Journal of Computing in Higher Education
Spring 1991, Vol. 2(2), 84-113.

Ten Commandments for the Evaluation of Interactive MultiMedia in Higher Education

Thomas C. Reeves
Department of Instructional Technology
The University of Georgia

ABSTRACT

A SET OF GUIDELINES for redirecting evaluation and research involving interactive multimedia (IMM) in higher education are presented in the form of "Ten Commandments." Each commandment is "illuminated" with anecdotes and stories to illustrate its importance and application. In light of the complexity involved in human learning via IMM and the politics of higher education, the commandments stress descriptive approaches to research and evaluation, including "modeling" methods that integrate quantitative and qualitative data.

INTRODUCTION

EVALUATION has many definitions. For example, evaluation has been defined as: the determination of the degree to which objectives have been attained by a program (Tyler, 1942); measurement of the critical components of an instructional program or product (cf., National Study of Secondary School Evaluation, 1960); comparison of the effects of competing programs (Campbell and Stanley, 1963); judgment of a program's worth (Scriven, 1967); description of a

program's inputs, processes, and outcomes (Stake, 1967); accounting for the use of project resources (Roueche and Herrscher, 1973); the ethnographic investigation of a program (Guba and Lincoln, 1981); and the critical analysis of a program's quality (Eisner, 1985). I define evaluation as the process of providing information to enlighten decision-making that will improve the quality of life. Although some people often feel threatened by evaluation, none of the definitions cited above imply any direct threat to the program, product, and/or people being evaluated.

After a decade of evaluations of "interactive multimedia (IMM)" such as interactive videodisc (IVD) (Reeves, Brandt, & Marlino, 1988), my colleagues and I have concluded that most of the reported studies do little to promote the use of IMM in higher education or any other context. A majority of these studies have been media comparison studies seeking to find differences in effectiveness between IMM and another method of instruction (sometimes a different form of instructional technology, or more often, the mythical "traditional classroom instruction"). Education, as a field, has witnessed a parade of comparative studies during the past sixty years, involving every conceivable instructional medium, e.g., educational films, programmed instruction, instructional television, teaching machines, computer-based instruction, and now various forms of IMM. Many of these studies show "no significant differences" in effectiveness among the media under comparison.

As a case in point, consider the evaluation report, "The Instructional Effectiveness of Interactive Video Versus Computer Assisted Instruction" (Tiene, Evans, Milheim, Callahan, & Buck, 1989) published in a new journal devoted to IMM called *Interact*. The authors reported no significant differences between two instructional treatments, i.e., college students using computer-assisted instruction (CAI) to learn basic photography skills were just as successful as those using interactive video (IV). This hardly seems surprising given the plethora of comparative media studies reporting similar findings (cf., Bayard-White, 1985; DeBloois, 1988; Schroeder, 1982). And yet anyone considering the application of IMM in higher education would find little support in this evaluation for a funding proposal or guidance for program design.

There have been numerous publications describing the prob-

lems with media comparison studies (cf., Clark, 1983; Hoban, 1958; Phillips, 1980; Reeves, 1986; Sanders, 1981), and indeed, some experts have begun to call into question the very assumption that human behavior can be predicted (Cronbach, 1982; Cziko, 1989), an assumption upon which most empirical research in education is based. In my opinion, educational research, as presently taught in graduate schools of education and published in the majority of educational journals, is a bogus enterprise, but the criticisms supporting this view will not be repeated here. Instead of extending the critique of educational research, the purpose of this paper is more constructive, viz., to present a set of guidelines which can be used to redirect evaluation and research involving IMM so that evaluation will be more of an impetus than an impediment to its development.

THE FIRST COMMANDMENT

DESCRIPTION SHALL BE YOUR FIRST AND ONLY GOD SINCE IT WILL BE THE PRESERVE OF COMPLEXITY.

A NEW SCIENCE traditionally passes through a series of stages which might be described as description, prediction, control, and explanation. First, scientists must describe the phenomenon under investigation, name its various components, and define their interrelationships. Second, they seek to predict phenomena, often in the form of "if - then" statements. Third, they would like to be able to control or modify the phenomena. And lastly, they desire to explain the underlying causes of the phenomena. Of course, this is not a strict hierarchical continuum of scientific goals, and each science has its own pattern of success with respect to the four stages.

For example, the science of meteorology has been fairly successful in describing weather phenomena, and even explaining the underlying causes of specific phenomena such as drought or blizzards. Yet, as we have all experienced, meteorologists enjoy only limited success in predicting the weather and almost no success in controlling it. In Lorenz' (1963, 1979) judgment, meteorology will never enjoy the predic-

tability of other sciences because of the phenomenon of chaos, a sensitivity of certain phenomenon (such as storms) to small fluctuations in initial conditions.

In contrast, chemistry has developed clear descriptions of the various chemical elements, and whole industries are built upon the power of chemists to predict and control the interactions among various chemicals. Their success in the realm of ultimate understanding and explanation, however, has been less successful, and several competing theories exist which try to link chemical phenomena to subatomic phenomena.

In my opinion, education as a science, and most particularly, instructional technology, suffers from a lack of fundamental work at the description level. Too often educational phenomena are given a name, but there is insufficient specification of what the name means. The Tiene et al. (1988) media comparison study cited above claimed to be comparing computer-assisted instruction (CAI) with interactive video (IV), but what do these terms really mean? How much alike or dissimilar was the CAI in this program to other examples of CAI? What were the similarities and differences between the so called IV in this study and other IV programs? Missing in this and similar studies is any effort to describe instructional treatments in terms of common dimensions or characteristics. What was the degree of content overlap between the treatments? To what degrees did the different treatments provide well established instructional events such as gaining attention, informing learners of objectives, recalling previous knowledge and skills, eliciting performance, and providing meaningful feedback?

In an effort to emulate older and more respected sciences such as physics and chemistry, educational researchers and evaluators have often regarded their treatments as distinct, cohesive wholes similar to a physical property or a chemical element. Unfortunately, instructional treatments lack this cohesiveness, and the types of experimental and quasi-experimental designs used in most comparative evaluations cannot reflect the degrees of covergence and divergence among instructional treatments. Although any number of evaluations may claim to have investigated a common phenomenon, e.g., interactive video, the truth is that each IV program is an extremely complex entity composed of many diverse instructional dimensions (Cooley & Lohnes, 1976).

Ironically, physicists, from whom educational researchers and evaluators have adopted methodologies and analytic techniques, have recently had to deal with much more complexity at the subatomic "quantum" level than they did with the phenomena and laws of classical Newtonian physics. For a fascinating account of how the perspectives of scientists in physics and other fields are changing, read *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity* by Heinz R. Pagels (1988). Pagels' work has profound implications for how research and evaluation of instructional phenomena might be conducted in the future through the use of modeling and computer analysis.

The proponents of comparative media studies who draw their inspiration from the research methods of classical physics may do well to reexamine the methodological implications of contemporary quantum physics research as described by Popper (1982), Herbert (1985), and others. These implications indicate that the traditional sources of scientific knowledge, first, theory construction, and second, experimentation, may have exhausted their complementary potential in fields of enormous complexity such as subatomic particles, global economics, and learning. Pagels (1988) describes a "third branch" of science, computational modeling, as joining the other two as an important and intellectually respectable method of inquiry.

> One of the ways that future science will progress is by a combination of precise observations of actual systems followed by computer modeling of those systems. This differs from the traditional notion of experimentation in which one actively alters the conditions of the actual system to try and determine what is going on. For many actual natural systems, such as the interior of stars, one cannot even do experiments, and computer modeling is the only route one can take. Likewise, in the social and psychological sciences, one cannot in many instances do experiments, for practical or ethical considerations, and once again computer modeling offers a powerful new method to see what is going on. Computer modeling is a new way to do "experiments." (pp. 43-44)

Pagels goes on to describe the fascinating computational and modeling work being undertaken in a variety of fields, which he calls "the sciences of complexity," e.g., quantum physics, cognitive science, evolutionary biology, and international econometrics. The essence of Pagels' thesis is that neither the human mind nor the experimental method can account for a sufficient number of variables in real world systems, and that the computational power of the computer is necessary in order to find meaning in the "deterministic chaos" of most natural phenomena.

What are the implications of the new physics for evaluation of IMM? I believe that the complexity and "deterministic chaos" involved in any type of human learning demand that we use evaluation as a tool for describing phenomena. Instead of hiding the complexity involved in a learning context such as the use of IMM, we should embrace the complexity and attempt to see it from many different perspectives. Instead of mindlessly applying classical experimental methods to compare the outcomes of IMM with other instructional treatments, we must adapt a multi-faceted approach to evaluation including the conduct of intensive case studies (Stake, 1978) and the application of computer modeling (Pagels, 1988). The basic message of this first commandment is that description, as the first stage of any type of scientific inquiry, has been ignored in education, and that we must develop far greater skill at this stage before we can move onto the stages of prediction, control, and explanation.

As an application of this first commandment, my colleagues (faculty and graduate students) and I have undertaken a series of participant observations of the use of IMM as a training vehicle for Apple Computer, Inc. A course called "Macintosh Fundamentals and Beyond (MFB)" has been developed by Apple to provide initial training in the use of the Macintosh microcomputer to all new Apple employees and dealers. The innovative MFB course which employs interactive videodisc training (delivered via HyperCard stacks encoded on a CD-ROM disc combined with laser videodiscs) replaced a traditional training course (delivered via two days of leader-led training during which all students were marched in lock-step through a series of computer exercises).

Initial discussions of a design for an effectiveness evaluation for

the MFB course included the suggestion that the new course should be compared with the old course using a quasi-experimental design. Fortunately, we were successful in convincing the Apple decision-makers of the futility of this type of evaluation approach. Further, we proposed and they accepted an evaluation approach emphasizing what Geertz (1973) calls "rich, thick" description.

Why is "thick description" recommended over an approach aimed at finding statistically significant effectiveness differences between two training approaches? Think about what Apple would do with the three possible outcomes of a comparative study. If the results showed that the IMM treatment outperformed the old course, the developers would get a pat on the back, but they would have little basis for understanding what worked and what did not, nor would they have much information about how to improve their next IMM training program. If the results showed that old training outperformed the IMM method, the cynics would smirk, and the decision-makers would be faced with the dilemma of choosing among several options including returning to the old training program, revising the IMM program, or attempting the development of a new approach. Finally, if the most likely result of "no significant differences" was found, no one would have a basis for decision-making (except perhaps to fire the evaluator).

The descriptive approach, on the other hand, will provide Apple with the information needed to "evolve" to a higher level of design and implementation of IMM for training. The word "evolve" is used because it seems clear that no single evaluation or research study can provide the "deterministic ingredients" or "laws" upon which to base future design and development. Instead, descriptive case studies can provide the fertile ground for discovering hints and developing creative ideas which can gradually improve the design and ultimate effectiveness of instructional treatments (Gruber, 1985).

THE SECOND COMMANDMENT

EMBRACE COMPLEXITY, DESCRIBING IT IN MANY TONGUES, AND YIELD NOT TO THE TEMPTATION TO OVERSIMPLIFY.

IT HAS BEEN MY FORTUNE to be called in as an evaluation consultant on a number of projects involving the application of IMM, primarily IVD, in a wide variety of contexts, including higher education, medical education, military training, and business and industry. Invariably, the main question posed for the evaluation is whether the instructional innovation out-performed some other instructional method, usually the education or training method previously used by the client. It is normal behavior for people, who have invested time, money, and often themselves in creating a new instructional program, to want to know if their investments have paid off. They desire a fast and simple answer to the question, "Did the students using the instructional innovation learn more?"

However, answering that question in any meaningful way is rarely fast and never simple. Consider the case of the Emergency Medical Conditions (EMC) IVD case simulations, which were funded by the U.S. Naval Health Science Education and Training Command (NHSETC) as part of their Computer-Assisted Medical Interactive Videodisc System (CAMIS) program and are currently in use at the U.S. Navy Hospital Corps "A" Schools in Great Lakes, Illinois, and San Diego, California. The program provides realistic simulations of emergency life support activities such as CPR and insulin shock. Naturally, the Navy wanted to know whether it was an effective training program; moreover, they wanted to know whether it was more effective than the existing training program which consisted solely of classroom instruction — what the military often calls "platform training." They contracted with me and one of my doctoral students, Mary Marlino, to evaluate the program and give them the answer.

That sounds simple enough. We had only to randomly assign some classes of students to use the IVD simulations and other classes to the traditional training, test them all, and analyze the results. If the students using IVD outperformed the students in the traditional classes,

the innovation was a success. And that is exactly what we did. A total of 448 students used the IVD materials in addition to the regular training, and 482 completed only the regular training. The results of the standardized test indicated that the test scores of IVD students were higher than those of the traditional training students, and that the difference was statistically significant (Reeves & Marlino, 1989). The Navy was very pleased with the results. Most likely, the statistics will be cited at budget hearings, and word will spread that IVD outperformed traditional training methods in teaching emergency medical care.

But when I think about it, the deceptively simple surface issues are actually extremely complex. First, I have a lot of questions about the reliability and validity of the standardized test used in this evaluation. And even if the test is reliable and valid, I have doubts about the educational as opposed to statistical significance of the four point difference (85% versus 81%) between the two groups. What new knowledge, skills, or attitudes do those four points represent? More critically, does that four point difference represent a valuable outcome of the four hours of additional training the IVD students experienced? Would not any type of instruction lasting four hours increase the test scores by a similar amount?

Please do not misunderstand. I think the EMC simulations provide very effective training, but that their effectiveness is much more complex than indicated by the evaluation results. It turns out that the EMC simulations were never designed to be part of the regular training program, but were designed as remedial materials for students failing the examination of EMC skills. We collected some evidence that the materials do have a remedial effect for less qualified students, and I would like to reanalyze the data in conjunction with a profile of student entering abilities. It also turns out that on average students only completed less than 25% of the simulations because of time and delivery system limitations. What would be the effects of completing all the case simulations?

This brings forth the question of the choice of outcome measures. The ultimate outcome measure in this context would be to assess the students' performance in the handling of actual medical emergencies. Practical and ethical reasons prohibit this, but certainly better measures than multiple choice tests could be devised, including the use of

moulage exercises or additional IVD simulations. In the initial evaluation meetings, these and other strategies for dealing with the complexities involved in this situation were discussed. But tradition and expediency carried the day, and because we thought getting to do even an inadequate evaluation was better than doing no evaluation, we did it. Fortunately, the results were positive, and it appears that we will have the opportunity to pursue some of the aforementioned complexities in future studies. All too often, however, inadequate evaluation methods, not only hide the complexities involved in human learning, they also report dissatisfactory results. The clients are displeased, the evaluators are dismissed, and the IMM being evaluated is given a "bum rap."

The essence of this second commandment then is to confront complexity head on and not give into the temptations to oversimplify the situation. Obviously, I violated this commandment with respect to the CAMIS evaluation, but I hope to get the opportunity to atone for this transgression. Of course, any evaluation must simplify to a certain extent. Complexity can be defined as a state which lies between complete order and utter chaos (Gleick, 1987). In a complex learning situation, scores, probably hundreds, of significant variables are involved. Whatever the number, it certainly exceeds the 7 ± 2 reported by psychologists to be the limits of human comprehension at any one time. The theoretical and experimental approaches of traditional science usually remain within the bounds of human attention. For example, classical experimentation involves controlling several variables, usually through random assignment of subjects, and manipulating one variable, and analyzing the results. All other variance is assumed to be taken care of by the randomization or attributed to the ubiquitous "error variance." As described above, this paradigm for research and evaluation has largely been more of an impediment than an impetus to the development of IMM.

What is the alternative to this oversimplification of complexity. Perhaps the most promising avenue to understanding the impact of IMM is computer modeling of the type described by Pagels (1988). The best that comparative experimental studies can hope for is the finding that one instructional treatment is more or less effective than another. Computer modeling, on the other hand, can be used to relate instructional treatment inputs and processes to outcomes in an effort to "ex-

plain" the effects of the IMM (Borich & Jemelka, 1982; Wang & Walberg, 1983). I believe that such models can be used to derive prescriptive principles of program design and implementation to guide the future development of IMM.

Computer modeling can take many forms (Dewdney, 1988), and much theoretical and statistical work needs to be done before this approach becomes as accepted as experimental methods. One approach, "program modeling," involves structured decomposition (the breaking of an instructional program into its component parts from the broadest level of global conceptualization to the lowest level of generality at which meaningful decisions can be made) (Borich & Jemelka, 1982). Subsequently, program components as well as input and output measures relevant to the educational context can be analyzed using statistical procedures such as commonality analysis (Kenny, 1979). There are currently both technical and conceptual problems associated with this type of "causal" program modeling, but it should be investigated if only because other methods of evaluation have proven so deficient.

Another approach to computer modeling, currently being investigated at The University of Georgia and the U.S. Air Force Academy, involves adapting existing theories of instruction (Carroll, 1963; Gagné, 1985) into a model tailored to interactive video for second language instruction (Marlino, 1989). Extensive measurements of each element in the model will be subsequently analyzed using a variety of path analysis and "soft modeling" approaches (Cooley & Lohnes, 1976). The goals of this analysis are to test the overall predictability of the model and to estimate the relative instructional effectiveness of each of the model's elements.

The difficulty of estimating the impact of IMM can be compared to the difficulty involved in measuring the development of black holes in space (Hawking, 1988). In both contexts, direct measurement is elusive. Astronomers are using computer models to predict where black holes (which are invisible) are and what effect they have on surrounding celestial bodies. They enter what data can be collected into computer models and gradually improve their understanding of these mysterious phenomena. Similarly, educators are advised to construct models of the effective dimensions of IMM such as interactive video, collect relevant

data, analyze the data with computer modeling methods, and thus improve our understanding of effective instructional dimensions "bit by bit." The uniquely powerful data collection capabilities of IMM lend themselves particularly well to the conduct of computational modeling. I believe that computer modeling may provide us with the new "tongues" we need to preserve complexity and advance the science and art of IMM.

THE THIRD COMMANDMENT

RENDER JUDGMENT WITH GREAT CARE. WHEREAS DESCRIPTION PRESERVES COMPLEXITY, JUDGMENT FORCES DECISIONS OF ACCEPTANCE OR REJECTION.

THERE IS A CONTINUING DEBATE in evaluation circles about the role of judgment in evaluation. Some authorities maintain that judgment is a major function of the evaluator (cf., Scriven, 1973), whereas others maintain that evaluators must refrain from expressing judgments, serving instead to provide the information with which others can make more informed decisions (cf., Stufflebeam, 1983). I lean toward the latter, believing that I have a much better chance of having a beneficial impact if I withhold my personal judgments and help others to express theirs.

I learned this lesson the hard way. As an apprentice evaluator working in the Center for Instructional Development (CID) at Syracuse University, I was responsible for evaluating a workbook series that had been developed to help disadvantaged students as part of the Higher Education Opportunity Program (HEOP). When I first looked at the materials in the workbook, they looked suspiciously familiar, and upon further investigation, I found that much of the material had been copied from elementary workbooks of the type found in a dime store.

Not long after this discovery, I was in the office of the HEOP director, reporting the results of the evaluation which were not good. As I went down the list of deficiencies in the workbooks, the Director's face became darker and darker. Finally, when I expressed my doubts

that the materials were original and that they looked like cheap copies of dime store materials, he burst out into an angry torrent of questions and accusations. "Don't you know that these materials have won awards?" he cried. He went on to inform me in no uncertain words, that my judgment of the materials was misinformed and that he was dismissing the evaluation. I should have known I was in trouble when upon entering his office, I noticed that he had the covers of the workbooks displayed in beautiful frames on the walls. The covers were nicely designed, and I later found out that the materials actually had won awards, for graphic design of the covers.

Thus, I formed the opinion that evaluators generally must keep their judgments to themselves, and that if they are asked to express their judgments, they should make a clear demarcation between their personal judgment and the information they are providing to improve the decision making process of others. For example, during the conduct of a beta test of an interactive videodisc training course for Apple Computer, I was asked to express my judgment of the various outcomes of the evaluation. In the final report, I ended each section with a clearly marked subsection called the "Evaluator's Interpretation." Underneath each of those subsections, I left a much larger blank space labeled. "My Own Interpretation." I used this simple device to encourage the reader to view my judgment for what it was, one person's interpretation, and that it was the reader's responsibility to form an independent interpretation of the data.

THE FOURTH COMMANDMENT

SEEK NOT TRUTH, BUT LOOK FOR OPPORTUNITIES TO IMPROVE THE LOT OF THOSE AROUND YOU. ULTIMATELY, THERE IS NO TRUTH, ONLY BETTER AND BETTER UNDERSTANDING.

IN SCIENCE, there is no truth, only theory. And to the extent that evaluation is scientific, it must develop a similar stance toward truth. Evaluation must result in reasonable hypotheses and supportable

theories, but not truth. Pagels (1988) offers an interesting analysis of scientific theories which he claims must be "logically coherent, universal, and vulnerable to destruction." To be logically coherent, a theory must conform to the rules of formal logic. In fact, many scientific theories, e.g., quantum physics, are rooted in formal mathematics. If a theory is not internally and logically consistent, one can "prove" anything with it and thus it is useless.

To be universal, a scientific theory must apply to everything within the realm of its context, e.g., cellular biology. A theoretical principle about the properties of membranes in a cell must be true for all cells of that type. Similarly, the "proofs" of a scientific theory must be true of everyone, regardless of their race, creed, color, or national origin. Religious, political, or cultural theories do not require this universality.

Finally, a scientific theory must contain the seeds of its own destruction. The logic of science prohibits direct proof of a theory. Instead, a theory is either supported by experimental results or it is not. As long as study after study supports a theory, the theory is accepted by the scientific community. When studies fail to replicate the findings reported by other scientists, as was recently the case in the much heralded and subsequently ridiculed findings of "table top fusion," a theory is banished to the realm of fantasy. This acceptance and rejection process is based upon judgment and consensus, not strict logic. Ironically, the ultimate goal of any science is to conduct an experiment which will not support the theory and thus destroy the consensus and force a reconstruction. Historically, this has often been traumatic for the scientists involved, and sometimes for whole societies. The ongoing controversy in this country between the creationists and the evolutionists is one such example. For another example, I recommend *Nobel Dreams: Power, Deceit, and the Ultimate Experiment* by Gary Taubes (1986) which provides an eyewitness account of the viciousness sometimes involved in scientific inquiry.

What does all this have to do with evaluation of IMM? First, as the commandment states, don't go looking for truth. No less an evaluation authority than Lee Cronbach has concluded that perhaps the best evaluation can hope for is to help people think more clearly about alternatives. In recent years, Cronbach (1982), once a strict empiricist, has argued eloquently for increased external validity in evaluation, even to

the violation of internal validity. The scientific side of evaluation stresses control and randomization to assure internal validity, but I share Cronbach's concern that this often prohibits the generalization of evaluation findings to any meaningful context. Generally, I opt to forsake experimental methodologies in favor of any creative way of revealing information which can illuminate decision-making.

For example, in a recent evaluation of an IVD medical simulation (Reeves, Marlino, & Henderson, 1988) my colleagues and I videotaped pairs of military physicians going through a program designed to teach them how to handle battlefield trauma. The "Acute Combat Trauma Life Support" program is quite intense, and the interactions between physicians attempting to solve the case are fascinating. Viewing those tapes provided more information about the effectiveness of the simulations than any test we could devise.

An evaluator must also be willing to subject his/her findings and conclusions to alternative explanations. It has always seemed ironic to me that whenever evaluators report "no significant differences," they expend much energy in generating alternative explanations of why the desired results were not attained, but when significant results are reported, few attempts to explore rival hypotheses are made. I am not arguing for a completely relativistic stance here. Instead, I am recommending that the findings of evaluation be exposed to as wide an audience as possible and subjected to as many plausible explanations that are logically supportable.

THE FIFTH COMMANDMENT

KEEP ALWAYS BEFORE YOU THE IMAGE OF MULTIPLE PUBLICS, READY TO EAT YOU ALIVE. EVALUATING IS A POLITICAL ACTIVITY.

I LEARNED the awful truth of this commandment on my very first job as an evaluator. As a new graduate student in the Area of Instructional Technology at Syracuse University, I was assigned to an evaluation team charged with assessing the effectiveness of an "in-

novative" math curriculum. The curriculum had been developed by a well known manufacturer of hand-held calculators (which were all the rage in the early seventies) and a distinguished science laboratory. Not too surprisingly, a major feature of the curriculum was the calculator which was used together with game boards, plastic dinosaurs, and workbooks to teach mathematics to first and second graders.

Our team spent a year evaluating the curriculum in several school districts in suburban Syracuse, and I can remember spending hours on the floor observing children using the curricular materials. Alas, the results of the evaluation were quite negative, and we recommended that the manufacturer "go back to the drawing board" as it were. We submitted our evaluation report in the spring, and in the early fall, we were surprised to see two-page full color advertisements for the curriculum appearing in all the major education journals. The ads featured pictures of children on the floor using the calculators together with the dinosaurs. The real shock came when we saw that at the bottom of the page were printed the following words: "This curriculum has been evaluated by Syracuse University."

Of course, what the advertisement said was true, we had evaluated the materials. What the ad did not say, however, bordered on criminal negligence, i.e., that our evaluation had concluded that the materials were ineffective. Ever since that first lesson, I have tried to expose my evaluation results to as wide an audience as possible in the belief that the more people who know about the information, the less likely the information is going to be misused. In the real world, this may involve some tough negotiation with clients who want to control the dissemination of evaluation results. One strategy is to identify as many relevant audiences as possible for the particular evaluation, and to get them involved in the evaluation to an extent that the results will have to be shared with them.

THE SIXTH COMMANDMENT

THOU SHALT NOT CONFUSE TESTING WITH EVALUATING.

MANY EDUCATORS AND PEOPLE in general place too much faith in tests. It amazes me how conditioned Americans are to putting high values on test scores, especially aptitude measures such as the Scholastic Aptitude Test (SAT) or Graduate Record Exam (GRE). As a member of graduate school admissions committees, I have often heard someone say, "We've got to admit this person. Just look at her scores!" Or perhaps you have heard this one, "His record isn't very good, but he tests well."

Testing is also highly over-valued in evaluation of instructional innovations. In many of the comparative evaluation studies reporting statistically significant differences in test scores between two groups, much is made of the findings without adequate analysis of the reliability and validity of the test. And I have learned firsthand that criticizing such a use of testing is a risky business.

Two of my colleagues and I are currently conducting a series of reviews of evaluations of interactive videodisc education and training programs for *The Videodisc Monitor* (Reeves, Brandt, & Marlino, 1988). In the very first review, (Reeves, 1988) I critiqued an evaluation of an IVD program designed to teach medical students how to assess infant neuromotor dysfunction (Huntley, Albanese, Blackman, & Lough, 1985). I believe that the program itself is an excellent example of IMM, but in the review, I questioned the educational significance of differences of 0.8 on a 10-item knowledge test and 1.3 on a 9-item diagnostic test, especially since both tests had extremely low reliabilities.

In a subsequent published response to my critique, Albanese and Huntley (1988) protested that my criticism of the test results was unjust. First, they asked how can a value be placed on an error in medical diagnosis such as might be involved in missing early signs of infant neuromotor dysfunction. Second, they described a reinterpretation of the test results which would make a reader think their program had achieved the most astounding results possible and that their results were all the more significant because the tests had been so unreliable.

Even if there are no fallacies in their analysis (as I believe there are), the only important difficulty is that Huntley and Albanese (and so many others) fail to make a link between the real world (where medical diagnosis errors are indeed terrible to contemplate) and the artifical world of tests (where a score may mean little or nothing). The students in the control sample scored an average of 6.6 on the "diagnostic test" and the experimental (IVD) students scored 7.9. My concern is not whether the difference between the two groups is significant, but whether the students completing the IVD training program have mastered the diagnosis of infant neuromotor assessment. I want to know how expert pediatric physicians would score on such a test. I want to anchor that test to something in the real world, not argue effective ranges and effect sizes.

Hence, this commandment is intended to caution against putting too much faith in test scores; in fact, it might be rewritten as "Thou shalt not trust tests." Whenever, test results are reported in an evaluation, try to find out as much as possible about the reliability and validity of the test before placing unjustified faith in the results. Fortunately, at a time when even such sacred cows as the Scholastic Aptitude Test are being questioned, a more realistic understanding of the limits of tests may be in the offing.

THE SEVENTH COMMANDMENT

ENTER NOT INTO THE EMPIRICAL SWAMP OF MEDIA COMPARISONS, AND RESIST THE TEMPTATIONS OF THE NUMBER CRUNCHERS.

AS STATED ABOVE, the first issue of the new journal, *Interact,* contains an article titled, "The Instructional Effectiveness of Interactive Video Versus Computer Assisted Instruction" (Tiene, Evans, Milheim, Callahan, & Buck, 1989). I maintain that this particular study is a classic illustration of the difficulties with comparative media studies. It was personally ironic when I read this article in the *Interact Journal* because I am listed as a consulting editor for this

publication. I have spent considerable time and effort during the past five years speaking out against exactly this type of evaluation.

I have not been alone in this critique. Certainly, the best known critic of comparative media studies is Dr. Richard E. Clark of the University of Southern California. Dr. Clark has compared these types of research and evaluation studies to studies that might investigate the relative nutritional value of food delivered by different types of trucks. According to Dr. Clark, the media (e.g., IVD vs. CAI) are just "vehicles" for instruction, and therefore, what really must be investigated are the effects of different instructional designs. Dr. Clark (1983), perhaps playing "devil's advocate," has gone so far as to claim that media are irrelevant in the delivery of instruction and that any instructional design can be replicated by any media.

Although I agree with Dr. Clark's criticisms of comparative media studies, I disagree with his conclusion that media are irrelevant. To extend his truck and food analogy further, in some cases the vehicle really does matter. For example, I would not want to deliver ice cream in a flat bed truck during a Georgia summer. Even if the nutritional value of the ice cream was not affected, there would be devastating effects on its appeal. More to the point, it seems that certain types of instruction demand specific types of media. For example, even if an instructional simulation of new arthroscopic techniques for orthopedic surgeons could be delivered by several different types of media, interactive videodisc (IVD) or some other variant of IMM would be the most effective and efficient medium because of the requirements for specific fast-paced interaction, immediate feedback, and the realistic video of real time arthroscopy. In my opinion, generating a list of instructional requirements which could be effectively and/or efficiently delivered *only* with IMM would be easy.

Clark's conclusion that media are irrelevant is also questionable given the reality that teachers, trainers, and others do develop and select different types of media, and that they invariably want to evaluate the effectiveness and efficiency of these materials. Developers of instructional innovations often have specific reasons for replacing existing instructional delivery systems. The choice of media are relevant to the developers, and they might choose to evaluate the relative effectiveness of two alternative media. Unfortunately, the adoption of experimental

and quasi-experimental evaluation designs is not adequate to provide sufficient information for making future media selections nor to guide design decisions. I recommend a more comprehensive approach to evaluation, considering many different levels and methods of evaluation. I have described specific recommendations for alternative research approaches in other publications (e.g., Reeves & Lent, 1984; Reeves, 1986; and Reeves, 1989).

My own experiences with media comparison studies have been dismal. One graduate student on whose committee I served sought to compare the effects of computer-based instruction with and without video support (Peters, 1988). As all too often occurs in such studies, there were no significant differences between effects of the two instructional treatments. This was hardly surprising since the role of the video in the instructional treatment was unclear, and further, because the relationship between the instructional content and the tests utilized was unspecified. Fortunately, the student included a number of other interesting and more successful aspects in his dissertation research and he was awarded his degree. But I shall be more reluctant than ever to approve any type of media comparison study, no matter how new or exotic the delivery systems involved in the inquiry.

THE EIGHTH COMMANDMENT

THOU SHALT FOREVER REMEMBER THAT DATA DON'T MAKE DECISIONS. PEOPLE DO THAT, AND YOU SHALL LABOR SO AS TO REQUIRE THAT EVIDENCE BE THE SOURCE OF DELIBERATION.

ELSEWHERE IN THIS PAPER, I argue for the exploration of computer modeling as a strategy for evaluating IMM. I do this not so much because I think such an approach will have an immediate and dramatic impact on the development of IMM, but because I believe the existing experimental and quasi-experimental paradigm has provided so little to IMM or education in general. I also do not want to leave the impression that I think the computer can evaluate phenomena

better than the human being. The computer is just an instrument for the evaluator to use. The evaluator must still interpret and report the information in ways which are meaningful to the educational community.

My perspective on the relationship between the computer and human evaluator is illustrated by the following quote from Carlo Rubbia, the winner of the 1986 Nobel Prize in Physics:

> You have all this computing, but the purpose of all this tremendous data analysis, the one fundamental bottom line, is to be able to let the human being give the final answer. It's James Rohlf looking at the fucking event who will decide whether this is a Z or not. (Taubes, 1986; pp. 137-138)

Although I am an advocate of further exploration of computer modeling as an evaluation strategy, I am also an advocate of other methods of inquiry, e.g., ethnographic methods (Fetterman, 1984). Any form of instruction involves numerous variables interacting with great complexity. As noted by Pagels (1988), computer modeling can be used "to distill that complexity down to a humanly manageable amount of information so that we can apply our intuition to it and see what is going in" (p. 41). At this time, this process of distillation is clearly more an art than a science, and traditionally trained experimental researchers and evaluators may experience considerable trauma in changing paradigms. Nonetheless, the poor record of previous efforts to advance instructional understanding demands that we try new ways.

The fact that people, not data, make decisions has negative as well as positive consequences. On the negative side, people can ignore the results of an evaluation if they conflict with their own agendas, as did the manufacturers of the mathematics curriculum described above. On the positive side, as long as creative human minds are involved, doubts and alternative hypotheses will always be explored. Ultimately, people must take responsibility for decisions, no matter how powerful the data supporting the decisions may be. This is difficult, and there is a growing tendency to blame mistakes on poor data or on computers. One example, is the "buck passing" that took place in the aftermath of the Challenger Space Shuttle disaster, as one contractor and government

agency after another blamed the system failure on bad information, never on poor human judgment.

It is important to stress again that evaluation provides no automatic formulas for establishing truth and that there is much more uncertainty involved in evaluation than the quantitative findings usually reported by evaluators might indicate. Evaluation cannot prove anything; it can only support or not support conclusions and decisions made by human beings. Furthermore, it is essential to remember that evaluations do not make decisions, people do. This last point is important because evaluation almost always occurs within a political context (House, 1980), and because the findings of evaluation must often compete with other sources of influence to affect decision making (Cooley and Bickel, 1986).

The cold reality of human decision-making hit me several years ago after my colleagues and I had completed the production and programming of a major interactive videodisc program for a very large computer company. The project had been extremely demanding for months, often requiring more than one hundred hours work per week. But we maintained this intensive effort, partly motivated by the belief that we would be able to conduct a substantive evaluation of the IVD program upon its completion. The lengthy program (100 hours of training) held forth the possibility of numerous kinds of evaluative inquiry and research.

However, in the initial meeting to discuss the evaluation of the newly released program, we were informed that the product would not be evaluated. After all, our client explained, the computer company did not evaluate other types of software, it just marketed them and let the consumer decide. The client went on to explain that there was little value in planning even a formative evaluation of the program because as soon as the program was released, the initial users would provide the basis for debugging and enhancement. Our training in instructional technology and personal experience told us that formative and summative evaluation were essential to the practice of systematic instructional design, but this was rejected in favor of the market imperative.

THE NINTH COMMANDMENT

THOU SHALT NOT COVET THE STATISTICIAN'S LANGUAGE; RATHER YOU SHALL SEEK TO SPEAK THE VULGATE.

STATISTICS are very useful; our modern life is replete with statistics. "Nine out of ten physicians prefer; the median price of a new home is; shooting 73% from the free throw line; a two percent growth in cost of living;" advertisements, the news, and sports thrive on statistics. This might be harmless if everyone had a solid grasp of statistics, but only a few are privy to the strengths and limits of statistical description and inference. For most people, a statistic, like a test score, takes on more value than warranted simply because it is a statistic. After all, if something can be quantified, it must be right.

I don't claim any great expertise as a statistician, but my limited study has left me with a healthy skepticism about the use of statistics. The literature on evaluation of IMM, scarce as it is, relies too heavily on the use of comparative experimental and quasi-experimental designs which conclude with statistical results of significance or non-significance. I suspect that when most people hear that the results of such a study are statistically significant, they interpret that to mean that the results were really important or really powerful. Actually, all that statistical significance tells us is that the results, however large or small, relevant or irrelevant, were not the result of chance. Statistical significance, therefore, is only the first step in reporting the "significance" of evaluation results; the evaluator and his/her audiences are still responsible for the interpretation of the social or educational significance of the findings.

While I encourage a healthy skepticism about statistics, I don't want to reject them. An aversion to statistics on the part of some evaluators has prompted them to reject all quantitative methods. As a result, there is an on-going battle between proponents of quantitative and qualitative evaluation and research in education. I saw this come to the fore at a recent meeting of a professional educational technology association when someone proposed the establishment of a special award for the best qualitative study done each year. The discussion of

the merits of such an award quickly became vehement resulting in the hasty departure of one of the proponents from the room. This type of debate seems silly to me because quantitative and qualitative evaluation methodologies are just two options from the list of inquiry methods which also includes historical, philosophical, and computational methods (Jaeger, 1988). I cringe when I hear someone describe him or herself as a qualitative evaluator; it sounds like someone saying that he or she is a hammer carpenter, forsaking the saw and chisel. The choice of evaluation methods is like the carpenter's choice of tools where the decision to select one must be based upon the nature of the job.

On the other hand, I can empathize with those evaluators advocating for qualitative as opposed to quantitative inquiry. During a recent computer-based education conference in Bulgaria, I witnessed a gross misuse of statistics and the experimental research paradigm. The offending investigator was reporting on research aimed at supporting her thesis that video games have positive effects on the development of children's thinking skills (Greenfield, 1989). The tests used to measure these effects were highly suspect with respect to reliability and validity, and to further confuse the issue, the researcher also chose to present her data in the form of levels of statistical significance unaccompanied by test scores. For example, she claimed that a comparison with a significance level of .01 was "more significant" than a comparison with a significance level of .05. This is a common fallacy. As even an elementary student of statistics knows, statistical significance simply means that a finding did not occur by chance at a predetermined level of significance. A research result either is significant or it is not; it is not more or less significant.

The particular conference was being simultaneously translated into Russian and Bulgarian. There was time for questions at the end, but I could not attract the attention of the session chair. I fear that scores of international conferees, many from developing nations, left Bulgaria believing that sound research had demonstrated the beneficial effects of PacMan, Castle Wolfenstein, and similar video games on the mental skills of children. In my opinion, the research provided little evidence for this thesis, but as is so often the case, statistics were misused to make inappropriate points. I foresee a dramatic rise in the sales of Nintendo in Peru, Sri Lanka, and Nigeria!

In most cases, evaluation of IMM will demand the choice of a range of methodologies, both quantitative and qualitative. The notions of convergence and triangulation are important here (Mark and Shotland, 1987). Since no one methodology can provide unequivocal support for a theory, evaluators should attempt to support their hypotheses from a number of perspectives. The numbers of the statistician can provide one perspective, and the stories of the ethnographer can provide another. Neither is inherently more valuable than the other. Both can have the potential to illuminate decision-making individually, but their combined impact is synergistically better than the simple addition of the two.

Evaluation has been defined as the process of providing the designers and users of IMM with timely, accurate information which will contribute to decisions about the improvement, continuance, and/or expansion of their programs (Anderson & Ball, 1978). This definition implies that the role of evaluation is to provide people creating or using IMM with any and all information to support their conclusions and improve their decision making. The rationale for this type of evaluation is the belief that informed decision-making is better that uninformed or misinformed decision-making. In the final analysis, the role of evaluation in IMM is nothing less than to provide practitioners with the information they require to actualize their expertise and creativity to design, produce, use, and improve IMM. A general premise of this stance is that the phenomena involved in learning are so complex and so difficult to measure that multifaceted evaluation methods are required to obtain meaningful information (Cronbach, 1982; Hunter, 1987).

Another perspective on the evaluation of IMM is provided by the following quote from Saul-Paul Sirag:

> The essential point in science is not a complicated mathematical formalism or a ritualized experimentation. Rather the heart of science is a kind of shrewd honesty that springs from really wanting to know what the hell is going on!

THE TENTH COMMANDMENT

LASTLY, REMEMBER TO READ FAR INTO THE NIGHT AND GO SOUTH IN THE WINTER. BETTER YET, LIVE IN THE SOUTH.

THESE COMMANDMENTS are derived from an earlier set of nine evaluation commandments written by my mentor, the late Dr. Edward F. Kelly. He taught me many things, including the fact that humor is often the best medium for a message.

REFERENCES

Albenese, M., & Huntley, J. (1988, July). Letter: Response to evaluation review of assessment of neuromotor dysfunction in infants. *The Videodisc Monitor, 6*(7), 26-27.

Anderson, S.B., & Ball, S. (1978). *The profession and practice of program evaluation.* **San Francisco: Jossey-Bass.**

Bayard-White, C. (1985). *Interactive video case studies and directory.* **London, Great Britain: National Interactive Video Center.**

Borich, G.D., & Jemelka, R.P. (1982). *Programs and systems: An evaluation perspective.* **New York: Academic Press.**

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research.* **Chicago: Rand McNally.**

Carroll, J.B. (1963), A model of school learning. *Teachers College Record, 64,* 723-733.

Clark, R.E. (1983), Reconsidering research on learning from media, *Review of Educational Research. 53*(4), 445-459.

Cooley, W., & Bickel, W. (1986). *Decision-oriented educational research.* **Boston: Kluwer-Nijhoff.**

Cooley, W.W., & Lohnes, P.R. (1976) *Evaluation research in education: Theory, principles, and practice.* **New York: Irvington.**

Cronbach, L.J. (1982). *Designing evaluations of educational and social programs.* **San Francisco: Jossey-Bass.**

Cziko, G.A. (1989). Unpredictability and indeterminism in human behavior: Arguments and implications for educational research. *Educational Researcher, 18*(3), 17-25.

DeBloois, M.C. (1988). *Use and effectiveness of videodisc training: A status report.* Falls Church, VA: Future Systems.

Dewdney, A.K. (1988). *The armchair universe: An exploration of computer worlds.* New York: W.H. Freeman.

Eisner, E.W. (1985). *The art of educational evaluation: A personal view.* Philadelphia, PA: Falmer Press.

Fetterman, D.M. (1984), *Ethnography in educational evaluation.* Beverly Hills, CA: Sage.

Gagné, R.M. (1985). *The conditions of learning* (4th ed.). New York: Holt, Rinehart and Winston.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Gleick, J. (1987). *CHAOS: Making a new science.* New York: Penguin.

Greenfield, P.M. (1989, May). *Information technology and visual literacy.* Paper presented at the Third International Conference on Children in the Information Age, Sofia, Bulgaria.

Gruber, H.E. (1985). From epistemic subject to unique creative person at work. *Archives de Psychologie, 53,* 167-185.

Guba, E.G., & Lincoln, Y.S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco, CA: Jossey-Bass.

Hawking, W.H. (1988). *A brief history of time: From the big bang to black holes.* New York: Bantam.

Herbert, N. (1985). *Quantum physics: Beyond the new physics.* New York: Anchor Press/Doubleday.

Hoban, C.F. (1958). Research on Media. *AV Communication Review, 6*(3), 169-178.

House, E.R. (1980) *Evaluating with validity.* Beverly Hills, CA: Sage.

Hunter, J.E. (1987). Multiple dependent variables in program evaluation. In M.L. Mark & R.L. Shotland (Eds.), *Multiple methods in program evaluation.* (pp. 43-56) San Francisco, CA: Jossey-Bass.

Huntley, J.S., Albanese, M., Blackman, J., & Lough, L. (1985, April). *Evaluation of a computer-controlled videodisc program to teach pediatric*

neuromotor assessment. Paper presented at Annual Meeting of the American Educational Research Association, Chicago, IL.

Jaeger, R.M. (1988). *Complementary methods for research in education.* Washington, DC: American Educational Research Association.

Judd, C.M. (1987). Combining process and outcome evaluation. In M.L. Mark & R.L. Shotland (Eds.), *Multiple methods in program evaluation.* (pp. 23-41) San Francisco, CA: Jossey-Bass.

Kenny, D.A. (1979). *Correlation and causality.* New York: Wiley.

Leinhardt, G. (1980). Modeling and measuring educational treatment in evaluation. *Review of Educational Research, 50*(3), 393-420.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20,* 130-141.

Lorenz, E.N. (1979, December). *Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?* Paper presented at the annual meeting of the American Association for the Advancement of Science, Washington, DC.

Mark, M.L., & Shotland, R.L. (1987). *Multiple methods in program evaluation.* San Francisco, CA: Jossey-Bass.

Marlino, M.R. (1989). *An examination of the effective dimensions of interactive videodisc instruction using a process modeling approach.* An unpublished doctoral dissertation. The University of Georgia, Athens, GA.

National Study of Secondary School Evaluation. (1960). *Evaluative criteria.* Washington, DC.

Pagels, H.R. (1988). *The dreams of reason: The computer and the rise of the sciences of complexity.* New York: Simon and Schuster.

Peters, C.L. (1988). *The effects of advisement, content mapping, and interactive video on learner control and achievement in computer-based instruction.* Unpublished doctoral dissertation, The University of Georgia, Athens, GA.

Phillips, D.C. (1980). What do the researcher and the practitioner have to offer each other? *Educational Researcher, 9*(11), 17-24.

Popper, K.R. (1982). *Quantum theory and the schism in physics.* Totowa, NJ: Rowan and Littlefield.

Reeves, T.C. (1989). The role, methods, and worth of evaluation in instructional design. In K. Johnson and L. Foa, (Eds.), *Instructional design: New strategies for education and training.* New York: MacMillan.

Reeves, T.C. (1988). Effective dimensions of interactive videodisc for training. In T. Bernold and J. Finklestein, (Eds.). *Computer-assisted approaches to training: Foundations of industry's future.* (pp. 119-132) Amsterdam, NR: Elsevier Science.

Reeves, T.C. (1988, April). Evaluation review: Assessment of neuromotor dysfunction in infants. *The Videodisc Monitor, 6*(4), 26-27.

Reeves, T.C. (1986). Research and evaluation models for the study of interactive video. *Journal of Computer-Based Instruction, 13*(4), 102-106.

Reeves, T.C., Brandt, R., & Marlino, M.R. (1988, April). Evaluation review: Guidelines, introduction, and overview. The Videodisc Monitor, *6*(4), 24-25.

Reeves, T.C., & Lent, R.M. (1984). Levels of evaluation of computer-based instruction. In D.F. Walker & R.D. Hess (Eds.), *Instructional software: Principles and perspectives for design and use.* Belmont, CA: Wadsworth.

Reeves, T.C., & Marlino, M.R. (1989, April). *An evaluation of the emergency medical conditions interactive videodisc.* Paper presented at Annual Meeting of the American Educational Research Association, San Francisco, CA.

Reeves, T.C., Marlino, M.R., & Henderson, J.V. (1988, April). *Evaluation of acute trauma life support interactive videodisc training.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Roueche, J.E., & Herrscher, B.R. (1973). *Toward instructional accountability: A practical guide to educational change.* Palo Alto, CA: Westinghouse Learning Press.

Sanders, D.P. (1981). Education inquiry as developmental research. *Educational Researcher, 10*(3), 8-13.

Schroeder, J.E. (1982). U.S. Army VISTA results. *Proceedings of the Fourth Annual Conference on Interactive Instruction Delivery.* Warrenton, VA: Society for Applied Learning Technology.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Curriculum evaluation.* American Educational Research Association Monograph Series on Evaluation, No. 1. Chicago: Rand McNally.

Scriven, M. (1973). The methodology of evaluation. In B.R. Worthen & J.R. Sanders, (Eds.), *Educational evaluation: Theory and practice.* Belmont, CA: Wadsworth.

Stake, R.E. (1967). The countenance of educational evaluation. *Teachers College Record, 68,* 523-540.

Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher, 7*(2), 5-8.

Stufflebeam, D.L (1983). The CIPP model for program evaluation, in G.L. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation.* **Boston: Kluwer-Nijhoff.**

Taubes, G. (1986). *Nobel dreams: Power, deceit, and the ultimate experiment.*

Tiene, D., Evans, A., Milheim, W., Callahan, B., & Buck, S. (1989). The instructional effectiveness of interactive video versus computer assisted instruction. *Interact Journal, 1*(1), 15-21.

Tyler, R.E. (1942). General statement on evaluation. *Journal of Educational Research, 35*(7), 492-501.

Wang, M.C., & Walberg, H.J. (1983). Evaluating educational programs: An intergrative causal-modeling approach. *Educational Evaluation and Policy Analysis, 5*(4), 347-366.

ABOUT THE AUTHOR

Dr. Thomas C. Reeves is an associate professor of Instructional Technology at The University of Georgia. He has been involved in the development and evaluation of new training technologies (primarily interactive videodisc) since completing a Ph.D. in Instructional Technology at Syracuse University in 1979. He teaches program evaluation, instructional design, and research methodology courses, and directs a wide range of funded projects for higher education as well as business, industrial, military, medical, and government clients. Author's present address: The University of Georgia, College of Education, 607 Aderhold Hall, Athens, Georgia 30602.