THE ROLE OF EVALUATION IN TRAINING SYSTEMS ACQUISITION

James R. Stonge
Instructional Systems Engineer
Displays/Trainer Products
Electronics Systems Division
Grumman, Bethpage, NY 11714

ABSTRACT

As the number and sophistication of training systems increases, there is a need to evaluate the effectiveness of the systems. The services have attempted to identify and define how that evaluation should be accomplished, primarily by looking at the concept described as "Transfer of Training." This paper identifies the two types of evaluation, formative and summative, that have received universal acceptance in government, industry, and academia. It describes the characteristics of each type of evaluation, and specifies when each should occur. Further, it suggests the role of both the contractor and the government. Those roles include the requirement for active government participation in formative evaluation, a role which has been traditionally ignored. The presentation format of the paper uses a model for evaluation as a foundation for the discussions. The description of the model includes the identification of each requisite step in the process. Each element of the model is amplified in the context of using evaluation for modern training systems.

STATEMENT OF THE PROBLEM

Historically, the rationale for using or acquiring simulators, or any other aspects of the modern training system, has been based on acquisition or replacement cost, or on the availability of training devices. Attempts to analyze effectiveness of training systems have normally occurred after production and delivery, in an evaluation process recently called "transfer-of-training." There has been a real problem in measuring this transfer, and even should this kind of evaluation be perfected, it is conducted too late to have an effect on the system delivered. In the model described in this paper, such evaluation comes under the heading of "summative" evaluation. The government and other procurers of training systems require a procedure, or model, which provides a process for the total evaluation requirements in new training systems acquisitions. In a break with current procedures and practices, this dictates a model which incorporates more of the evaluation activities into the development stage of the program, using "formative" evaluation prior to production and delivery. With this approach to evaluation there is the opportunity for greater cost effectiveness and training efficiency though early intervention. To date, no consistent formative evaluation has been integrated into the acquisition process.

INTRODUCTION

The history and literature about evaluation is primarily found in scholarly research and theory. Therefore, a glossary is provided at the conclusion of this paper, defining many of the terms used in the paper. In addition, a reference section is provided, with an annotated bibliography for those who wish to read more on the subject. The specific requirements of this paper dictate that the emphasis be on a useful model, rather than the theoretical discourse common to the literature. The organization consists of three major areas of discussion: the introductory material, including a short history of evaluation and existing evaluation models; detailed descriptions on the pertinent aspects of the model suggested for training systems acquisitions; and discussions on the rationale for using this model and the role of both customer and contractor with this approach.

History of Evaluation

Prior to 1930, evaluation was almost exclusively concerned with the administration of standardized tests. From that period through 1957, the basic pattern of evaluation activities was established, and were specifically used for higher order cognitive and affective objectives. Since then, evaluation has attempted to affect improvement in the design and development process, to cause improvement in the training outcome. As evaluation evolved, two discrete but related kinds of evaluation have been identified. They are formative evaluation and summative evaluation. These should not be confused with formal and informal evaluation. The difference between formative and summative is timing. Formative evaluation occurs during the early, or formative, stages of a project, before all decisions have been finalized, and before large-scale production has begun. Summative evaluation comes after, to summarize results after implementation. Transfer-of-Training is an example of a task evaluated in summative evaluation, although such transfer is not an objective in every situation. Formative evaluation should be used during the process, summative checks how well the process worked.

Existing Evaluation Models

Various models have been developed to assist managers in following the evaluation process. These models are designed to demonstrate the kinds of activities involved, and the sequence of those activities. None are in the format of a checklist, or guide, to allow the novice to perform evaluation unaided. The major drawback to reviewing these models is differences between the models, and the difficulty in deciding which are appropriate. In addition, the most comprehensive of these models does not include media selection, and none are specifically formatted for acquiring and producing training systems. Rather, they have attempted to address the needs of the processes required for public education, and primarily are designed to work for cognitive information.

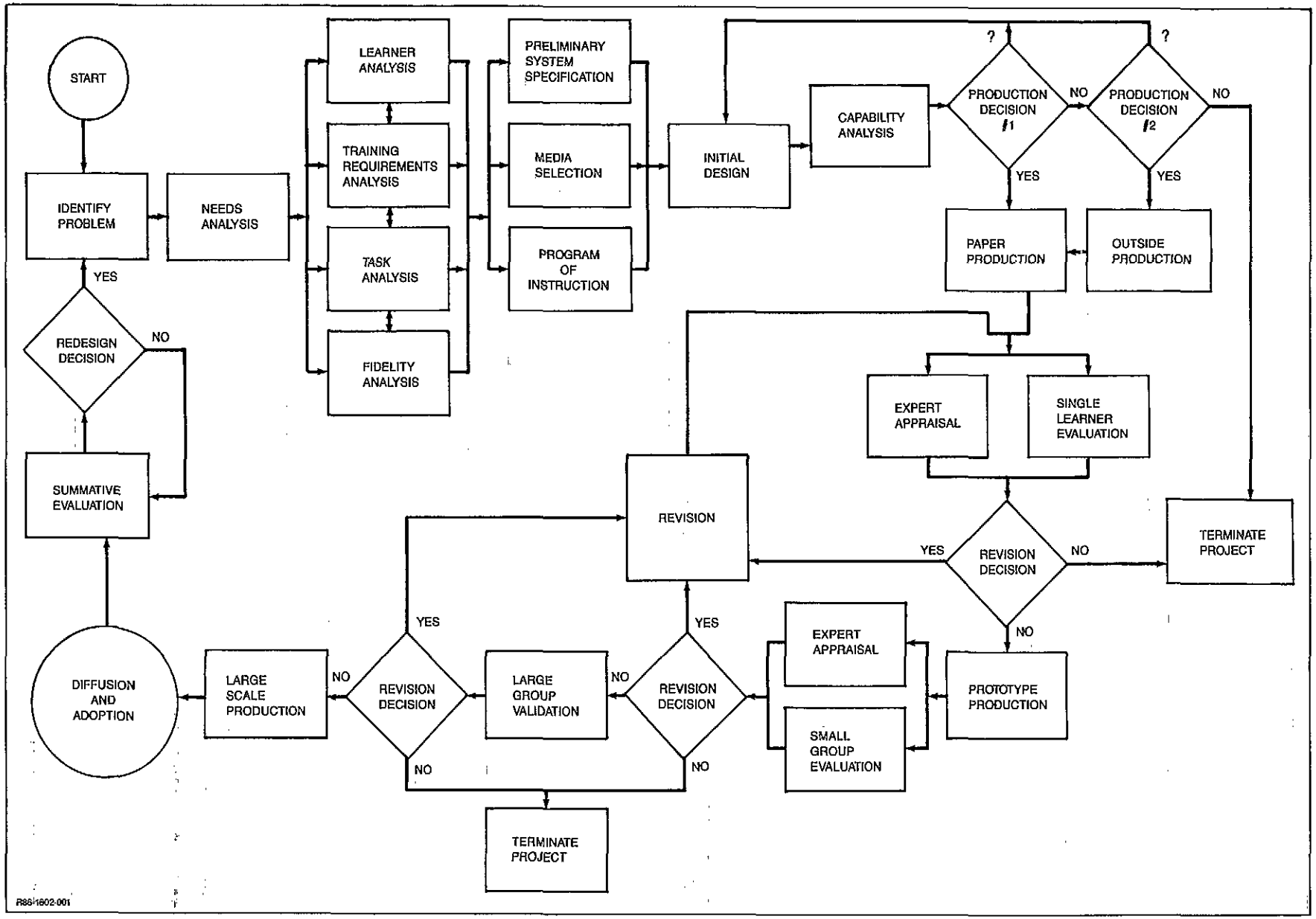

Figure 1 Evaluation Model

Evaluation in Training Systems

In 1977, the author developed a model for evaluation and product development. This model for using evaluation in the acquisition of training devices is based on that previous model, and incorporates the realities of today's environment and technologies. This new model is presented in Figure 1. The model includes the total acquisition process, starting with problem identification through the complete life cycle of a system. However, discussion of every step in the total process is beyond the scope of this paper. The discussions will be limited to evaluation, and those steps in the process associated with evaluation. Discussions will start with the production decision steps, and conclude with summative evaluation.

THE MODEL

Traditional Instructional Systems Development (ISD) procedures and technologies are employed in the initial stages of the model. In a large percentage of acquisitions, these are completed by the customer prior to going to potential customers. While in some instances the analytic processes are included in the contracted system, it is not common to all acquisitions. The first common step in virtually every procurement is that of the production decisions.

Production Decisions

The first production decision is to determine whether you have the capability to accomplish the production of the design in-house. (See Figure 2) For most military organizations, in most instances, the answer is no. That can often be the decision for corporations as well. If the answer is no, you still have two choices, to redo the design, or go to the second choice. The second choice is to decide whether to go outside of your organization for production. This is when RFPs are released by the military, or teaming arrangements are made by contractors. If the answer is no, you still have two choices. You can look at the initial design again, or you can terminate the project. When contractors decide to terminate, the evidence is a no-bid decision. For the military, it is the delay or non-release of an RFP.

Whether you decide to do the project yourself, or go outside for help, you will end up in the same place, the start of production. Many times, when someone decides to go outside for production, they skip all the intervening steps until delivery, which is called Diffusion and Adoption in the model. This is a serious mistake. Your responsibilities do not end. Regardless of who builds the system, the steps in the process must be completed. You cannot absolve yourself of responsibility for the final product by shipping out the work. You are the office of responsibility, and often the user as well.

Paper Production

This is the rough draft, or storyboard stage of a project. It is the creative stage based on the analyses, designs, specifications, and decisions that have preceded it. It has been called intuitive, because it is conceptual, and has not been verified. It is likely to be full of imperfections. In its simplest forms, it will consist of typewritten manuscripts in place of audio information; paper storyboards for visual depictions; drawings or sketches to represent simulators or other hardware. Generally, the representation at this stage is the cheapest form of the finished project, as changes are anticipated, and costs are reduced by keeping the representations simple.

In some instances, this "paper" representation may be more sophisticated, because of the newer processes employed in storyboarding. Due to the nature of the project, it may include non-professionally narrated audio tapes, black and white photographs, and cardboard mockups of equipment. It may even include some rough edited videotapes or other more sophisticated media. Regardless of the representation, for the processes and techniques used it should remain the cheapest form of representation of the finished product.

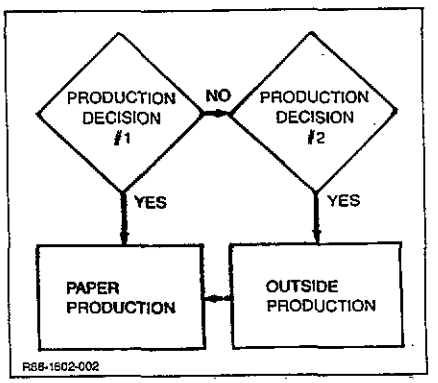

Figure 2

First Evaluation Stage

The production of validated materials requires a series of successive approximations, although the exact number of iterations required is project dependent. The earlier the revisions are made, the greater the advantages, to prevent unnecessary waste of time and resources. There are three stages in formative evaluation: the individual student stage, the small group stage, and the final stage, the field test. The first stage includes both single learner and expert opinion. (See Figure 3)

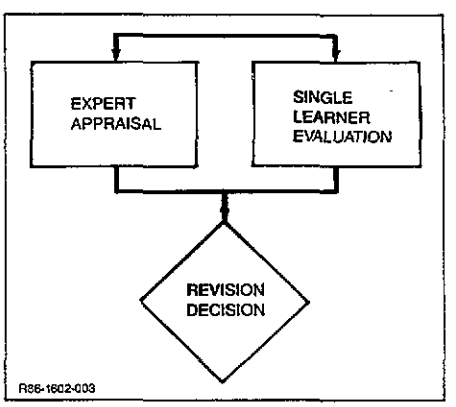

Figure 3

Expert Appraisal. There are several types of experts required for expert appraisal. One is the subject matter expert (SME). The SME will identify inconsistencies, unintelligible jargon, format errors, and all of the other problems found in first drafts that would frustrate a student representative of the target population. Their domain is content. They identify missing information, ambiguous information requiring clarification, and redundant information. As a result of their appraisal, the information presented is likely to be more complete and accurate.

A second kind of expert is the instructional design expert. These experts look at the overall design of the system. To some extent, this is peer approval. The instructional design expert will look at such aspects as consistency with learning principles, design effectiveness, individual differences, and whether the objectives have been met.

The third kind of expert is, for lack of a better term, a media expert. Various media are used in a system, and this expert could be a hardware or software engineer, as well as a traditional media specialist. The goal of these experts is to evaluate the media being planned, to determine if it is feasible to do it that way, to identify alternatives that can do it better and/or cheaper and.or quicker. More importantly, for those who are working with limited budgets or under the constraints of a firm, fixed price, these experts evaluate whether the design can be implemented within the budget and schedule constraints that have been imposed.

Single Learner Evaluation. The rationale for single learner evaluation is extensive and persuasive. Important in using single learners is the selection of the individual learner. One suggestion is that students familiar with the task are most appropriate. These students should have skills similar to that of the target group. This sophisticated learner offers several advantages similar to that of the subject matter expert: the ability to identify inconsistencies, unintelligible jargon, format errors, and other first draft problems. While some may argue that larger groups are needed, the research has found no additional advantages to using more than one learner at this stage. Further, it has also been found that using large groups in this stage would be wasteful of both time and effort. Not surprisingly, developers are unwilling to make changes without extensive review, but the literature does not support that bias, and states that this expert appraisal and single learner verification are sufficient to make the first decision.

The single learner is most valuable in the early stages, when large scale revisions are feasible, and the system has not been greatly refined. It is a shortcut, finding problems without extensive usage of student personnel, evaluations, and materials. But, although it is a shortcut, you should not shortcut the process. Multiple single learners should be used, for the results are used in the revision decision, and one student may not provide sufficient information. A good recommendation is to use learners who meet the average characteristics of the target population, as well as those who represent the two extremes. The bright, average, and slow learner all have different characteristics to bring to the evaluation process, and there are advantages to each.

Revision Decision

It is at this point that the first revision decision must be made. Each time the developers come to a decision point, there are three choices: proceed as is, revise the system, or quit. If the decision is to proceed, then you proceed to the next step in the process. At this point it would be prototype production. If the decision is to terminate the project, then you exit the process, as there is no place else to go, except maybe to start over at the beginning. A common mistake is to go back to some other place in the process, such as back to the production decision. The flaw is that something has obviously gone wrong somewhere in the total process, and you may not go back far enough. Then you would have someone else produce a system with the same flawed concepts, and still not have a viable system.

The first time through you will unlikely be able to choose either to go on or to terminate. Because of problems that will be identified you will need to make revisions. After making revisions you may be able to choose one of the other options. You should not expect to get through this point the first time without going through the revision process. Unfortunately, few people understand this, and managers and customers alike are often frustrated that revisions are necessary. Without the revisions the results are likely to be no better, and more likely to be worse, when the system is delivered. Then nobody is happy, resulting in contractors getting bad reputations and customers feeling that this kind of training does not meet their needs.

One problem that may occur is that the single learner verification produces insufficient data to make a decision. The learner may have had problems, but you are unable to pinpoint why. In that case, use a second, or third, or fourth individual learner. If you are unable to make a decision, continue evaluation until you can.

As it is unnecessary to discuss the revision process every time it occurs, there is one more point that should be made about revision decisions. Regardless of when the decision is made, it is important that you go back to the single learner evaluation after the revision. The same reasons for using single learner evaluations in the first place are still valid after revisions. It is relevant to remember that the single learner will often provide information that is lost in larger-group evaluations.

Prototype Production

When the individual learners have sufficiently debugged the system, it is time to move into prototype production. The prototype allows additional evaluation in a more realistic setting, and the production itself may help to bring additional problems to light. If the movement from initial design to final production is seen as a continuum, the prototype production can have reduced fidelity. If you have the time, commitment, and financial resources to do so, you

may consider having successive approximations toward reality during the prototype production process. Using videodisc as an example, the paper production used in the single learner evaluation may have been line drawings on storyboards. For the first prototype production, photographs could be used. Successive prototypes may move from photography to videotape to videodisc. The rationale for this is that it is not unreasonable to expect that revisions will have to be made after the first prototype is complete.

Small Group Evaluation

Following completion of the prototype, the second stage in formative evaluation occurs. (See Figure 4) The expert appraisal that occurs here is extremely similar, if not identical, to the appraisal that occurs in the single learner stage. It is important, then, to use different experts, so that the bias of the first experts can be offset by new appraisers. Otherwise, if new experts were not used, it would be the non-validated design of the first experts that would be delivered, in which case there is no added value, and you could have used the original design without expert appraisal. The inclusion of a second set of experts allows for a consensus of expert opinion, as there would be three sets of experts: the initial designers, the first experts, and the second experts.

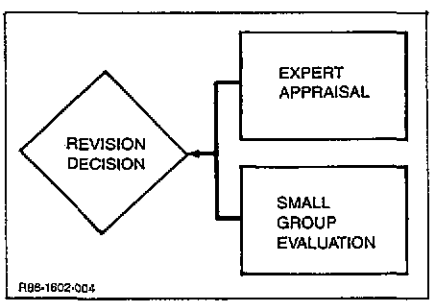

Figure 4

The small group is a sample of the target population. Most of the characteristics of the target population should be present. That does not mean that the sample has to be scientifically selected. It does mean that any obvious differences, such as age, sex, mental ability, prerequisite knowledges, etc., should be incorporated into the sample, especially when those differences could have some impact on the results. There is no rule on the size of the small group, and smallness could be relative to the target population. A general guideline is to keep it as small as possible while maintaining the major characteristics of the larger audience, yet having it large enough to allow group interaction with the system.

In actuality, the small group could be used prior to construction of a prototype, but it is not deemed best. The best time to use small groups is when major problems have been resolved, when the product has more closely approximated the final system form, but while it is still economical to make major revisions. It is for that reason that formative evaluation is completed prior to large scale production. Because of the prior single learner verification, the producer has some confidence in the product, so there is justification for producing the prototype, even though it is anticipated that additional changes are likely to occur. It is a half-way point, and successful evaluation of the prototype leads to the next step. The revision decision/revision process will not be discussed again at this point.

Large Group Validation

Successful completion of the small-group evaluation leads to the final stage of the formative evaluation process. (See Figure 5) Another common name for large group validation is field testing. The objective here is to conduct the evaluation in an environment as close to the actual training as possible. It should be considered as analogous to a final dress rehearsal for a theater production. Although modifications still are possible, by this time only minor changes should be required. For this reason, the prototype can be extremely close to the deliverable system. Because it is costly and time consuming to have large group validations, it is imperative that the system has been thoroughly tested in the earlier stages, and revision decisions were not deferred until now. However, it is still possible to need major changes, even at this point in the process. What worked in a single learner or small group validation may fail the large group validation. But with each successful approximation, the confidence level should be greater. To have reached this point, the system must have just successfully completed both single learner and small group evaluations, with subsequent revisions based on the findings of those evaluations. Again, however, following this validation a revision decision must be reached, and, if required, revisions must be made.

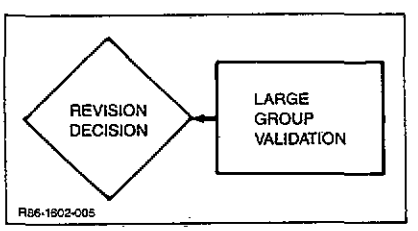

Figure 5

Large Scale Production

After the formative evaluation is complete, full scale production occurs, with confidence in a long and useful life for the system. Actual production shall not be discussed in this paper. If formative evaluation has done its job, this step is technical in nature, and does not warrant detailed discussions in this forum. Suffice it to say that full quality control and periodic checks by both contractor and customer are essential, for even the best designs can be negated with faulty workmanship.

Diffusion and Adoption

With production comes dissemination and use. That means normally that the system is shipped to the ultimate user for use as designed. Unfortunately for both customer and contractor, most people think the process stops here. There are two steps beyond production, Diffusion and Adoption, and Summative Evaluation. Although both are important, the most important may be Diffusion and Adoption. Some organizations consider Diffusion and Adoption to be part of production, but here it is discussed as a separate element in the process.

Although Diffusion and Adoption are used together as a single phrase, each word has specific meaning. Diffusion can be thought of as distribution, but it is more than that. It includes such things as an instructor/operator course, setting up the system and demonstrating it, and providing support to the user. Adoption includes some of the things in Diffusion, but more specifically means getting the user to accept and use the system. It is this last part that is often least understood and accomplished. For a system to be useful, it not only has to be designed well, produced professionally, and delivered in a timely manner, it must also be accepted by the instructors and students who are to use it. That acceptance leads to increased use, increased familiarity, and, as a result, superior performances by the students.

Within the general area of Diffusion and Adoption, a critical element is the instructor/operator course. While the course is designed to enable instructors to use the system, the real purpose is to provide familiarity. The instructors are like all other people, and are most comfortable with things they understand. As the instructors learn to use the system, they accept it, understand it, and encourage additional uses and applications.

Summative Evaluation

The terminal stage of evaluation is the summative evaluation. It occurs after the system has been fully implemented. This is the market test. Summative evaluation provides a history of the system's effectiveness over time. It is a continuous effort, and provides the perspective of time that can be critical in determining true effectiveness. At any one point in time, a system may be more or less effective than at other times. Generalizations based on a single point in time may be misleading, good or bad. It is the summation of this data that provides the information that leads to the final revision decision. That final decision, based on evolving conditions, is to continue using the system as is, or to reinitiate the entire process.

The purpose of summative evaluation is twofold. One reason for summative evaluation is to prove that the system does indeed accomplish its purpose. This is often called criterion referenced evaluation. The students' performances are compared to the specified objectives, and success or failure is identified. One example of summative evaluation as criterion referenced evaluation is the transfer-of-training comparison. Here students are tasked to perform on actual equipment. The evaluation determines whether the training system prepares them for this objective, performance of the tasks in a non-training environment.

Summative evaluation may initially meet the objectives, but fail to keep up with changing situations. Therefore the second reason for summative evaluation is the continuing validation of the effectiveness of the training system. Sometimes obvious adjustments are identified, such as major weapon system modifications. Other times there is no one single event that can be pinpointed, but rather the evolution of equipment and personnel may slowly compromise the effectiveness of the system, and summative evaluation will aid in highlighting the trend toward system incompetence. While the first reason has obvious advantages, it is the second use of summative evaluation that has the greatest validity over time. Because of these evolving conditions, it is important that the process be reinitiated, as you have come full circle, and the original data, which lead to the initial design, is obviously no longer valid.

RATIONALE

While many reasons can be postulated for using formative evaluation, according to this or any other model, all of the reasons fall into two main categories. You can reduce overall training systems acquisition costs, and you can create more effective training systems. Ideally you can do both, making future training systems more cost effective and training efficient. With current constraints in resource acquisitions, it had become imperative to achieve these results. Each of these reasons are discussed in detail below.

Cost Efficiency

There is no argument that there is an additional investment in time and money at the early stages of a program using full formative evaluation, at least in comparison to existing methodologies of system acquisition. However, some examination of those additional costs are relevant. First of all, a "paper" version is almost always completed prior to production. During the development of the Grumman trainer for the A-6E Detecting and Ranging Set (DRS), the drawings required for production were used to make a "paper" version of the trainer for formative evaluation. Technical manuals and other existing training guides were used in lieu of media presentation. The only additional costs were the time and personnel required for the single learner and small group evaluations, which lasted together less than one week. The DRS was delivered on time and cost schedules, despite the additional formative evaluation. Although delivered in 1979, the DRS as designed is still capable of training Navy and Marine Corps students. No changes to the production units were required and learning effectiveness is high.

In another example, the Army Maintenance Training and Evaluation Simulation System (AMTESS) used storyboards for expert appraisal and other forms of formative evaluation. These storyboards had to be created regardless of any evaluation to be performed. The costs associated with the additional steps were only that of the personnel involved in the evaluation. No schedule slippage occurred because the evaluation was conducted

simultaneously with other activities. The AMTESS story is incomplete, because it still is following the model. Small group evaluations and field testing are continuing following prototype production, and upon successful completion of these evaluations, full scale production is anticipated.

In both instances mentioned, there was no significant additional cost, and both are successful training system procurements. Potential savings are indeed the result, although they can not be accurately measured. The savings are those realized because the systems are working, that they do the job, and do not have to be replaced. Savings are realized because students receive the training they need, and therefore are able to perform on the job. If it would be appropriate, and it isn't, programs could be identified where they did not do the job. There are measurable losses on those projects, the expenses for the system that didn't work, the costs associated with ineffective training, and the costs associated with inefficient performance on the job and potential or real loss of equipment and/or personnel because of that poor performance.

Training Effectiveness

Training effectiveness is the ultimate goal. Some attempts to define training effectiveness have been made, but two objectives are consistent in every definition: that the training takes less time than with the alternative methodology, and that students are better trained when they arrive to work in the field than they would have been with other training. The first instance is an example of efficiency, rather than effectiveness. The second is a true example of effectiveness. The efficiency of decreased training time is measured as cost effectiveness. Savings can be realized though less student time to train, which is measurable, but also through the need for fewer training systems and devices, which can also be measured.

Training effectiveness on the job has been addressed as transfer of training (or sometimes as transfer of learning). The Army Research Institute, among others, has been attempting to find a way to measure this prior to the student going to the field. In this model, the summative evaluation makes this measurement, but does it historically, based on field performance. By using the formative evaluation stages, with small and large group evaluations, the need for transfer of training evaluation is reduced. The desire for this transfer measurement has been based on a lack of confidence in the acquired systems. With the increased confidence, and a scheduled summative evaluation, other forms of measurement are not required. Knowing that the system has met the criteria of the evaluations conducted during the formative stage should alleviate most fears.

THE ROLES

The role of the contractor is straightforward. Each step as identified must be followed. It is the role of the military or any other customer that warrants discussion. First, and foremost, the customer must be willing to pay for any costs associated with formative evaluation. Secondly, they must specify in advance that they require this evaluation, that they will pay for it, and that they will support it fully with assets and personnel. Thirdly, they must follow through by incorporating the results of the evaluations and supporting the revision decisions. When a contractor asks for SME support, or for students in individual, small group, or large group situations, the customer must provide the support promptly and to the agreed upon specifications. Delays, insufficient numbers of personnel, or the wrong kinds of personnel, violate the concept. Failure to respond appropriately places an unfair burden on the contractor, and jeopardizes the entire system. The customer must provide their best people, preferably dedicated to the project, for the duration of the project. It is in their own best interest to do so.

The implementation of the model will not happen voluntarily. Winning contractors normally are not going to take some dollar amount and dedicate it to formative evaluation, unless it is imposed by the contract and proceed into the proposal. Otherwise, they would probably have to take it out of profit, and that would never happen. They will also not add it to their costs when making bids, as then every thing else being even, their bid will be higher than the competition who does not add the cost to the bid. Since contractors are by definition profit motivated, there has to be an incentive to encourage the use of formative evaluation, and the military customer has to provide that incentive, either through contracted cost reimbursement or through the promise of future procurement potential.

There are several ways incentives can be used to promote the use of formative evaluation. The most obvious way is to make it a requirement of the contract. Then those who do not include provisions in their proposal will be non-responsive. If they make provisions for it but do not actually perform, they are in non-compliance, and should suffer the penalties involved. If it is included, companies will comply, to remain competitive. Perhaps, formative evaluation should be on a cost plus *basis, even for firm fixed fee contracts.*

Simpler, with longer lasting effects, is to incorporate formative evaluation as a standard contract item, just like environmental requirements, equal opportunity, and other standard requirements. Then it would be required of everyone, and contractors would not be browsing through RFPs to see whether and where it is mentioned in the statement of work. If it were a standard, it would receive ready acceptance. Contractors could not pick and choose. They would have to conform or get out of the business.

Even with ready acceptance, it is important that the customer remains involved. If no participation, observation, and/or documentation is *required, then the intent could be destroyed,* albeit unintentionally. Ideally, the customer will fulfill the roles of observer and participant, whether to provide students, instructors, or both. The customer should be actively involved in any revision process as well, not only to safeguard their interests, but also to provide the expertise necessary in the decision making.

CONCLUSION

The government, like all customers, must

procure quality products in training systems as well as other products. A method for ensuring quality in training systems acquisition has been described. This method incorporates the technology of formative evaluation. Through formative evaluation the customer gains some assurances that the training system being procured will meet the needs of the using community. This in turn eliminates the need for artificial measurements to determine training effectiveness.

To obtain satisfactory training systems products with formative evaluation, the government must provide the methodology for contractually requiring the process, as well as taking an active part during the process. In doing so, there will be savings obtained through lower total costs in system acquisition, as well as the savings available through a more efficient training system. The model provided is one means to reach that goal. Additional research and comment are desired to make formative evaluation a reality.

REFERENCES

Abedor, Allan J. "Second Draft Technology." Viewpoints, Vol. 84, No. 4, July 1972, pp. 9 - 43. This article includes discussions on single learner verification, and the attributes required for this evaluation stage.

Allen, William H. "Research on Instructional Media Design." Educational Media Yearbook 1974. Ed. James W. Brown. New York: R. R. Bowker Company, 1974. Among the many excellent attributes of this article is a discussion on the history of evaluation.

Allen, William H. "Intellectual Abilities and Instructional Media Design." AV Communication Review, Vol. 23, No. 2, Summer 1975, pp. 139 - 170. Dr. Allen was one of the most respected scholars on individual differences of learners, and this article provides substantial discussions on these differences and how they impact the learning process. This is the best source for understanding more about Learner Analysis. This is also one of the earliest works specifically discussing media selection from the practical viewpoint of learner characteristics.

Baker, Eva L. and Marvin C. Alkin. "Formative Evaluation of Instructional Development." AV Communication Review, Vol. 21, No. 4, Winter 1973, pp. 389 - 418. In their article on Formative Evaluation there is a discussion of the history of evaluation. They continue with specific discussion on the aspects of formative evaluation, to include the requirement for early intervention in the whole process.

Campbell, Sam C., et al. Final Report, Army Maintenance Training and Evaluation Simulation System. Bethpage, New York: Grumman Aerospace Corporation, 1980. While much of this publication references specific applications, it does provide insight into many of the traditional ISD processes, including detailed descriptions of how specific analyses were conducted, including Task Analysis, Training Requirements Analysis, and, very specifically and importantly, how a Fidelity Analysis could be and was conducted on that project.

Cronbach, Lee J. "Course Improvement Through Evaluation." Teachers College Record, Vol. 64, No. 8, May 1963, pp. 672 - 683. Cronbach is the father of Formative Evaluation, and this article provides insight into his concept of this evaluation concept. He has written other articles; this one was chosen because it is representative of his work.

Cunningham, Donald J. "Comments on the Case Studies of Formative Evaluation: The Sources of Information." Viewpoints, Vol. 48, No. 4, July 1972, pp. 111 - 118. This is a specific discussion on using students as experts in the formative evaluation process. Among the findings in this study is the ability of single students in the first stage of the formative evaluation process to be as useful as large groups, without the additional time and effort required for large group evaluation.

Davies, Ivor K. Competency Based Learning. New York: McGraw-Hill Book Company, 1973. Among many instructional topics, Davies provides in this book a detailed description of many of the analytic processes, including Task Analysis and Needs Analysis.

Deterline, William A. "Applied Accountability." Educational Technology, Vol. XI, No. 1, January 1971, pp. 15 - 20. This early article includes discussion of the need for single learner verification.

Gropper, George L. Diagnosis and Revision in the Development of Instructional Materials. Englewood Cliffs, New Jersey: Educational Technology Publications, Inc., 1975. Gropper provides discussions on formative evaluation stages and procedures, including specific requirements for single learner verification.

Haney, John B., Phil C. Lange and John Barson. "The Heuristic Dimension of Instructional Development." AV Communication Review, Vol. 16, No. 4, Winter 1968, pp. 358 - 371. Among the important concepts to be found in this article is the thought that good training is a result of successive approximations, which provides justification for formative evaluation early in the development process.

Kandaswamy, Subramanian, Harold D. Stolovitch and Sivasailam Thiagarajan. "Learner Verification and Revision: An Experimental Comparison of Two Methods." AV Communication Review, Vol. 24, No. 3, Fall 1976, pp. 316 - 328. This is one of the specific analytic looks at the effectiveness of single learner verification in formative evaluation. It provides statistical evidence for accepting the validity of the concept.

Kemp, Jerrold E. "Which Medium?" Audiovisual Instruction, Vol. 16, No. 10, December 1971, pp. 32 - 36. At one time, this was one of the most definitive models in media selection. It still has relevance in some applications. Its major benefit is in understanding the thought processes and algorithmic approaches used. It also provides a historical reference to compare to newer media selection models.

Komoski, P. Kenneth. "An Imbalance of Product Quantity and Instructional Quality: The Imperative of Empiricism." AV Communication Review, Vol. 22, No. 4, Winter 1974, pp. 357 - 386. Included in this article is some rationale for single learner verification in the initial stage of formative evaluation.

Lucas, Robert J. "The Relationship of Training and Experience Variables to Teachers' Skill in Judging the Effectiveness of ITV Programs." AV Communication Review, Vol. 22, No. 1, Spring 1974, pp. 79 - 89. Lucas incorporates the need for single learner verification in creating effective video presentations.

Lynn, Robert A. Marketing Principles and

Market Action. New York: McGraw Hill Book Company, 1969. This book from the business world provides insight on Diffusion and Adoption, from the marketing standpoint. It is a corollary to the concept in training.

Mierhenry, Wesley C. "Research in Educational Media and Technology." Educational Media Yearbook 1974. Ed. James W. Brown. New York: R. R. Bowker Company, 1975, pp. 88 - 94. Among other things, this article includes specific information on the history of evaluation.

Myers, Charles R., Jr. "Media Selection: A Contextual Model." Deer Park, New York: np, 1982. This is a detailed, definitive description of the media selection process, with specific steps and guidelines to assist in the process. It is comprehensive, and includes the requisite ingredients to keep it up to date and incorporate the latest media technologies.

Nemsik, James A. The Effectiveness of Selected Learner Characteristics in the Formative Evaluation of Instructional Materials. Unpublished Doctoral Dissertation. Bloomington, Indiana: np, 1973. This is a detailed look at the specific requirements associated with the single learner in the first stage of the formative evaluation process.

Newman, Joseph W. Motivation Research and Marketing Management. Boston: Harvard University, 1957. This book includes discussion on marketing of new business products, which is related to Diffusion and Adoption of training systems.

Scott, Roger C. and Stephen L. Yelon. "The Student As Co-Author - The First Step in Formative Evaluation." Educational Technology, Vol. IX, No. 10, October 1969, pp. 76 - 78. Scott and Yelon offer details about the stages of evaluation. They identified the three stage approach. They provide specific guidelines for the first stage, including both single learner evaluation and expert opinion.

Scriven, Michael. "The Methodology of Evaluation." AERA Monograph Series on Curriculum Evaluation. No. 1. Eds. Tyler, Gagne, and Scriven. Chicago: Rand McNally & Co., 1967, pp. 39 - 89. Scriven has the distinction of having given Formative Evaluation its name, and his monograph provides important insight to his concept of this evaluation strategy.

Stolovitch, Harold D. From Models to Modules: Application of an Adapted Model for Instructional Development to the Design, Production and Evaluation of Audiovisual Training Modules. Unpublished Doctoral Dissertation. Bloomington, Indiana: np, 1975. Primarily, this thesis provides an in-depth analysis of evaluation models, culminating in a recommendation for a "new" model, that of Dr. Stolovitch. There is also some specific detailed descriptions of steps in the process, including that of Needs Analysis.

Stonge, James R. Evaluation and Product Development. Bloomington, Indiana: Privately Published, 1977. This monograph is the basis for the model discussed in this paper. In addition, it provides additional descriptions of many of the steps in the process, as well as more on the literature of formative evaluation.

Thiagarajan, Sivasailam and Harold D. Stolovitch. Evaluation Workshop: Coordinators Manual, Participants Manual. Bloomington, Indiana: University Consortium for Instructional Development and Technology, 1975. Provides amplifications on the requirement for both single learner verification and subject matter expert during the initial stage of formative evaluation.

GLOSSARY

AFFECTIVE - Related to affective domain, specifically those areas of learning related to mood, interest, enjoyment, and/or other aspects that allow the student to like or accept the related material. One of the three domains of learning (affective, cognitive, psychomotor).

CAPABILITIES ANALYSIS - Determines whether the design selected can be accomplished. In full implementation, will likely include cost tradeoff analysis. Provides data for making production decisions.

COGNITIVE - Related to the cognitive domain of learning, specifically the intellectual acquisition of theory, concepts, relationships, etc. that are learned by thinking, not by doing.

FIDELITY ANALYSIS - Looks at fidelity required for training, important for system specification.

INITIAL DESIGN - Based on everything preceding, this is often in the form of a proposal. It is a description of the implementation of the system specification, using selected media and course content.

LEARNER ANALYSIS - Investigation of needs, abilities, capabilities, and attitudes of target audience.

MEDIA SELECTION - Often predetermined, process to select appropriate media for presenting information.

NEEDS ANALYSIS - A further definition of the problem, providing a blueprint for action.

PRELIMINARY SYSTEM SPECIFICATION - Often appears as a statement of work (SOW). Identifies basic components of the system. It is a requirements documentation, pre-design.

PROGRAM OF INSTRUCTION - Lesson content, minimally an outline, provides some degree of specificity on actual course content.

PROTOTYPE - A single version of an intended production run. Used in most production processes, the term usually means a version used for testing, not necessarily identical to the intended final product. In manufacturing, breadboards and brassboards are terms that have similar meanings.

RFP - Request for Proposal, a specific document issued by a customer to prospective contractors, detailing the characteristics and requirements for the contract, as well as the contractual rules to be followed in bidding the contract.

STORYBOARD - Usually a paper product, this is the rough draft of a lesson or script. In most instances, this is the guiding document for media production. The actual characteristics and level of detail will vary with the using organization.

TASK ANALYSIS - Traditional analysis, identifying tasks or skills required to perform training objectives.

TRAINING REQUIREMENTS ANALYSIS - Also called context analysis, looks at when and how final product will be used, as well as evaluating any pre-existing training available.

TRANSFER-OF-TRAINING - Also called Transfer-of-Learning. A term coined to describe the ability of a student to transfer the skills and knowledges learned via one medium or equipment to the new job or equipment. It is also used to refer to the success of a medium to train as measured by the student's ability.

ABOUT THE AUTHOR

Mr. James R. Stonge is an Instructional Systems Engineer at Grumman. As Instructional Systems Development (ISD) manager on the Grumman AMTESS program, with similar responsibilities on their MITS project, he is responsible for all ISD activities for that project. Prior to joining Grumman nine years ago, he was on the faculty of Indiana University, in the department of Instructional Systems Technology. He also served in the U.S. Air Force, where he served as an Instructor *Missile Combat Crew Commander* and a Training Squadron Commander. Mr. Stonge holds an AB in Education, Lycoming College; MBA, University of Missouri; MS in Education, Southern Illinois University at Edwardsville; and additional studies at Indiana University, where he completed all work for a doctorate except completing dissertation (ABD). He has numerous publications including previous Interservice/Industry Training Equipment Conference Proceedings.