that influence the experts’ judgments. In addition, those interested in the creative aspects of multimedia for interactive learning will find the art-based concepts and ideas that Eisner adopts of considerable interest.

**Postmodern Evaluation Model**

According to Jencks (1989), postmodernism advocates multiple, often contradictory ways of knowing in contemporary society. The “post” in postmodernism reflects a belief that we are existing in a time beyond the failed modernity of the industrial age. “Modern” is meant pejoratively to refer to the misplaced faith in science, technology, and progress that have left the world with the threat of nuclear holocaust, environmental destruction, and global racism and intolerance (Hlynka & Yeaman, 1992). Hlynka and Belland (1991) present multiple perspectives on
postmodernism and related evaluative perspectives such as critical theory. They criticize the field of educational technology for overemphasizing modern technologies and positivist modes of inquiry. They point out that educational technology can be viewed as a series of failed innovations including motion pictures, television, programmed instruction, instructional systems design, computer-based instruction, and intelligent tutoring systems. They recommend the postmodern perspective as an approach to revealing the political agendas hidden in each of these “innovations.”

Hlynka and Yeaman (1992) describe how to be a postmodernist:

1. Consider concepts, ideas and objects as texts. Textual meanings are open to interpretation.
2. Look for binary oppositions in those texts. Some usual oppositions are good/bad, progress/tradition, science/myth, love/hate, man/woman, and truth/fiction.
3. “Deconstruct” the text by showing how the oppositions are not necessarily true.
4. Identify texts which are absent, groups who are not represented and omissions, which may or may not be deliberate, but are important. (pp. 1-2)

What are the lessons of the postmodernist perspective for designers of interactive learning systems? Basically, instructional designers should not automatically assume that their systems models and instructional technologies are the best methods for establishing conditions for teaching and learning. Further, they should constantly examine and re-examine their motives and methods to ensure that minority perspectives are included. They should attend to critics of educational technology such as Cuban (1986) and Postman (1995). Even reputed “cranks” such as Clifford Stoll (2000) should not be ignored because, in the midst of the hyperbole, an important idea may surface occasionally. In addition, designers of interactive learning systems should invite alternative views that can be used to rethink and deconstruct the programs and products they develop (Cuban, 2001; Yeaman, 1994).

Not surprisingly, critics of the postmodernist model see it as anti-technology, anti-progress, and anti-science. Postmodernists respond that positivism and science have had their chance to perfect the world and have failed miserably. Postmodernists seek to empower the disenfranchised in contemporary society, especially female, third-world, and non-white interests (Anderson, 1994). It is difficult to be postmodernist within the context of instructional design and evaluation because postmodernism largely rejects such systems-oriented modes of development and inquiry. Instructional systems design (ISD) is criticized as a tool of
positivists who hold onto the false hope of linear progress. Further, criticism is valued over evaluation because of its emphasis on identifying “dysfunctions as well as functions” (Hlynka & Yeaman, 1992, p. 2). Although the incongruency between instructional design and postmodernism is certainly problematic, there is some value in the postmodernist perspective as a method of checking interactive learning systems for aspects that may be racist, sexist, and/or culturally insensitive.

**Stake’s Responsive Evaluation Model**

Upon first consideration, Stake’s (1980) Responsive Evaluation Model might appear to fit within the qualitative paradigm. However, according to Stake (1990), this model does not adhere to one methodology: “There is a common misunderstanding that responsive evaluation requires naturalistic inquiry or qualitative research” (p. 76). Stake claims that the methods for responsive evaluation are negotiated by the “stakeholders” in the evaluation. Sometimes, they will be naturalistic or phenomenological; other times they will be quantitative or goal-oriented. More often, responsive evaluators will attempt to “triangulate” issues and findings using multiple methods.

According to Stake (1990), the key distinction of this model is “that of emphasizing the issues, language, contexts, and standards of stakeholders” (p. 77). Stakeholders are defined as the diverse people who are affected by a program, including administrators, teachers, students, parents, community members, program developers, and representatives of funding sources. The evaluators themselves are also viewed as critical stakeholders, and their judgments are given weight equal to the perspectives of others in evaluation reports. However, the evaluators must endeavor to expose the subjectivity of their own judgments as rigorously as they do those of other stakeholders.

Another essential element of responsive evaluation is the continuous nature of observations and reporting. Responsive evaluators begin to provide feedback to stakeholders as early as possible during an evaluation, fully expecting to please some people and displease others. The questions and methods of evaluation are modified throughout the evaluation in response to the evolving concerns of the stakeholders.

Critics of the Responsive Evaluation Model argue that “too much attention is given to subjective data, for example, the testimony of participants” (Stake, 1990, p. 76). Stake defends the approach by maintaining that subjectivity is inherent in any type of observation or measurement. He claims that responsive evaluators endeavor to expose the origins of their subjectivity, whereas other types of evaluators may disguise their subjectivity by using so-called objective tests and experimental designs.
Responsive evaluation can also be criticized because it requires a level of trust and cooperation within an evaluation context that may be unrealistic, especially when the financial or political stakes are high.

The utility of responsive evaluation within the context of design may rest more in the values espoused by Stake (1980, 1990) than in specific methodological aspects of the model. Responsiveness to the multiple audiences likely to be affected by an instructional program or product under development should be a key ingredient of the analysis phase of any ID model. This responsiveness is an ideal rarely attained in many ID projects, and attention to the ideas and values inherent in Stake’s Responsive Evaluation Model may improve the degree to which minority issues are considered by designers of interactive learning systems.

Multiple Methods Evaluation Model

In 1987, the American Evaluation Association published an important volume in the New Directions for Program Evaluation Series called *Multiple Methods in Program Evaluation* (Mark & Shotland, 1987). What Mark and Shotland and their co-authors present is not so much a unitary model for evaluation as a set of guidelines for applying an eclectic approach to evaluating instructional inputs, processes, and outcomes. They maintain that the primary benefits of “multiple methods” are addressing different but complementary questions in an evaluation and reducing the misplaced belief in the certainty assumed to be associated with traditional experimental methods.

Using multiple methods is not just a simple matter of two or more methods being better than one. Two poorly designed and sloppily conducted evaluation strategies will yield no better picture of the findings than one poor study. Mark and Shotland (1987) maintain that multiple methods are only appropriate when they are chosen for a particular purpose, such as investigating a particularly complex program that cannot be adequately assessed with a single method.

Key issues in Mark and Shotland’s (1987) approach are “triangulation” and “bracketing.” Triangulation involves using multiple measures to converge on a more accurate estimate of the “true” value of a variable. For example, if an instructional designer needs to determine the motivation of trainees to engage in Web-based learning, a questionnaire, an interview protocol, and a personality scale might be used in an effort to gain a more precise assessment of their motivation. The triangulation approach assumes that errors in one type of measure are canceled out by errors in another type. According to Mark and Shotland, this is an optimistic view, especially in education and other social sciences.
Bracketing, on the other hand, involves using multiple measures to provide a range of estimates of the value of a variable. This is a subtle but important difference. The instructional designer interested in student motivation might use the same three measures (questionnaire, interview, and personality scale), but the results would be examined differently. Whereas from the triangulation perspective, the designer would assume convergence of the measures toward a more accurate assessment of motivation, the instructional designer employing a bracketing perspective would interpret differences among the measures as reflecting a range of motivation. The data might be used to make an interactive learning system more responsive to learners with different sources and levels of motivation. Mark and Shotland (1987) maintain that the bracketing approach is more practical in evaluation because of the complexity and uncertainty involved in education and other social science contexts.

Probably the most valid criticism of the multiple measures approach is that it is so complex in terms of design, implementation, analysis, and interpretation that few evaluators or instructional designers will be able to apply the approach without the assistance of a team of methodologists and statisticians. However, the benefits of considering the approach and employing aspects of it are considerable. Evaluations focused on the effectiveness of instructional innovations in comparison to other programs have often led to findings of no significant differences (Clark, 1992). Although methodological issues may have invalidated many of these studies, there is also a strong possibility that a single weak outcome measure was used. Bracketing the outcomes of instructional innovations with multiple methods may yield a more realistic portrayal of the outcomes of these systems. Increased emphasis on accountability for the large investments being made in learning technologies in schools and business demand that evaluators consider alternative methods, even when they are complex and challenging.

Summary

This chapter has reviewed the four major paradigms that influence evaluation practice today. In addition, brief descriptions of several noteworthy evaluation models were included. We expressed a preference for an eclectic paradigm, and we are convinced that no single existing evaluation model is sufficiently comprehensive to guide the effective and efficient evaluation of interactive learning systems. In the next chapter, we present our analysis of the functions of evaluation as they relate to the processes of designing and implementing interactive learning.
References


Interactive Learning Systems Evaluation


