A Comparison Between Predictive and Formative Cost-effectiveness Evaluation Techniques for the Assessment of Lecture and Computer-based Multimedia Training

by

Vicki L. Neale

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

INDUSTRIAL AND SYSTEMS ENGINEERING

APPROVED:

Dr. Walter W. Wierwille, Chair

Dr. Thomas A. Dingus

Dr. Mary Beth Rosson

Dr. Antonio A. Trani

Dr. Robert C. Williges

May 2, 1996

Blacksburg, Virginia

Keywords: Cost-effectiveness, Predictive, Formative, Lecture, Multimedia, Training

A Comparison Between Predictive and Formative

Cost-effectiveness Evaluation Techniques for the Assessment of

Lecture and Computer-based Multimedia Training

by

Vicki L. Neale

Dr. Walter W. Wierwille, Chair

Industrial and Systems Engineering

(ABSTRACT)

In an effort to validate a predictive (as opposed to a traditional formative or summative) cost-effectiveness model, a study was conducted to evaluate Kearsley and Compton's (1981) *Benefits Model*. Costs were input into the model as they applied to the design, development, and dissemination of two training programs on the topic of teaching individuals how to detect the level of drowsiness of their colleagues during team operations. The benefits of the training programs were identified, classified, and quantified as they applied to two media: lecture and computer-based multimedia.

The experimenter identified the training system parameters, training benefits, and operational benefits. Then, for the predictive approach, the relationships between training system parameters, training benefits, and operational benefits were classified based on expert opinion. Quantification concerned the assignment of values (-1 or +1) based on expert opinion. The costs to design, develop, and disseminate the training programs were determined based on the parameters of the project. Finally, based on all information present, experts determined which of two training programs would be the most cost-effective to disseminate.

To determine the accuracy of the *Benefits Model* as a predictive assessment tool, the same identified training system parameters, training benefits, and operational benefits were evaluated from a traditional formative evaluation approach. An empirical evaluation was conducted for the two training programs and a determination of the most cost-effective

training medium was made. The data collected in the traditional formative evaluation approach was then compared to the experts' ratings and choice of training programs.

For both the predictive and formative evaluation approach to determining cost-effectiveness, the computer-based multimedia was chosen as the most cost-effective training medium. However, for the predictive approach, the experts' choice was based either solely or heavily on dollar amounts associated with design, development, and dissemination, while the data obtained through the validation process were given little or no weight. All experts stated that it would not have been possible to use the information gathered through application of the *Benefits Model* to determine cost-effectiveness with any confidence.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Walter W. Wierwille, for his always thorough critique of my work. I have appreciated and admired his ability to "see the forest for the trees" and hope that experience will give me the same ability. I would also like to extend my appreciation to my other committee members, Dr. Robert C. Williges, Dr. Mary Beth Rosson, Dr. Antonio A. Trani, and Dr. Thomas A. Dingus.

Throughout this process my husband, Dennis C. Neale, has been supportive and encouraging. For that I would like to thank him and let him know how much I appreciate his confidence in me.

Furthermore, I would like to thank my parents, Michael and Cheryl Lewis, and my brothers, Michael Jr., Matthew, and Mark. I wish everyone could have such a completely supportive family.

TABLE OF CONTENTS

LIST OF FIGURES

Figure:

LIST OF TABLES

Table:

CHAPTER 1

INTRODUCTION

Cloyd S. Steinmetz (1967) wrote:

> As man invented tools, weapons, clothing, shelter, and language, the need for training became an essential ingredient in the march of civilization. Whether man stumbled upon or invented these facets of civilization is of relatively little significance. What is more important is that man had the ability to pass on to others the knowledge and skill he had gained in mastering his circumstances. This he did by deliberate example, by signs, and by words. Through these devices the development process called "training" was administered; and when the message was received by another successfully, we say that learning took place and knowledge or skill was transferred (p. 1).

As human beings we spend a great deal of our lives "in training" for some purpose or another. This acquisition of information may take place through the course of our daily activities, or through some type of structured education or training program. Examples of structured education are attending high school, attending trade school, participating in an informal on-the-job training routine (for example, learning the logging trade tends to be informal), or participating in a formal job-oriented training program. As the body of knowledge we need to get through our lives becomes larger and more complex, the structured approaches to acquiring knowledge and skills become necessary.

The structured approaches to learning are called both "education" and "training." Even though the terms "education" and "training" denote the same idea of a structured approach to learning, a distinction is made in the literature between the terms. Hawthorne (1987) notes that education is a process of learning to think and to examine and solve problems, where as training is specific and skill oriented. Education tends to be used to imply instruction in an academic setting, such as a high school, trade school, university, etc. Training tends to imply instruction through an organization, such as industrial on-the-job training or courses designed for a specific task. The contrast between the two settings is that the structure of information is different. As Gordon (1994) states:

> Traditionally, educators have been asked to teach general knowledge in formal educational institutions, and training professionals develop more specific job-oriented or task-specific training programs for employees... There has also been a difference in the length of time for instruction. Educational institutions have emphasized teaching of more general knowledge for longer periods of time. Training, on the other hand, has mostly consisted of short-term activities such as workshops that provide knowledge and skills for immediate use (p. 2).

Regardless of the terminology used, the structured approaches to learning lend themselves to an evaluation of the method of instruction to improve the created system and improve future teaching technologies. However, this document will focus on the specific issues regarding the evaluation of training programs. The reason is that the evaluation of education requires special considerations due to the large numbers of students who are taught, various socio-economic considerations, and the involvement of several levels of government. Conversely, the evaluation of training programs is a relatively defined issue in which a company's concern is whether or not the returns of a training program are worth the costs. This concern lends itself to an evaluation of the technologies used in the design, development, and dissemination of training programs. This method to evaluate training technologies may be divided into two categories: cost-benefit and cost-effectiveness.

COST BENEFIT VERSUS COST EFFECTIVENESS

There are different definitions of cost-benefit analysis and cost-effectiveness analysis. Cullen et al. (1978) define cost-benefit as the analysis of training costs in monetary units relative to the benefits derived from training in non-monetary terms, such as trainee attitudes, health and safety, etc. Cost effectiveness is defined as the analysis of training costs in monetary units as compared to benefits derived from training in monetary terms, such as production increases, production waste, etc.

Eicher (1980) defines the terms in the opposite manner, stating that a cost-benefit analysis works with variables that can be measured by the same standard: money. Thus, a rate of return can be established. On the other hand, cost and effectiveness do not have that same unit of measurement. Training effectiveness is generally defined with respect to (a) the time required for the average pupil to attain the level of knowledge considered normal on completion of a given course (as measured by objective testing), and (b) the percentage of pupils attaining a given level of knowledge. Consequently, cost-effectiveness analysis allows only the separate comparison of cost and effectiveness in two different situations. It can tell us if the cost of system A is higher or lower than system B or the same for its effectiveness. However, if the cost and effectiveness of A are both 10 percent lower than B, this analysis does not make it possible to conclude that both training solutions are absolutely equivalent. Eicher concludes that cost-effectiveness analysis only provides information that must be evaluated by the decision-maker; it does not provide objective and definite criteria of choice.

Another way to define a benefit is to consider the goals of training. For example, while some organizations will prefer to know a dollar return on an investment, still other organizations may find it useful to express outcomes in terms of increases in output per trainee (Cascio, 1989) or improved safety and health. Yet other uses of the terminology use cost effectiveness to imply cost benefit and the terms are used interchangeably. For the purposes of this document, cost-benefit analysis and cost-effectiveness analysis will be defined based on terminology adapted from Kearsley and Compton (1981): training system parameters, training benefits, and operational benefits.

Training system parameters are the lowest level of analysis for a training program. These parameters would include student learning capacity, availability of instruction, presentation capabilities, interactive capabilities, testing capabilities, management capabilities, development requirements, revision requirements, and so on. In essence, training system parameters are all the constraints that are present in the training system.

Impacted by the training system parameters are the training benefits. The training benefits include improved outcomes with student completion time, retention, student motivation and attitude changes, level of mastery, development and revision time, attrition/failure rate, stress level, and work relationships. Training benefits are often seen as an artifact of the training medium chosen, or at least impacted to some degree by the training medium. Although rarely quantified, training benefits are considered a valuable aspect of training that should be maximized. For example, Rosenthal and Mezoff (1980) give ten suggestions for before, during, and after training to maximize training benefits. Examples include designing the training to address issues of the organizational role, providing an opportunity for trainees to informally discuss work related issues, and providing certificates of completion at the conclusion of the training.

Impacted by the training system parameters and the training benefits are the operational benefits. Examples of operational benefits are increased job proficiency, higher production, reduced equipment failure, improved safety records, reduced customer complaints or service calls, increased sales, and so on. Operational benefits can be defined as improvements that can be seen as direct dollar returns to a company or as the goals the company has for the training program. Based on this terminology, a cost-benefit analysis will be defined as:

> The input into the training program based on the training system parameters as compared to the output from the training program based on the operational benefits.

A cost-effectiveness analysis will be defined as:

> The input into the training program based on the training system parameters as compared to the output from the training program based on the operational benefits and the training benefits.

These definitions will provide a criterion to categorize cost-benefit and cost-effectiveness models such that models may be examined. Before cost-benefit and cost-effectiveness models are considered specifically, it is helpful to look at models which assess only the costs of a training program.

Cost Models

The actual costs that go into a cost-benefit model vary from one analyst to another depending on the specific needs of an organization or the benefits to be assessed. The following cost models are two examples of cost assessment tools typical in the training literature.

Kearsley and Compton (1981). One very simple cost assessment model is defined by Kearsley and Compton (1981) called the *Resources Requirement Model.* The model is stated as:

$$\text{Total Training Costs} = \sum_{\text{across all stages of training cycle}} (\text{Total Personnel Costs} + \text{Total Equipment Costs} + \text{Total Facilities Costs} + \text{Total Materials Costs})$$

Although Kearsley and Compton do give examples of the kinds of costs that would fit into each category, they do not specify exactly what costs go into each category. This is reasonable in light of the fact that various training programs are going to incur costs specific to the training situation.

Another model by Kearsley and Compton is the *Life Cycle Model.* With this model, the analyst takes into account the entire life cycle of the training system. The total costs of a training system for a life cycle of months or years are given by:

$$\text{Total Life Cycle Costs (t)} = \text{Total Start-up Costs} + \text{Total Transition Costs (n)} + \text{Total Steady State Costs (m)}$$

where start-up costs are a one time expenditure, transition from a previous training method to a new training method is summed across a period of n months or years, and the operation steady state cost is multiplied across a period of m months or years.

Mirabel (1978). A model called the *Training Cost Model* used by the U.S. Civil Service Commission is discussed by Mirabel (1978). The *Training Cost Model* is a model to simulate the behavior or training costs under various specified conditions. This process uses a step-by-step procedure which enables one to either predict the cost of a proposed training course or, in the absence of accurate cost data, to reconstruct the cost of a course previously conducted. This model accounts for the trainees' cost, trainee travel and per diem costs, instructor costs, instructor travel and per diem costs, facilities costs, development costs, and production costs. Mirabel provides detailed worksheets to account for all costs in these categories.

There are other cost accounting models available (for example, Finkle, 1987), with each model being more or less encompassing. These models can be used to compare the costs of one training approach to another or to compare the costs of one training medium to another. However, although cost accounting models are necessary, they do not provide an accurate representation of the training investment since, obviously, there should be benefits to training. It is more accurate to compare both costs and benefits of a training program in order to assess the training investment as a cost-benefit ratio.

Cost-Benefit Models

The cost-benefit analysis was at one time a much simpler assessment tool. The methods used for training were simple (usually a person demonstrating how to do the task correctly) and the benefit was easy to see (increased productivity in terms of widgets produced). What has changed over the years is that the costs and benefits have become much more difficult to define. Costs now include assessing advanced and competing technologies, such as computer hardware and software costs, and benefits now include assessing less quantitative parameters, such as the value of reduced customer complaints.

The most basic cost-benefit model considers the projected cost input (C) and the projected benefit output (B). Then:

if $-C + B > 0$ the training program is acceptable.

if $-C + B < 0$ the training program is not acceptable.

Cohen (1985). An example cost-benefit model is proposed by Cohen (1985) in which the model requires the input of commonly obtainable data at the firm level. The model has been applied to data collected from interviews of training and personnel managers in ten manufacturing companies in the Malaysian Peninsula. The model is:

$$\sum_{t=1}^{T} \frac{COST_t + EDUR_t - p_t PVQW_t}{(r + rem)^t} = \sum_{t=1}^{T} \frac{(PVQW_t - EAFT_t)l}{(r + rem)^t}$$

where COST = cost of training, EDUR = individual earnings during training, EAFT = individual earnings after training, p = production rate of a trainee, PVQW = average earnings per qualified worker; r= retention rate, rem = rate of return to the employer, and t = years of training and work.

This model is directly applicable to manufacturing companies in which increased production is of primary concern. Stated another way, this model assesses the operational benefit of increased production. As Cohen defines benefit, it is the difference between the productive value of the graduated worker (which is approximated by taking the average earnings of a fully qualified worker) and the individual earnings paid to the newly graduated worker. These benefits are multiplied by a retention rate for the trained worker to account for the fact that not all benefits would materialize due to a portion of trained workers abandoning the company. Simply stated, the training program is considered a success if the post-training production rate meets a level that pays for the training and produces profit.

Cascio (1989). Although Cohen's model is designed to assess the operational benefit of increased production, there are operational benefits aside from increased production that a company may consider valuable. Many industrial training programs are designed to meet the operational goals of improved safety records, reduced customer complaints, or other such improvements in the company. An example of a more encompassing model is a utility analysis method described by Cascio (1989). The acceptance criterion to be applied to the training program is that the net present value (NPV) of the program must be greater than zero (as with the basic model). The equation is:

$$NPV = -C_0 + \sum_{t=1}^{n} B_t (1/1+i)^t > 0,$$

where C_0 is the cost of the program, B_t denotes the program benefits (in incremental cash flow) in period *t*, *i* is the discount rate, and *n* is the number of periods over which the program benefits last.

Any benefits from the program ultimately must be stated in terms of direct, measurable changes in the firm's cash flow. The discount rate reflects the fact that a dollar received in the future is worth less than a dollar received today, because today's dollar can be invested to earn interest. Discounting is used to express the present value of future cash flows in terms of today's dollars.

The model works in three phases. First the minimum required annual benefit is calculated for a period of time by solving for B. Calculation of the minimum annual benefits also specifies what the minimum annual net payoff (ΔU) from the training program must be with the formula:

$$\Delta U = n(d_t)SD_y$$

The second phase is to use break-even analysis to determine the minimum effect size, by solving for d_t, that will yield the minimum required annual benefit. The minimum effect size would determine the necessary performance improvement from each employee. Phase 3 uses meta-analysis results to determine the expected effect size and expected payoff from the intervention. The training program is considered successful if changes in the post-training period cash flow are close to the expected payoff within some degree of accuracy.

Summary of Cost-benefit Models

In general, researchers tend to use cost-benefit models as a summative form of evaluation; that is, the evaluation is conducted after the program is already in place. Summative evaluation can be useful in assisting decision makers in revising or modifying programs since repeated summative evaluation over time can be used for program modification (Mohr, 1988). However, summative evaluation is not often the most cost-effective way to choose a training approach or training medium. For example, if a group of employees work through company training and then the program was determined to be insufficient, the trainees would need to be retrained. Instead, the most cost efficient cost assessment method is often a formative method of evaluation; that is, an evaluation designed to assist program staff in understanding how to improve a program by providing information relevant to explaining the intent of the programs impact (Hawthorne, 1987). A formative method of evaluation would provide a program staff with the information to choose the most effective training approach or training medium before the trainees are exposed to the program.

Cost-effectiveness Models

As previously defined, a cost-effectiveness model takes into account the training system parameters, the training benefits, and the operational benefits. The level to which any one cost-effectiveness model includes all three sub-components varies.

Cullen et al. (1978). Cullen et al. (1978) present an example of a cost-effectiveness model that combines three cost-accounting models that have been utilized in the training profession. The model is broken into training program costs and training benefits. Costs are split into three groups; fixed, variable, and total. Fixed costs are costs that do not vary even though the number of trainees, training time, or training program development vary. Variable costs are costs that change as the number or trainees, training time and training program development vary.

Major costs for the training program are the expected standard costs: training development time, materials, etc. To evaluate the training returns, Cullen et al. (1978) state that one must "detail the competencies" of the worker and evaluate them. The following outline summarizes the procedures for assessing the returns of a plastic-extruder machine operator for both a structured and an unstructured training program.

1) Production Task Performance
   - A. Trainee has reached a job competency via training (structured or unstructured training program).
   - B. Trainee is satisfied with his or her training and the job.

2) Collect Data on Task Performance Returns
   - A. Costs of reproducing copies of developed training program.
     1. Time (to reach competency, production curtailed, startup)
     2. Production Rate
     3. Performance Test
     4. Product Quality
     5. Raw Material Usage
   - B. Measurement of trainee attitudes toward his or her training and the job.

3) Monetary Value of Returns
   - A. Convert trainee performance data to a monetary value.
   - B. Returns of structured training program and unstructured training program are totaled.

Two training benefits are listed in the outline that were included in the model: trainees' satisfaction with the training and the job; and, time to reach competency. Also included, but not listed in the outline provided by Cullen, et al. (1978) is a comparison of levels of production worker competency by time intervals between the two training methods; the total development costs and returns of two training programs; the production losses of the structured training program versus the unstructured training program; the reactions to production problems (malfunction performance test); and, the attitudes of the trainees toward their training, trainer, and job.

Cullen, et al. suggest that this model could be used to find the most appropriate training program by applying the model to two different programs. However, there are a few shortcomings to the model when used for this purpose. First, the model does not take into account that one training approach may have benefits that another training approach does not possess. In this case, the effectiveness of a training approach is not being fully explored. Second, although the authors advise to find the monetary equivalency of non-monetary indices whenever possible, there is no method in place to quantify each benefit as it relates to one training approach over another. The following cost-effectiveness model addresses this issue.

Kearsley and Compton (1981). As stated previously, using cost models, such as the *Resource Requirements* model, to compare two training programs does not consider the performance issues (or other benefits) associated with two different training approaches. A new training method or technology may achieve better performance, higher motivation, etc., in which case a training approach that at first glance appears more costly could save money in the long term. A model which encompasses performance variables is the *Benefits Model* as defined by Kearsley and Compton (1981). The intent of the model is to show the causal relationship between training system parameters, training benefits, and operational benefits as defined previously. The causal relationships can then be used to help quantify the worth of a benefit to a training program.

Figure 1.1 illustrates the structure of the model. Each node represents either a parameter or benefit of the training program. Each connecting line represents a causal link, or relationship, between variables. At the bottom of the tree are the major parameters of the training system. At the next level are training benefits which are affected by the training system parameters. At the third level are the operational benefits which are derived from the training benefits. A unique set of system parameters, benefits, and causal relationships are developed for each training situation.

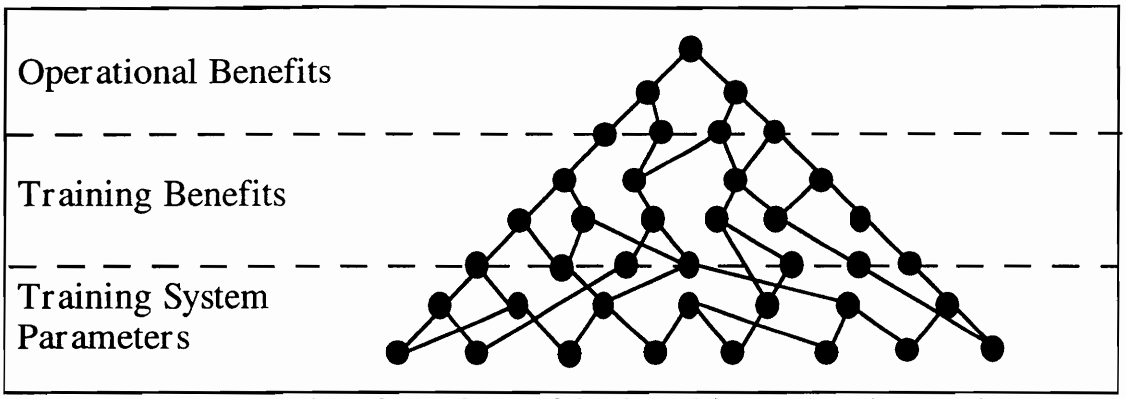


Figure 1.1. Representation of *Benefits Model* (adapted from: Kearsley and Compton, 1981).

As Kearsley and Compton (1981) explain, once the relevant parameters, benefits, and causal relationships have been defined, values are assigned to the strength of the causal links. Although Kearsely and Compton suggest using a range of values from -1 to +1, the examples they provide seem to contradict the idea of using a "range" of values. Instead, it appears that they use only -1 and +1 as the values. For example, if multimedia capability slows down revision time then a value of -1 is applied to represent a negative affect. Likewise, if modularized design speeds up revision time then a value of +1 represents a positive effect. The assigned values can be the result of expert judgment or decisions based on the current literature. Kearsely and Compton represent this model as:

$$B_j = \sum_{i=1}^{k} V_i B_i$$

where B_j is the j^{th} benefit and V_i is the value associated with B_i, the i^{th} benefit or system parameter. Each benefit is represented by the summation of a set of k benefits which affect it. As this computation is made for each successive benefit, B_j's of lower levels become the B_i's of higher levels.

For example, consider the scenario in which a multimedia and lecture training program are being compared. Assume that the impact of the training medium on the operational benefit of increased trainee performance is being considered. Further assume that expert opinion has determined that the variables which affect increased training performance for a multimedia training program are the training system parameter of a large trainee population (B_{i1}), the training benefit of trainee motivation (B_{i2}), the training benefit of revision time (B_{i3}), and the training benefit of time to complete training for each trainee (B_{i4}). Further assume that expert opinion has determined the following:

Multimedia capability is well suited to a large trainee population (+1).

Multimedia capability increases trainee motivation (+1).

Multimedia capability slows down revision time (-1).

Multimedia capability decreases time to complete training (+1).

Based on this information,

B_j (increased trainee performance) = (+1 x B_{i1} (large trainee population)) +
(+1 x B_{i2} (trainee motivation)) +
(-1 x B_{i3} (revision time)) +
(+1 x B_{i4} (time to complete training)) = +2

Based on the formula, the value of the operational benefit of increased trainee performance for the multimedia training program is +2. This value can then be compared to the calculated value of B_j for a lecture training program. The training medium which has the highest value is, in theory, the most suitable training medium for the job. Once all the B_j values are calculated and the costs for the training program are determined, an item for item comparison can be made and the experts can determine which training program is the most cost effective.

Furthermore, Kearsley and Compton explain that the k values associated with any particular benefit need not combine to unity since the specified benefits may not account for the higher level benefit. For example, a set of benefits, B_1, B_2, and B_3 may only account for half of benefit B_4; the remaining half being unknown or due to factors outside of the training domain. As a consequence of allowing partial causality, the model exhibits attenuation of effects as the distance in the tree increases. Thus, a system parameter or benefit may exert strong influence at a particular level but this effect will become weaker at each higher level in the tree. On the other hand, at a higher level, factors may become strong due to the accumulation of factors that comprise it.

In theory, this model has an advantage over the cost-benefit models discussed previously in that it includes training benefits and is therefore designed to undertake a more thorough analysis of the outcomes of a training program. Also, it is more thorough than the cost-effectiveness model proposed by Cullen et al. (1978) for the purpose of evaluating the effectiveness of two or more training programs since it includes a method to assign a value to benefits as they relate to a particular training approach or medium. Furthermore, the theory behind the *Benefits Model* is that it considers the relationship that benefits have on each other, which helps assess the value of a benefit. Finally, the *Benefits Model* is designed as a formative evaluation method to assess the impact of two or more training programs before choosing one program as the most effective. This is in comparison to a summative evaluation to assess the impact of one training program after it has been fielded.

In relation to this last fact, the *Benefits Model* is, in theory, a <u>predictive</u> assessment method. This is a valuable tool since typical formative evaluation requires some design and development in order to collect data and compare the programs, thereby being more costly. A predictive approach on the other hand requires gathering cost and effectiveness information on two possible systems before any system is implemented so that only the training system that is determined to be the most cost-effective would ever be developed. Therefore, more resources could be focused on design, development, and dissemination of the one best system.

<u>Summary of Cost-effectiveness Models</u>

When an analyst is choosing a method to evaluate a training program, there are several reasons for and against using each kind of model. The advantage to using a cost-benefit model is that for many situations it will be the least time consuming method of analysis. Costs of the training program will be documented for the most part, and the benefits are typically easily quantifiable. Conversely, there are two disadvantages of using the cost-benefit model. One is that it does not account for training benefits. This can be a serious downfall if the training benefits are contributing to the operational benefits of a training program and then not being maintained in the training program. A second disadvantage to using a cost-benefit model is that it tends to be used in a summative manner (after the program has been disseminated) to assess productivity or final net gain. The problem with the summative form of cost-benefit evaluation is that the success or failure of the program can not be attributed to any particular aspect of the program, in which case the method to improve the program is not specified by analysis. In the case of an unsuccessful or inadequate training program, a summative method of evaluation increases the overall cost of the program by not informing the analyst of the problem until a group of trainees have already completed the discontinued program.

The disadvantage to using a cost-effectiveness method of evaluation is that it may be initially more costly to assess both the training and operational benefits of a training program, especially those that are not easily quantifiable. However, the advantage to evaluating effectiveness is that if the benefits of the program are identified and considered valuable, they can be maintained. Likewise, if an expected benefit is not being produced in training, an effort can be made to correct the problem before program dissemination. A further advantage to cost-effectiveness evaluation is that it lends itself well to formative evaluation; that is, evaluation before the program has been disseminated in an effort to improve the program before fielding it. Also, formative evaluation can be used to choose the most effective program between two or more training approaches. Therefore, when making a comparison between programs, a formative method of evaluation will be considerably less expensive in the long run since only the most effective training program or medium will be fully developed and employed.

However, a more valuable aspect of a cost-effectiveness model is the idea of a predictive approach. The *Benefits Model* requires that training system parameters, training benefits, and operational benefits be identified, classified and quantified via expert opinion in order to predict the cost-effectiveness of two or more training programs. This would be in contrast to a traditional formative approach which requires some design and development in order to collect data and compare the programs, which in the long term is more costly. Nonetheless, a problem exists in that the model has not been appropriately validated. What is needed is a comparison between the outcome of a predictive application of the model to a traditional formative evaluation approach in which all cost and performance data are collected for two or more programs.

THE IDENTIFICATION, CLASSIFICATION, AND QUANTIFICATION OF VARIABLES FOR VARIOUS MEDIA

As stated, a major advantage to using a cost-effectiveness model is that it would serve to compare the relative benefits of a number of different training approaches or delivery systems (Kearsley and Compton, 1981). This ability to choose between media can be valuable for the reason that training media can be very complex and choosing a training medium is a difficult process. Furthermore, although the training medium chosen is often dictated by the resources present in the training budget, a training approach that at first glance appears more costly could save money in the long term. The following sections review the costs and benefits of various training media.

Costs of Media

As one would expect, almost every book designed to guide a program designer through the process of developing a training program discusses the issue of choosing a training medium that is most appropriate for the goals of the program. When cost is discussed regarding the selection of a training medium, the issue addressed is whether the value of the course justifies the cost of the medium or media considered (for example, Reynolds and Anderson, 1992). The fact that this issue exists further supports the need for a cost-effectiveness model to help determine which training medium to choose for a particular training program.

The issue of which training media will require the most resources over the life of a training program is rarely discussed. This is not to say that those writing training development handbooks are not thorough, rather, it is simply that technology is advancing so quickly that a static guideline for choosing media is not feasible (Gordon, 1994). Therefore, the program analyst is left to consider not only the goals of the system and which media will meet those goals, but also the most efficient and effective use of resources. The problem is that, regarding media, a method to choose an efficient and effective use of resources is not clear.

Determining costs is often seen as a product of determining resources; that is, if an analyst knows the resources then he or she can determine which media will help meet the performance goals of the program, then see which medium fits the resource budget. This reasoning is faulty for three reasons. First, it may be the case that there is more than one medium that will meet the performance goals and fit the budget. Second, it may be the case that of two competing media, one obtains only a small increase in performance but costs considerably more. In this case, remedial training to bridge the gap in performance could actually be less expensive than the more costly medium. Finally, it may also be the case that the analyst does not understand all of the benefits possible from a training medium, and how benefits interact with each other to help meet the goal of the system.

The question then is how an analyst should determine what the media costs are? A possibility is to apply the resources information gathered to the overall training program, and also to each training medium. For example, the following list of resources is recommended by Gordon (1994) to be evaluated prior to selecting a training medium.

- Facilities (space, power, lighting, etc.)
- Financial resources available for the project
- Time
- Available personnel and their skills
- Equipment
- Cost of resources at each prospective development site
- Similar or usable existing programs
- Storage and delivery services
- Dissemination resources
- Costs of delivery or dissemination resources.

Gordon recommends that the information be gathered by interviewing knowledgeable people and performing site visits. Furthermore, she states that the analyst should be sure to obtain estimates of future resources as well as currently existing resources. These same categories that are applied to overall training consideration can also be applied to a training medium to get an impression of the types of costs to be considered. For example, a stand-alone multimedia computer system would require a different type of facility than a lecture medium; the total financial resources would assuredly impact the medium chosen; time resources for development would be different for a lecture format versus a multimedia format; etc.

However, this list is not complete enough to estimate all costs. It requires some stamina on the part of the training analyst to obtain all the sub-categories of costs. For example, when examining the costs of time resources for a lecture presentation, one may need to consider several costs of the lecturer's time: total curriculum hours per year, average number of instructors per session, total in-class instructor hours per year, total hours of instructor preparation, etc. (categories taken from Mirabel, 1978). In addition to these costs, one may need to also consider the lecturer's salary, travel time, travel per diem, and so on. It is easy to see that the categories of information for "time" are many just to consider the lecturer's time, and the lecturer's time is only one sub-category of time. There is also the time to design, develop, and disseminate the program, along with corresponding sub-categories that must be considered.

There are other more complete cost assessment methods. For example, Mirabel's (1978) cost assessment method discussed previously shows a clear amount of effort to consider all the costs, and many of the categories can be used to assess the training medium. Realistically, there will probably never be a strict guideline to determine costs of media since the costs are unique to each training program. However, some guidelines that are specific to the categories of costs would help an analyst in an attempt to be thorough.

Benefits of Media

As previously stated, there are two categories of benefits that are being addressed; training benefits and operational benefits. Training benefits have been defined as an derivative of the training medium chosen, or at least impacted to some degree by the training medium. Operational benefits have been defined as improvements that can be seen as direct dollar returns to a company or as the goals the company has for the training program.

Identification of benefits. The first step to comparing benefits of various media is to identify the benefits of each training program. Identification of the operational benefits of a training program is somewhat less elusive than identifying the training benefits. The operational benefits can be identified by definition – that is, is the training program buyer getting what he or she paid for, or are the training program goals being met? For example, a lecture-only style training program is not optimal for training an individual of the intricacies of piloting an airplane; therefore, a lecture-only program would not provide that particular operational benefit. This is not to say that the lecture-only program would not provide some training benefits. It simply indicates that the operational benefit would not be met, in which case the training program is considered a failure.

The identification of training benefits is more difficult. Kearsley and Compton (1981) provide training benefit examples of student completion time, retention, student motivation and attitude changes, level of mastery, development and revision time, and attrition/failure rates; however, this list is not complete. One method to determine the benefits of a training program is to consider what variables would affect the effectiveness of a training program. For example, if high trainee motivation positively affects the outcome of a training program then that variable is a benefit of the training program.

After an extensive literature review, Tannenbaum, et al. (1993) provide a list of variables that affect training effectiveness as shown in Table 1.1. The list is somewhat confusing since training effectiveness itself is one of the major categories. However, their concept is that the other 10 categories of variables affect the training effectiveness variables.

This list shows some of the characteristics of the training program that will impact the effectiveness; however, the list does not consider all benefits that affect efficiency, or in other words, would impact a cost-effectiveness ratio. For instance, a variable of time to complete training is not considered. There are also other short-comings of using the variables in Table 1.1 as a list of training benefits. First, the list is too cumbersome to consider the variables and the relationships between them. Secondly, these variables as benefits will not always be present in every training program, since several factors will affect whether or not these variables are provided by a training program (Salas, Burgess, Cannon-Bowers, 1995). Finally, each medium will vary in the kind and amount of training benefits it will provide.

Table 1.1. Variables in the training effectiveness model (from Tannenbaum, et al., 1993).

Individual Characteristics (Pre-training)
- Abilities
  - • cognitive ability
  - • psychomotor ability
  - • learning rates/trainability
- Attitudes
  - • commitment
  - • intent to remain
  - • career planning
  - • job satisfaction
  - • reactions to previous training
  - • coworker/teammate relations
- Self-efficacy
  - • physical self-efficacy
  - • cognitive self-efficacy
  - • task-specific self-efficacy
- Personality
  - • locus of control
  - • ego strength
  - • need for achievement, affiliation
  - • conformity
- Demographics
  - • family history
  - • age
  - • gender
  - • education
- Experience
  - • tenure/experience with company
  - • with task
  - • with previous training

Organizational/Situational Characteristics (Pre-training)
- Organizational Climate
  - • participatory versus centralized
- Trainee Selection/Notification Process
  - • voluntary versus mandatory attendance
  - • reward vs. punishment
  - • communication medium, accuracy
- Purpose of Training
  - • maintenance vs. advancement
- Task or Job Characteristics
  - • task complexity
  - • task type
  - • task difficulty
  - • feedback
- Organizational History
  - • management-labor relations
  - • growth/decline
- Organizational Policies, Programs, & Practices
  - • other human resource practices
  - • other company practices

Trainee Expectations
- Trainee Performance Expectations
- Training Expectations
  - • training format
  - • challenge
  - • degree of interactions
  - • focus of content

Pre/During Training Motivation
- Motivation to Attend
- Motivation to Learn

Training Program Characteristics
- Training Needs Analysis
  - • accuracy of need identification
  - • involvement of potential trainees
- Training Method/Process
- Use of Training Principles
- Training Content
- Instructor characteristics
- Use of Technology

Expectation Fulfillment
- Perception/Expectation Match

Programmed Interventions
- Relapse Prevention
- Transfer Support Programs

Training Effectiveness
- Training reactions
  - • training relevance/perceived value
  - • affective responses/happiness index
- Learning
- Training Performance
- Job Performance
- Results/Organizational Effectiveness

Post-training Individual Characteristics
- Attitudes
  - • commitment
  - • intent to remain
  - • job satisfaction
  - • coworker/teammate relations
- Ability
  - • task specific ability
- Self-efficacy
  - • physical self-efficacy
  - • cognitive self-efficacy
  - • task-specific self-efficacy

Post-training Motivation
Motivation to Transfer and Maintain

Organizational/Situational Variables – (Post-training)
- Transfer Environment
  - • supervisor support
  - • co-worker support
  - • resource availability (time, equipment)
  - • workload
  - • job security
  - • authority/autonomy

Organizational Culture
  - • openness to innovation/risk taking

Many books which are created to teach training program design and development list various benefits of media as advantages. As an example, Table 1.2 shows the following list of advantages for lecture and computer-based training (CBT).

Table 1.2. Possible advantages of lecture and computer-based training programs (adapted from Gordon, 1994).

Advantages of Lecture
1. Lecture is relatively easy and inexpensive to develop.
2. Lecture material is easy to modify at any time before presentation, making it good for information that changes frequently.
3. Lecture can be somewhat adaptive (more than text or videotape). It can be modified according to the needs of the class; However, this can still only be done for the*whole* class and not for individuals to any great extent.
4. Lecture has the potential to be interactive, students can ask questions, and be asked by the instructor to generate ideas or answers.
5. Lecture can be relatively motivating, certainly more so than reading text.
6. The instructor has control over the material to which the learner is exposed.
Advantages of Computer-based Training
1. There is an improving ease of development.
2. There is a potential for a high level of interactivity.
3. There is potential for adaptivity.
4. It is convenient for instructional delivery.
5. It is motivating.
6. There is a high speed of response and feedback for the trainee.
7. Computer-based multimedia supports course management.
8. There is potential for insuring topic coverage (there is no instructor to forget something).
9. Training time tends to be significantly shorter.
10. Low-cost relative to on-the-job training.
11. There is potential for practice.
12. There is potential for testing.
13. Trainees usually have a great deal of learner control.

Kearsley (1983) provides a somewhat different list of benefits for computer-based training. According to Kearsley, some of the advantages of CBT are that it can

1. provide increased control over training activities in terms of improving utilization or completion of learning materials, increasing standardization of instruction, or monitoring of student progress.
2. reduce resource requirements. An example is providing instruction at branch offices or field sites.
3. give individualized training, allowing each student to learn at a speed and in a fashion most suited to his or her particular learning style.
4. provide instruction "on demand" instead of when instructors and employees can schedule time and facilities.
5. reduce training time due to the individualized instruction.
6. improve job performance or eliminate job performance problems by providing a more active (as opposed to passive) method to learning.
7. be convenient, insofar as employees already use computers for their jobs.
8. increase learning satisfaction since learners typically find CBT more motivating.
9. reduce development time.

What should be noted about the lists of advantages (benefits) is that they are merely potentials for each medium. For example, lecture may have the potential to be interactive, but if students do not ask or answer questions, the lecture training program will not be interactive. The implication is that an analyst cannot rely on the literature to determine what training benefits to expect out of each medium; therefore, not all training benefits can be added to every cost-effectiveness model.

Classification of benefits. Aside from the problem of the identification of the training benefits that should go into a cost-effectiveness model, is the problem of the classification of the benefits as they would go into the *Benefits Model.* Classification refers to the determination of how the benefits affect each other. For example, a training program that provides the benefit of being highly motivational may in turn provide (or at least augment) the benefit of taking less time to complete. The problem is, of course, that stating a relationship based on the literature is not reliable since relationships will vary depending on characteristics of the trainees in a training program, organizational characteristics, the training media, etc. (Salas, Burgess, Cannon-Bowers, 1995).

Looking again to the work by Tannenbaum, Cannon-Bowers, and Salas (1993), the classification of the relationship between benefits can be considered similar to the relationship of the variables that affect training effectiveness. Their model of training effectiveness which considers the relationship of variables can be seen in Figure 1.2. The model is fairly simple and does point out several key relationships; however, the empirical examination of these relationships of variables as benefits is not a simple issue. To use this model of the relationship of variables that affect training effectiveness brings about the same problems that were stated for using these variables for the identification of benefits: the list is incomplete, cumbersome, and each variable (benefit) is not always present in every training program.

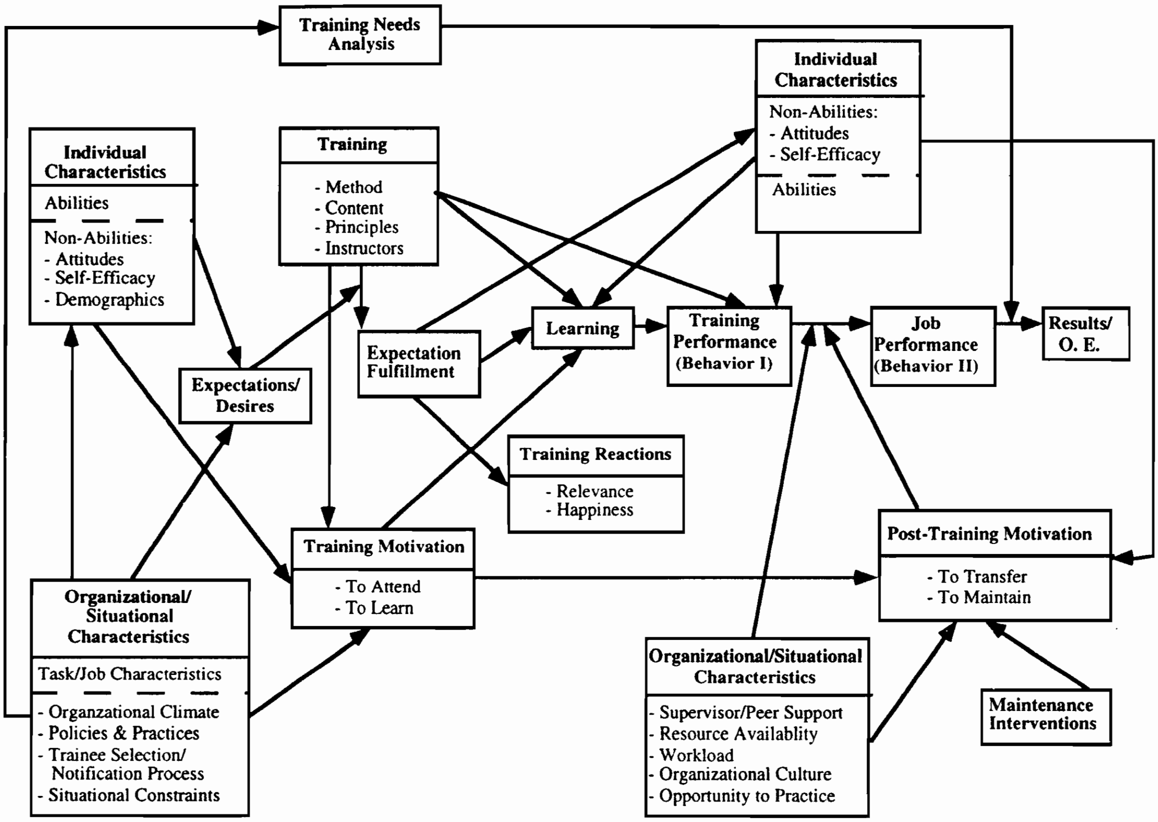


Figure 1.2. Model of training effectiveness (Tannenbaum, Cannon-Bowers, Salas, 1993).

Quantification of benefits. Quantification of the training benefits refers to Kearsley and Compton's (1981) -1 to +1 weighting scheme for each benefit. In other words, how should the weighting scheme be applied? Although it is not made clear in their paper, it would appear that Kearsley and Compton use only the numbers -1, 0 (does not apply), and +1. The application of the value system would work as such:

	Benefit	
Media	Revision Time	Interactive Capabilities
Multimedia	-1	+1
Lecture	+1	-1

The logic for the numbers is that, for example, revision time is slower for a multimedia training program and therefore a -1 would be assigned. On the other hand, revision time would be faster for a lecture training program and therefore a +1 would be assigned. However, the lack of guidelines for quantification becomes apparent when assigning values for a benefit such as interactive capabilities. For example, although multimedia has very high capabilities, multimedia is only interactive if designed to be so. Likewise, lecture has the possibility to be interactive if trainees are willing to participate and the instructor is good at soliciting participation; however, in practice, lecture is often not interactive. In these circumstances the question is how the numbers should be applied. Furthermore, there are varying degrees of interactivity, in which case, a simple three point numbering system may not be adequate.

Summary of the Variables for Various Media

Concerning the identification of costs and benefits, and the classification and quantification of benefits, there are no clear answers or literature for an analyst to consult; However, if guidelines were available, they would provide little assistance since they will not apply to all situations or will not be kept current with technology.

In theory, the *Benefits Model* is a predictive cost-effectiveness evaluation method; however, it is has not been validated as a predictive tool to aid program designers in choosing between two media. An important first step towards the validation of this predictive tool would be its application to choose the most cost-effective of two competing media. The outcome could then be compared to a traditional formative evaluation approach in which performance data is collected and analyzed to determine the most cost-effective training program. The predictive application of the *Benefits Model* should include the identification, classification, and quantification of the benefits for the model. Such an effort would provide either validation for the model or possible insight into the kind of elaboration necessary to create a cost-effectiveness model that would work predictively.

Lecture and computer-based multimedia training programs were designed and developed in the Vehicle Analysis and Simulation Laboratory at Virginia Polytechnic

Institute and State University. Using these training programs, it was proposed that an evaluation of the *Benefits Model* be conducted.

LECTURE AND COMPUTER-BASED MULTIMEDIA DROWSINESS DETECTION TRAINING PROGRAMS

A series of studies on sleep deprived drivers was conducted at the Vehicle Analysis and Simulation Laboratory at Virginia Polytechnic Institute and State University. While running the experiments, it was discovered that the experimenters could estimate the level of drowsiness of the drivers by direct observation of the drivers' faces as seen on a video monitor. Because of this anecdotal phenomenon, a subsequent experiment was conducted (Wierwille and Ellsworth, 1995) to determine the scientific validity of the concept of observer rating of drowsiness. Based on the results of that experiment, which showed that observer ratings were consistent and reliable, it was proposed that a training program be developed to teach individuals how to rate the level of drowsiness of their colleagues to improve team safety. Furthermore, in an effort to select the most effective training medium, it was determined that both lecture and computer-based multimedia formats would be developed for formative evaluation.

The design model utilized in the development of the training programs was that outlined by Gordon (1994). The model is a three-step process which includes a front-end analysis, a design and development phase, and system evaluation. The model is shown in Figure 1.3. With the concurrent design and development of the Drowsiness Detection Training programs, relevant information from the training programs were available for an evaluation of Kearsley and Compton's (1981) *Benefits Model.*

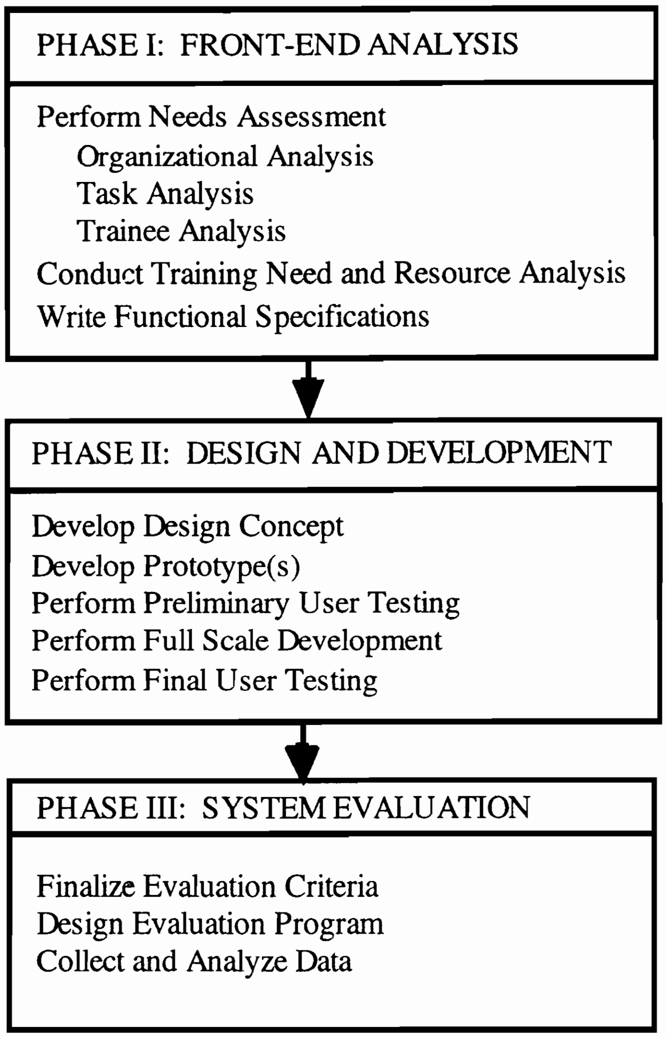


Figure 1.3. Design model for training and performance support systems (adapted from Gordon, 1994).

THE PROBLEM STATEMENT AND PURPOSE OF STUDY

A validated predictive cost-effectiveness model would serve numerous purposes. First, it would serve to compare the relative benefits of a number of different training approaches or delivery systems (Kearsley and Compton, 1981). Second, it would allow training media to be compared in a predictive evaluation manner, rather than in a traditional formative evaluation manner, which is a more cost-effective method to finding the most effective media for the training dollar. Third, a predictive model would encourage the training program designer to consider what benefits are inherent in a program and what benefits need to be designed into the program. The problem is that although a predictive cost-effectiveness model has been proposed, all elements of the model and the assessment of those elements have not been clearly defined, nor validated.

To validate a predictive (as opposed to a traditional formative or summative) cost-effectiveness model, a study was conducted to evaluate Kearsley and Compton's (1981) *Benefits Model*. The costs that were input into the model were the costs as they applied to the design, development, and dissemination of two training programs on the topic of teaching individuals how to detect the level of drowsiness of their colleagues during team operations. The benefits of the training programs were identified, classified, and quantified as they applied to two media: lecture and computer-based multimedia.

For these training programs, the training system parameters and operational benefits were identified from the front-end analysis of the training programs. The training benefits were identified as they applied to the program. The classification of the relationships between training system parameters, training benefits, and operational benefits were defined based on expert opinion. Quantification concerned the assignment of values (-1 to +1) based on expert opinion. The costs to design, develop, and disseminate the training programs were determined. Then, based on all information present, experts determined which of two training programs would be the most cost-effective to disseminate.

To determine the accuracy of the *Benefits Model* as a predictive assessment tool, the same identified variables were evaluated from a traditional formative evaluation approach. Data were collected from trainees who participated in one of the two training programs and a determination of the most cost-effective training medium was made. The data collected in the traditional formative evaluation approach were then compared to the experts' ratings and choice of training programs.

RESEARCH INVESTIGATION

This research was conducted to investigate the scientific validity of a predictive cost-effectiveness assessment tool. Based upon the design of two training programs (lecture and computer-based multimedia formats), expert opinion, and performance data collected from the two training programs, the variables for a predictive cost-effectiveness model were identified, classified and quantified, and experts determined the most cost-effective training approach. The outcome of the predictive model was then compared to a traditional formative evaluation approach in which the most cost-effective training program was determined from performance data collected. From the results of this investigation, information was gathered to provide insight into the tasks of identifying, classifying, and quantifying training variables, evaluate the value of the *Benefits Model* as a predictive cost-effectiveness assessment tool, and determine the necessary directions research must take to enhance the usability of such a model.

It was not the purpose of this research to determine the most cost-effective way to use each training medium. Instead, the purpose was to determine the potential of the *Benefits Model* as a predictive approach to determining the most cost-effective training program as the programs were intended to be designed, developed, and disseminated.

CHAPTER 2

THE VARIABLES IN THE COST-EFFECTIVENESS MODEL

COSTS INCLUDED IN THE MODEL

The resources necessary for a training program were divided into three major categories: design, development, and dissemination. Each category was defined as follows:

Design - all resources necessary to perform a front-end analysis (needs assessment, training need and resource analysis, and develop the functional specifications) and develop the design concept, including storyboards of the training program.

Development - all resources necessary to prototype, perform optimization, perform initial usability testing, and perform formative evaluation.

Dissemination/Facilities - all resources necessary to put the training program into the environment of the trainee, including facilities, personnel, equipment, storage and delivery, etc.

IDENTIFICATION OF VARIABLES

Which variables should be considered for a cost-effectiveness analysis should depend on the particular training program and the reason for the creation of the program. From this viewpoint, the training system parameters and the operational benefits were determined from the front-end analysis of the training program development and by individuals related to the development of the program. This was a somewhat informal process, meaning that the training system parameters and operational benefits were determined in the normal course of program development, without the goals of outlining what the cost-effectiveness parameters and benefits would be. Instead, the designer thought in terms of "specific behavioral objectives" of the training program and "program constraints due to available equipment and tools" as the design model would dictate. Therefore, some of the parameters and operational benefits were not stated in terms that would fit easily into a cost-effectiveness model.

The identification of the training benefits was a more difficult process since the individuals who wanted the training program developed did not specify this category of information. As such, the training benefit variables were determined by the designer of the training program with the help of individuals in the area of driver fatigue studies. For example, the program designer did not specify that it would be good to design a multimedia training program in a manner which would promote shorter revision time; indeed, the designer did not think in terms of revision time at all, but thought in terms of concentrating on how to design a good interface for the multimedia program through any means possible.

Identification of Training System Parameters

The sponsors of the program requested that two training programs be designed, developed and tested, and the one with the best performance be disseminated. Furthermore, the buyers suggested that the resources for dissemination of the training program to the users not be considered – resources for dissemination would be allocated after the sponsors knew which training program would solicit the best performance. Table 2.1 shows the "constraints" that were determined during the front-end analysis (Neale and Wierwille, 1995a).

As can be seen in Table 2.1, there was a fairly large number of constraints on the training program development. However, it should be noted that there are no constraints given for the dissemination of the program. The program sponsors were not concerned with dissemination, but only with finding an effective training medium. Nonetheless, even though there are no constraints given for the dissemination of the training program, there are still too many to evaluate effectively. Therefore, from the list given in Table 2.1, a list of training system parameters can be created which is a subset of the constraints and considered representative of the constraints of the training program.

Table 2.1. The constraints of the Drowsiness Detection Training Program (adapted from the front-end analysis).

1) Constraints relevant to the training program development:
  a. What financial resources are available?
    • Development sponsored by the Federal Aviation Administration through NHTSA.
  b. What human resources are available?
    • The principal investigator, one graduate student working on a full time (20 hours per week) assistantship, and trial subjects. There are 11 months total available for the project.
  c. What equipment and other tools are available?
    • Equipment provided by the cooperative agreement, equipment and laboratory facilities available in the Virginia Analysis and Simulation Laboratory at Virginia Tech.
  d. What are the time limits?
    • Grant project time is 11 months total. Although there is no limit on the time for the trainee to complete the training program module, it can roughly be set at not more than four hours total. A shorter training time would probably be better.

2) Constraints relevant to training program delivery:
  a. When do the trainees need to be trained?
    • As soon as possible.
  b. How many trainees need to be trained?
    • Possibly thousands.
  c. Where do trainees need to be trained?
    • In the towns and cities in which they live, possibly at flight schools or local government offices.
  d. What is the frequency of tasks being trained?
    • At this stage, it is not confirmed whether trainees will need follow up training, but at this point, only one time training is being considered.
  e. What is the time span from training to actual job performance?
    • Probably days, most likely not more than a couple of weeks.
  f. Are there any other constraints on method of delivery?
    • Needs to be as convenient and as inexpensive as possible considering the large numbers that need to be trained. <u>However</u>, it is requested that at this time the main goal of the research and program development is to concentrate on the most effective method of training, and not the monetary concerns of disseminating the developed program.
  g. Will instructional content change over time or is it stable?
    • Probably fairly stable. The only possibility is further refinement in the technique as research uncovers more information.
  h. What is the estimated life span of the training program?
    • Years.
  i. What financial resources are available for delivery?
    • This is not being viewed as a concern for this phase of the project. It has been stated that the concern of this project is finding the best method of training and that resources for dissemination will be based on the best method of training.
  j. What human resources are available for delivery?
    • This is not being viewed as a concern for this phase of the project. It has been stated that the concern of this project is finding the best method of training and that resources for dissemination will be based on the best method of training.

k. What equipment is available for delivery?

• This is not being viewed as a concern for this phase of the project. It has been stated that the concern of this project is finding the best method of training and that resources for dissemination will be based on the best method of training.

l. What are the needs or biases of management or other personnel involved?

• Not known

3) Constraints imposed because of certain characteristics of the trainees:

a. Who are the trainees and what are their characteristics?

• People with diverse backgrounds and interests. The minimum education level is high school plus specific trade training. Trainees have a general knowledge of operation procedures of air traffic systems with specific training in aviation. Knowledge directly related to the ability to identify the level of drowsiness of colleagues is not assumed.

b. Is there anything in trainees background or attitudes that would constrain the program?

• It is unknown whether trainees have any biases that would constrain the program.

c. What are the relevant knowledge and skills of the trainees?

• Trainees have a general knowledge of operation procedures of air traffic systems with specific training in aviation. Knowledge directly related to the ability to identify the level of drowsiness of colleagues is not assumed.

From the list of "constraints" for the Drowsiness Detection Training program, the list of training system parameters that adequately represent the program are shown in Table 2.2.

Table 2.2. Training system parameters for the Drowsiness Detection Training program.

Training System Parameters
Financial Resources
Human Resources
Equipment and Tools
Time Resources
Large Number of Trainees
Many Locations for Training

Identification of Training Benefits

Because training benefits were not specified either by the sponsors or the front-end analysis, there were more training benefits to be considered. The list of training benefits (Table 2.3) was determined based upon what would be considered assets of the system that would make one medium more effective than another for the Drowsiness Detection Training Program.

Table 2.3. Training benefits for the Drowsiness Detection Training Program.

Training Benefits
Time to complete training session with practice
Number of remediation sessions necessary
Time to revise training program (after formative evaluation)
Trainee motivation
Self-efficacy
Convenience of attending training

Some clarification of this list of training benefits is in order. First, time to complete training session with practice does not include any remediation that may be necessary. This will remove the confound between time to complete the training session and the number of remediation sessions, since increased remediation naturally means increased time. Second, for the lecture training program, the time to complete training is dependent to some degree on the presentation style and level of preparation of the lecturer, and the dynamics of the class participants. These issues are artifacts of the lecture medium. Therefore, since the resources for this project do not allow multiple lecturers to be compared and contrasted, an assumption is made that the lecturer is representative of the lecturer population, and the generalizability of the study is restricted to the level to which the one lecturer represents the population.

A final clarification for the list of training benefits is a definition for "trainee motivation." For the purposes of the experiment, trainee motivation is defined as the level to which the trainee wants to learn the material as it is presented by the respective training medium. Stated another way, did the particular training medium encourage the trainee to learn about drowsiness detection training?

Identification of Operational Benefits

The operational benefits of the system are listed in the front-end analysis as "Post-training Goals and Objectives." Table 2.4 shows the operational benefits that were determined during the front-end analysis (Neale and Wierwille, 1995). The post-training goal is stated generally, whereas the specific behavioral objectives are more precise.

Because of time constraints, it was not feasible to evaluate all the operational benefits. Of the specific behavioral objectives, the single most representative objective which could be used to determine the effectiveness of the training program is the last objective, "Demonstrate the ability to rate the drowsiness level of a person via video to a set standard of performance." The other behavioral objectives could be put together as one behavioral objective, "Demonstrate the ability to recall important facts." Therefore, these behavioral objectives were the operational benefit used in the *Benefits Model.* All of the identified training system parameters, training benefits, and operational benefits are summarized in Table 2.5.

Table 2.4. Post-training goals and objectives (adapted from front-end analysis).

What are the goals and specific behavioral objectives with regard to trainee performance after training?

• The main goal is to train individuals to monitor a colleague to determine if the colleague's level of drowsiness is severe enough to warrant intervention. This consists of several task components, namely monitor the colleague, determine if the colleague exhibits signs of drowsiness, determine the general level of drowsiness from five general levels, determine the level of drowsiness between the general levels, determine if intervention is necessary, and determine appropriate intervention. The instructional goals can be further specified by writing behavioral objectives.

• Specific Behavioral Objectives
- Demonstrate the knowledge that monitoring colleagues for drowsiness is important for crew safety.
- Describe how often to monitor.
- Demonstrate knowing what normal facial tone is.
- Demonstrate knowing what normal fast eye blinks are.
- Demonstrate knowing what short ordinary glances are.
- Demonstrate knowing what occasional body movements or gestures are.
- Demonstrate the skills to determine the difference between blinks or purposeful eye closures and slow eye-lid closures.
- Demonstrate knowing what the criterion level for intervention is.
- Demonstrate knowing what the appropriate action is considering the task.
- Demonstrate knowing if alerting the colleague of his or her condition will be sufficient to bring the colleague out of his or her drowsy state.
- Demonstrate knowing if the person needs a rest break period.
- Demonstrate knowing if the person should finish the shift.
- Demonstrate knowing if the work crew is in a position to be able to work without the drowsy person.
- Demonstrate the ability to rate the drowsiness level of a person via video to a set standard of performance.

Table 2.5. The identified variables for the *Benefits Model*.

Training System Parameters	**Training Benefits**	**Operational Benefits**
Financial Resources	Time to complete training session with practice	Demonstrate the ability to rate the drowsiness level of a person via video to a set standard of performance.
Human Resources	Number of remediation sessions necessary	
Equipment and Tools	Time to revise program (after formative evaluation)	Demonstrate the ability to important facts.
Time Resources	Trainee motivation	
Large Number of Trainees	Self-efficacy	
Many Locations for Training	Convenience of attending training	

CHAPTER 3

METHOD AND RESULTS OF APPLYING THE PREDICTIVE COST-EFFECTIVENESS MODEL (*THE BENEFITS MODEL*)

EXPERIMENTAL DESIGN

The *Benefits Model*, a predictive cost-effectiveness model, solicits information from experts for the classification and quantification of identified variables. Therefore, three data collection sessions were held for each of 14 experts. Session 1 was held for each expert to obtain the variable classifications; in other words, determine between which variables a relationship existed. A binomial test was used to determine when there was significant agreement among the experts on the existence of a relationship. After data from Session 1 was compiled, Session 2 was held for each expert to obtain the variable quantifications; that is, to assign either a positive or negative value to each of the causal links. A binomial test was used to determine when there was significant agreement of either a postive or negative relationship. Afterwards, the experimenter input the quantified benefits (B_i) into the *Benefits Model* to determine a value for each B_j. Finally, Session 3 was held for each expert to review the predictive data collected and determine which training program should be considered the most cost-effective. "Experts" were defined as individuals who had formal coursework in training system development at the graduate level and had experience in designing training systems.

SESSION 1

Subjects

The same experts were used in each of the three experimental sessions. The experts were 14 graduate students (4 women, 10 men) from Virginia Polytechnic Institute and State University who were enrolled in the Industrial and Systems Engineering graduate program with a Human Factors Option. All 14 of the experts had taken a graduate level course in

Training Systems Development in the Human Factors Option and were familiar with the basic Macintosh computer interface.

Session 1 took less than one hour. The experts were paid $6.00 for their time.

Apparatus

Classification statements were arranged in a HyperCard software program on a Macintosh PowerBook Duo 230 laptop computer. The program was converted to a HyperCard Player version (a play only program that does not rely on the existence of the original software to operate, and does not allow the user to alter the program) and designed such that all expert responses were compiled on a single "card" that could then be accessed for data analysis. The experts used the same laptop computer to respond to statements in the Hypercard Player program.

Procedure

Session 1 was conducted in a private office with the laptop computer set up on a desk. Each expert began the first session by reading and signing an informed consent form (Appendix A). Because expert opinion was collected via computer software, instructions for the task were given as part of the software program (Appendix B). A description of the drowsiness detection training programs was included in the instructions. Also, the experimenter paraphrased the instructions to the expert and encouraged the expert to ask questions before starting.

Appendix C shows a hardcopy version of a HyperCard Player program that was used to collect data from the experts. The program was designed such that the experts could go forward but they could not go back for two reasons. First, it was desirable for the experts to concentrate on the relationship of the variables presented on each card instead of making decisions based on past comparisons. Second, the experts were asked to concentrate on the variables as they related to each training medium, without making a decision based upon how the expert considered the relation for the other medium. The fact that the user could go forward but not back was explained in the instructions, and the experimenter also explained the reasoning to each expert. A list of variable definitions was provided on a single sheet of paper and given to the experts to refer to as needed (Appendix D). The definition list was provided in hardcopy format so that it would always be available to the expert, as opposed to requiring the expert to "jump" to a definition list in the program. For each variable, the expert was instructed to select every variable listed that would be affected by the first variable. This was accomplished by moving the cursor to select and highlight the variables, causing an "x" to appear in the corresponding check box.

Results

A two-tailed binomial test was performed to determine if there was a significant agreement between the experts for each response. The outcome was significant at the $\propto$ = 0.05 level if 12, 13, or, 14 experts agreed ($p = 0.0132$). When there was not significant agreement on the response (11 or fewer experts agreed), then the relationship between the variables was assigned a "0" to represent that a relationship did not exist. The variables that had a significant number of experts agree on a relationship were put into a HyperCard program for Session 2. These variables are shown in Table 3.1.

After the expert finished the task, the experimenter asked the expert what he or she thought of the task. When the expert left the room, the experimenter wrote down those comments.

SESSION 2

Subjects

The experts who volunteered for Session 2 were the same who volunteered for Session 1. Session 2 took less than one hour. The experts were paid $6.00 for their time.

Apparatus

The same apparatus was used for Session 2 as that used for Session 1.

Table 3.1. The variable relationships that were significantly agreed upon by the experts.

Lecture Training Program Variables Affected	Multimedia Training Program Variables Affected
Financial resources affect: human resources equipment and tools number of locations at which trainees can be taught	Financial resources affect: human resources equipment and tools time resources developer's time to revise
Human resources affect: time resources number of trainees that can be trained number of locations at which trainees can be taught	Human resources affect: time resources developer's time to revise
Equipment and Tools affect: developer's time to revise	Equipment and Tools affect: time resources number of trainees that can be trained number of locations at which trainees can be taught developer's time to revise
Time resources affect: number of trainees that can be trained time to complete the training program developer's time revise	Time resources affect: no variables affected
Number of trainees that can be trained affect: human resources equipment and tools time resources	Number of trainees that can be trained affect: equipment and tools
Number of locations at which trainees can be taught affect: financial resources human resources equipment and tools time resources number of trainees that can be trained convenience to attend training	Number of locations at which trainees can be taught affect: equipment and tools number of trainees that can be trained convenience to attend training
Time to complete the training program affect: trainee motivation self-efficacy	Time to complete the training program affect: number of remediation sessions trainee motivation
Number of remediation sessions affect: time to complete training trainee motivation	Number of remediation sessions affect: time to complete training trainee motivation self-efficacy
Time for revision affect: no variables affected	Time for revision affect: no variables affected
Trainee motivation affect: number of remediation sessions self-efficacy ability to rate drowsiness ability to recall information	Trainee motivation affect: time to complete training number of remediation sessions self-efficacy ability to rate drowsiness ability to recall information
Self-efficacy affect: number of remediation sessions trainee motivation ability to rate level of drowsiness	Self-efficacy affect: time to complete training number of remediation sessions trainee motivation
Convenience to attend training affect: no variables affected	Convenience to attend training affect: trainee motivation

Procedure

The results of Session 1 created the experimental testing materials for Session 2 (Appendix E). For example, for the lecture training program, there was a significant level of agreement that financial resources affect human resources, equipment and tools, and the number of locations at which trainees can be taught. Therefore, a HyperCard was included in Session 2 to allow the experts to quantify those relationships (for example, see page 111).

There are consistent features of each HyperCard in the program. At the top of each card was a statement to inform the expert of what relationships were significantly classified by the experts as a group. In the middle of the card is basic information about the amount of resources that would be required by each training medium. This was considered a minimal amount of information that the expert would need to make a quantification. At the bottom of each card was the area where the expert would make a quantification by using the computer mouse to check on one of the two choices.

Approximately one week had passed since Session 1, so the experimenter first reminded the expert what he or she had done in the last experimental session. The expert was shown a hardcopy version of the last program, then it was explained that the current task was to assign a value to the causal links (relationships) that the experts agreed upon in Session 1. The task instructions were included in the program, just as in the previous session (Appendix F). Once again, the program was designed such that the experts could go forward but could not go back through the training program. Also, the experts were given a list of variable definitions to refer to as needed.

With the *Benefits Model*, Kearsley and Compton (1981) suggest assigning a value to each benefit as it relates to a training approach or medium. Although Kearsley and Compton suggest assigning values in a range of -1 to +1, the examples they provide seem to contradict the idea of using a "range" of values. Therefore, the numbers were assigned as follows:

-1: Variable 1 negatively affects Variable 2.

+1: Variable 1 positively affects Variable 2.

To determine the quantification, the experts first read a brief description of the lecture and computer-based multimedia training programs to provide a frame of reference for the decision making. Each card in the HyperCard program referred to either the lecture or multimedia training program, and the expert was reminded to keep this in mind while responding to the quantification statements. The expert was also told that in situations in which a statistically significant relationship among experts was found, but the individual expert did not concur, the expert was asked to quantify the relationship with the assumption that the significant outcome of the classification session was valid.

After the expert had finished the task, the experimenter asked the expert what he or she thought of the task. When the expert left the room, the experimenter wrote down those comments.

Results

A two-tailed binomial test was performed to determine if there was significant agreement among the experts for each response. The outcome was significant at the $\propto$ = 0.05 level if 12, 13, or, 14 experts agreed (p = 0.0132). When there was not significant agreement on the response, then the causal link was assigned a "0" to represent that a relationship did not exist. When there was a significant relationship between the variables, the values were calculated from the inclusion of the quantified variables in the formula:

$$B_j = \sum_{i=1}^{k} V_i B_i$$

For example, for the lecture training program the B_j of "the number of trainees that can be trained" the variables that affect the B_j are B_{i1} "human resources" with an assigned value

(V_{i1}) of +1, B_{i2} "time resources" with V_{i2}=+1, and B_{i3} "the number of locations at which trainees can be taught" with a V_{i3}=+1. Summing the B_i values, the B_j for "the number of trainees that can be trained" is +3. Tables 3.2 and 3.3 show the complete results of the quantification for each training medium. Variables not listed in the B_i column for a B_j were either determined not to be related in Session 1, or did not receive sufficient agreement on variable quantification in Session 2.

Table 3.2. Summary of lecture training program quantification.

B_j Values	Corresponding V_is and B_is
0 Financial Resources	no related variables
+1 Human resources	+1 financial resources
+1 Equipment & Tools	+1 financial resources
0 Time resources	no related variables
+3 Number of trainees that can be trained	+1 human resources +1 time resouces +1 the number of locations at which trainees can be taught
+2 Number of locations at which trainees can be taught	+1 financial resources +1 human resources
0 The trainee's time to complete the training program	no related variables
-2 The number of remediation sessions necessary	-1 motivation -1 self-efficacy
+1 The developer's time to revise the training program	+1 time resources
0 Trainee motivation	-1 the number of remediation sessions necessary +1 self-efficacy
+1 Trainee self-efficacy	+1 motivation
+1 Trainee convenience to attend training	+1 the number of locations at which trainees can be taught
+2 Trainee's ability to rate level of drowsiness	+1 motivation +1 self-efficacy
+1 Trainee's ability to recall information	+1 motivation

Table 3.3. Summary of computer-based multimedia quantification.

B_i Values	Corresponding V_is and B_is
0 Financial resources	no related variables
+1 Human resources	+1 financial resources
+1 Equipment & Tools	+1 financial resources
+2 Time resources	+1 financial resources +1 the number of locations at which trainees can be taught
+1 Number of trainees that can be trained	+1 equipment and tools
+1 Number of locations at which trainees can be taught	+1 equipment and tools
0 The trainee's time to complete the training program	no related variables
0 The number of remediation sessions necessary	no related variables
0 The developer's time to revise the training program	no related variables
+1 Trainee motivation	-1 the number of remediation sessions necessary +1 self-efficacy +1 convenience to attend training
+1 Trainee self-efficacy	+1 motivation
+1 Trainee convenience to attend training	+1 number of locations at which trainees can be taught
+1 Trainee's ability to rate level of drowsiness	+1 motivation
+1 Trainee's ability to recall information	+1 motivation

SESSION 3

Subjects

The subjects who volunteered for Session 1 and 2 were the same as those who volunteered for Session 3. The third experimental session took less than 30 minutes. The experts were paid $3.00 for their time.

Apparatus

All information for the third session was on paper.

Procedure

Approximately five weeks had passed between Session 2 and Session 3; therefore, the experts were first reminded of the tasks that they had performed in the two previous sessions. This included reviewing the variables, looking over a hardcopy version of the two HyperCard programs, being reminded what it meant to classify and quantify variables, and looking over the results of Sessions 1 and 2. Then the experts received a hardcopy version of the training costs and benefits summarized (Table 3.4).

Table 3.4. Predicted cost-effectiveness of the lecture and computer-based multimedia drowsiness detection training programs.

Costs and Benefits	**Values Assigned**	
	Lecture	**CBM**
Predicted Cost to Design and Develop:*	$26,403.52	$50,225.20
Including:		
Financial resources **	0	0
Human resources	+1	+1
Equipment and tools	+1	+1
Time resources	0	+2
Predicted Cost to Disseminate		
Including:		
Number of trainees that can be trained	+3	+1
Number of locations at which trainees can be taught	+2	+1
350 trainees (1 training circuit)	$193,578.00	$214,004.00
700 trainees (2 training circuits)	$387,156.00	$390,439.00
1,400 trainees (4 training circuits)	$774,312.00	$743,309.00
2,100 trainees (6 training circuits)	$1,161,468.00	$1,096,179.00
3,500 trainees (10 training circuits)	$1,935,780.00	$1,801,919.00
7,000 trainees (20 training circuits)	$3,871,560.00	$3,566,269.00
Benefits:		
Trainee's time to complete the training session	0	0
Number of remediation sessions necessary	-2	0
Developer's time to revise the training program	+1	0
Trainee's motivation	0	+1
Trainee's self-efficacy	+1	+1
Trainee's convenience to attend training	+1	+1
Trainee's ability to rate level of drowsiness	+2	+1
Trainee's ability to recall information	+1	+1

* Predicted costs to design and develop were determined based upon the time originally allocated by the design and development contract for the project to be completed. All other numbers were determined based upon expert opinion as determined by data collected in experimental Sessions 1 and 2.

** Positive numbers denote that the cost or benefit is positive for the particular training program. Negative numbers denote that the cost or benefit is negative for the particular training program. A difference between the values in the lecture column and those in the computer-based multimedia column denotes the degree of predicted difference in performance between the two training programs. A zero denotes that the variable will neither positively nor negatively impact the training program.

For the predictive approach, the design and development costs were based upon the time and resources allocated by the contract to produce training programs. The cost-accounting forms shown in Table 3.5 were provided to allow the experts to see how the dollar amounts had been derived.

Table 3.5. Predicted costs for design and development of the Drowsiness Detection Training Program.

Lecture Cost Category	Amount	Computer-based Multimedia	Amount
A. Salaries and Wages		A. Salaries and Wages	
1. Designer/developer, M.S. Salary 45,000 440 hours	$9,519.18	1. Designer/developer, M.S. Salary 45,000 680 hours	$14,711.46
2. P.I./Project Manager, Ph.D. Salary 60,000 22 hours	634.62	2. P.I./Project Manager, Ph.D. Salary 60,000 34 hours	980.77
3. Secretarial Support Salary 22,000 22 hours	232.69	3. Secretarial Support Salary 22,000 34 hours	359.62
B. Fringe (30% of Salaries)	3,115.95	B. Fringe (30% of Salaries)	4,815.56
C. Supplies	150.00	C. Supplies	150.00
D. Travel Expenses 2 day trips to Washington D.C.	900.00	D. Travel Expenses 2 day trips to Washington D.C.	900.00
E. Equipment Purchases (3,174/3)	1,058.00	E. Equipment Purchases	10,069.95
Total Direct Cost (sum of A through E)	15,610.44	Total Direct Cost (sum of A through E)	31,987.36
Overhead @ 60% (sum of A – D)	8,731.46	Overhead @ 60% (sum of A – D)	13,150.45
Equipment Overhead @ 10% (Item E)	105.80	Equipment Overhead @ 10% (Item E)	1,007.00
Sum of Overhead	8,837.26	Sum of Overhead	14,157.45
Sum of Direct and Overhead	24,447.70	Sum of Direct and Overhead	46,504.81
Fee @ 8%	1,955.82	Fee @ 8%	3,720.39
Grand Totals	$26,403.52	Grand Totals	$50,225.20

Because the extent of the dissemination of the Drowsiness Detection Training program had not been determined in time to be incorporated into the study, the actual scenario and conditions in which the training programs would be disseminated was determined based on a likely dissemination scenario (Appendix G). For the lecture training program, the scenario was that a trainer travels to 7 U.S. cities and instructs 50 individuals at each site. For the multimedia training program, the scenario was that a single multimedia worksite is set up in 7 U.S. cities at training facilities to accommodate 50 individuals at each site who come in for individual training. Costs were then determined for the incremental training of more individuals. These cost-accounting sheets were also shown to each expert to explain how costs were determined.

Finally, the benefit B_j values shown in the summary sheet (Table 3.4) were explained to the experts as the values calculated from the inclusion of the quantified variables in the formula:

$$B_j = \sum_{i=1}^{k} V_i B_i$$

Each expert was then asked to review the information to determine which training program he or she would choose as the most cost effective. After the expert gave a response, the experimenter asked questions regarding the expert's selection of the most cost-effective training medium, and regarding the expert's opinion of the usability of the methods used to determine cost-effectiveness (Appendix H). These questions were asked in an informal interview style with the experimenter writing down responses while the expert spoke freely.

Results

All 14 of the experts agreed that the computer-based multimedia training program would be the most cost-effective, making the two-tailed binomial test significant at p =

0.0132. However, the experts' responses to the interview show that their responses were not based on information gleened from the classification and quantification of variables, but based either soley or heavily on the dollar values associated with cost of designing, developing, and dissemination of the training programs. All experts stated that it would not have been possible to use the information gathered through application of the Benefits Model to determine cost-effectiveness with any confidence.

In response to question 1 in Appendix H, the experts stated that the list of data shown in Table 3.4 simply did not give much information. Several points were brought up. One is that the numbers themselves have no meaning. For example, several experts asked what a B_j of +3 meant in relation to other values such as +2 or -1. Likewise, a second issue was that the scale of numbers was very small, making the differences seem inconsequential. Third, the B_js when totaled for each medium come out to be exactly the same, +11. Overall, it appeared that although the experts could see how the numbers were derived, they could not associate enough meaning to them to derive a valid answer regarding cost-effectiveness of the training media.

Questions 2 and 3 concerned the identification of the variables. The experimenter reminded each expert that the variables were chosen without expert input, and basically asked how appropriate or inappropriate they thought the variables were considering that the goal was to compare cost-effectiveness of two training media. About half of the trainees said they would not revise the variables. Suggestions from those who would alter the list included adding variables regarding technology advancements, long-term maintanence, the trainees' familiarity with technology and comfort level with computers, more variables regarding the goals of the training system, and including a variable for long-term effectiveness of the training program. Regarding variables that should not be included, several experts said they did not see the cost-effectiveness relevance of trainee motivation and trainee self-efficacy (the trainee's level of confidence in his or her ability to perform the task), or that motivation and self-efficacy were too similar for both to be included. Finally, one expert said that the variables regarding time seemed irrelevant.

Questions 4 and 5 regarded the experts opinion of the task of classifying variables. In an effort to get the experts to be as honest as possible, the experimenter reminded each expert that the experimenter's goal was to critique the model, and the expert's honest evaluation was important. At this point, the experimenter read two comments that were made by several of the experts after finishing Session 1. These comments were:

> "You could come up with an instance of how one [variable] affects the other [variable] if you try hard enough.
>
> "Some [classifications] could be read in more than one way. For example: Does the number of trainees affect financial resources? In one way yes, because more trainees require more resources. In another way no, because there is only so much money to spend on the training program, and the number of trainees will not make that amount go up or down."

Many of the experts said they agreed with those statements, and added more coments:

- The task was too abstract, with too many open-ended issues left to be decided by the expert.
- Sometimes the expert was left to decide where to draw the line [when trying to determine if the variables were related], and had to create a personal threshold.
- Trying to relate the variables to one another was too difficult.
- The relationships were easy enough to determine if a causal relationship existed.
- The wording of the statements greatly affected classification. For example, resources 'available' vs. resources 'required' for the training program.
- I tended to think in generalities, not just the particular training program, so I'm not sure of the relevance of the task regarding the particular training programs.
- The whole process seemed fuzzy.
- It needed more structure, more rules.
- The definition of "related" [as in whether the variables are related] is hard to understand.

An often made comment was the experts' expression of their dislike of the "check all that apply" type of task. Some of the comments made were:

- I felt constrained by only yes or no options – I would have liked to rate the strength of the relationship.
- Some classifications apply in both directions, and I felt constrained by having to answer in one direction or the other.

About half of the experts made some type of referral to wanting a rating scale for the task. This would allow the experts to specify in which direction the relationship existed and how strong that relationship was; in effect, the experts wanted to classify and quantify the variables at the same time. In fact, one expert said specifically that the task would have been easier if Sessions 1 and 2 were combined.

Questions 6 and 7 regarded the experts' opinion of the quantification of variables. Once again, the experimenter read two comments that were made by several of the experts after finishing Session 2. These comments were:

> "I started out doing a correlation between the variables to decide if the relationship was positive or negative, but that rationale stopped working when I got to the more qualitative variables."
>
> "Some variables did not have a cause and effect relationship, so doing a forced choice didn't seem right."

All the experts said they agreed with the first statement, with some of the experts adding that they did not think they were answering consistently. This is interesting input from the experts when considering that all of the experts answered using the same rationale and were consistent in their method. Although the experts started out using a correlation method which worked well as long as the idea of a correlation did not conflict with the fact that "positive" meant something good for the training program, and "negative" meant something bad for the training program. The method broke down when the qualitative variables were quantified. For example, consider the statement, "The trainee's motivation will (positively) (negatively) affect the time to complete the training program." The idea is that as motivation goes up, the time to complete training will go down. This would be a negative correlation, but good for the training program. So the experts, instead of using the correlation method, stopped to consider the training program. Therefore, most experts chose to quantify the relationship as positive.

However, when quantifying the variables, few of the trainees seemed to remember to quantify the variables in terms of the training medium. In fact, when giving their opinion of the quantification process, two of the trainees said that it was difficult to respond to the statement while keeping in mind which training medium was being considered. Furthermore, the experts seemed to ignore the basic information on each card regarding the amount of resources each medium would require (the statements in the middle of each card). It appeared that the task of quantifying the variables was simply too difficult.

Most experts found the quantification process more confusing and ambiguous than the classification process. However, there were three experts who thought the quantification process was easier to understand. As one expert put it, the quantification was clearer because it became apparant how the other experts were thinking about the variables during the classification process. This seemed to give the expert some comfort to see that his thinking aligned with other experts.

Another interesting comment made by one expert was that it would be easier to make the classifications and quantifications if the experts could sit through both training programs. The thought was that seeing the training programs would alleviate any ambiguities in the classification and quantication process. The experimenter explained to the expert that the idea of the cost-effectiveness model was to determine cost-effectiveness before the training programs were developed to save time and resources.

Question 8 asked the experts of their overall impression of the method. Many experts restated the fact that the resulting data were not very informative. Two of the experts simply said they did not like the method. The remaining experts said that although the method was new, putting value to qualitative benefits is a good idea, and time and research might help refine the method to make the data more worthwhile. In general, the experts reiterated that it would not be possible to make a decision of cost-effectiveness with any confidence without the dollar amount for design, development, and dissemination.

CHAPTER 4

DESIGN, DEVELOPMENT, AND TESTING OF THE DROWSINESS DETECTION TRAINING PROGRAMS[1]

The two training programs used in this research were a group-paced lecture training program and a self-paced computer-based multimedia training program. The training programs were titled "Drowsiness Detection Training" and were intended for flight crew personnel and air traffic controllers. The training programs consisted of an information session divided into five sub-categories, a practice session for rating the drowsiness level of individuals in video segments, and as necessary, remediation which consisted of more practice at rating video segments. The video segments of drowsy automobile simulator drivers were included in the training program. These segments were videotaped during research previously conducted by Wierwille and Ellsworth (1994).

As previously discussed, the training programs were designed following the design model for training and performance support systems advocated by Gordon (1994). After the training programs were prototyped, a two iteration optimization process took place for each training program.

The first optimization process conducted for each training program consisted of sending the storyboards for both training programs to the government agencies that requested the training programs be developed. The feedback received consisted of suggestions for improved narrative. Phrases that could be interpreted as harsh by the potential user group were reworded or eliminated. For example, the phrase "relieved of duty" was in the training programs to suggest that a drowsy person should take an extended break or leave his or her shift early. However, for flight crews, many of whom are ex-military personnel, that phrase is used to denote a particularly harsh punishment.

---

[1] Chapter 2 is adapted from Neale, V. L. and Wierwille, W. W. (1995b, December).

The second optimization process conducted for each training program consisted of bringing in an individual to act as a participant for the training programs. In these cases, the individual was a graduate student in Industrial and Systems Engineering specializing in Human Factors Engineering from Virginia Polytechnic Institute and State University who did not have prior training in drowsiness detection training. The person was encouraged to constructively criticize (not just find a problem, but make suggestions for improvement) the training program. After the second optimization process, the training programs were updated and considered ready for formative evaluation.

METHOD

Experimental Design

The study evaluated the between-training group differences for the following variables:

1. Ability to rate drowsiness level. This is the ability of the trainee to accurately and consistently rate the level of drowsiness of individuals in 24 videotaped segments. This includes absolute error from previously obtained expert ratings (Wierwille and Ellsworth, 1994), intrarater reliability, and interrater reliability.
2. Ability to recall presented information. This is the ability of the trainee to accurately recall critical information about the training program other than how to rate the level of drowsiness, such as, ability to know when to intervene on behalf of a drowsy person.
3. Time to complete the training session including practice. This time was obtained for the lecture group as a whole, and for the multimedia group individually. This time does not include the time for remediation.
4. Number of remediation sessions necessary. Remediation was given if the results of the practice sessions were not satisfactory. This number was obtained by counting each remediation session that the trainees needed. This number was

obtained for the lecture group as a whole, and for the multimedia group individually.

5. Trainee motivation. Trainee motivation was determined for each trainee by his or her response to a question on the post-test questionnaire. Trainee motivation is defined as the level to which a trainee wants to learn the presented material.
6. Self-efficacy. Self-efficacy is defined as the ability to believe in one's capability to perform a task and was determined for each trainee by his or her response to a question on a post-test questionnaire.
7. Convenience to attend training. Convenience for the trainee to attend a training session was determined by a response on the post-test questionnaire. This was assessed by presenting two training scenarios to the trainees which explained the two methods of receiving training and asking which method the trainee would find most convenient to attend.

Subjects

Because it was not possible to test the actual user group for whom the training programs were intended (flight crew members and air traffic controllers), trainees were chosen as naive and matching the educational level of the user group; that is, individuals with at least two years of college (junior and senior level undergraduate students) who had not participated in previous drowsy driver studies or drowsiness detection studies. The trainees were matched for computer experience and any piloting experience they may have had.

It was intended that sixteen volunteer participants, 8 females and 8 males, be assigned to each training medium group (lecture and computer-based multimedia) for a total of 32 participants. However, there was a concern that if all of the participants were not present for the lecture training program, the session would need to be canceled and rescheduled. Therefore, three extra participants (2 extra males, and one extra female) were signed up to participate in the lecture training session, for a total of 19 participants. Since all 19 lecture participants were present at the scheduled time, the data of two males and one female were discarded from analysis. The subject data to be discarded were selected based on the fact that two males had experience piloting an airplane (no one in the multimedia group had any experience piloting an aircraft), and one female admitted to being a graduate student when a request was made for only undergraduate junior and senior level students.

To account for the possible confound of running the trainees at different times during the university semester, all trainees were recruited and asked to participate during the same two week time period in the semester. The entire experimental session, whether lecture or multimedia, took approximately two and one-half hours, for which the participants were paid $6.00 per hour.

Apparatus

The two training programs required a different equipment list. The lecture program required the lecturer's overheads and notes, the trainees' handouts, an overhead projector for the lecturer, a video-cassette player, and a television monitor. The computer system for the computer-based multimedia program was a Power Macintosh 8100/100AV computer, a keyboard, mouse, and 20" high resolution computer monitor. The multimedia program was programmed with Director 4.0, a multimedia authoring package by MacroMedia.

The video segments used in the training programs and for the post-test rating session came from previous experiments performed in the Vehicle Analysis and Simulation Laboratory at Virginia Polytechnic Institute and State University which involved drowsy drivers. In these experiments, low-light level video recordings of the drivers' faces had been made. The videotaped segments showed drivers driving a computer-controlled, moving-base simulator, and contained episodes of various levels of drowsiness.

A digital stopwatch was used by the lecturer/experimenter to record the time to complete training. The number of remediation sessions was also recorded by the lecturer/experimenter for both training media.

Procedure

All trainees first read and signed an informed consent form to participate in the study (Appendix I). Thereafter, the trainees read an introduction to the study and any questions the trainees may have had were answered (Appendix J). Following the answering of any questions, a pre-test questionnaire was filled out by each trainee (Appendix K). The pre-test questionnaire was used as the basis to match the participants' computer experience and any piloting experience they may have had.

The lecture training program was conducted in a seminar format[2], and the computer-based multimedia program was conducted on an individual basis. As indicated previously, the training programs consisted of an information session divided into five sub-categories, a practice session for rating the drowsiness level of individuals in video segments, and if necessary, a remediation session consisting of more practice at rating video segments.

It is worth noting that both training programs were designed such that the trainees were never aware that they were receiving remediation. During the experiment, whether or not the trainees received remedial training was dependent upon their response to the five practice sessions. For the lecture training program, the lecturer asked the trainees to tell what their rating was by a show of hands. A consensus of less than 75% on each practice session required that one remediation session be given, for a possible total of three remediation sessions. For the computer-based multimedia training program, the program was designed with an algorithm that gave the trainee remedial practice if the responses to previous ratings were not in an acceptable range. After completion of the training program

---

[2] Lecture included the use of visual media such as video as explained in the Apparatus section.

and the practice sessions, participants were tested for their ability to rate the level of drowsiness of individuals in 24 video segments.

The recording of the time to complete training began after the informed consent, introduction to the study, and the pre-test questionnaire were filled out, and when the actual training program began. The timer was stopped when the subject(s) finished the last of five practice segments and before the first remediation session began. The number of remediation sessions was recorded by the lecturer/experimenter for both training media.

The procedure to test for the trainees' ability to rate the level of drowsiness of drivers in this study was similar to the testing procedure used in the Wierwille and Ellsworth (1994) study to determine the validity of expert raters. The rating task consisted of viewing 24 video segments of drivers at various levels of drowsiness and subjectively rating each segment on its corresponding rating-scale form.

The rating scale used was that developed for the Wierwille and Ellsworth (1994) study, which is a form of the Likert (descriptive graphics) scale. The continuous scale contained five descriptors: Not Drowsy, Slightly Drowsy, Moderately Drowsy, Very Drowsy, and Extremely Drowsy. One scale was used for each segment to be evaluated, totaling 24 scales. The scale is shown in Figure 4.1. As in the Wierwille and Ellsworth study, each copy of the scale was provided on a separate slip of paper, approximately 22 cm wide and 7 cm high, to minimize influences from previous ratings.

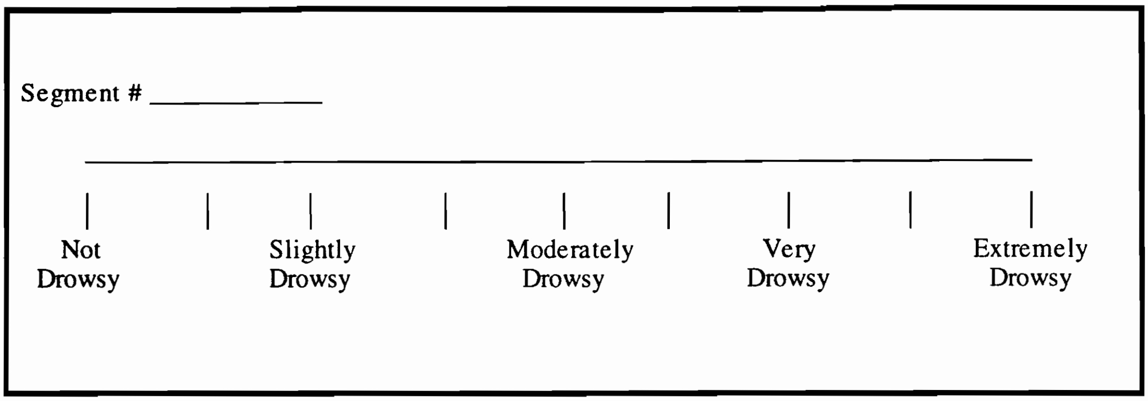


Figure 4.1. Rating scale used by the trainees to rate drowsiness level.

The rating task was as follows: When the testing session began, the first videotaped image appeared on the screen. A short time thereafter, a recorded voice instructed the rater to begin the evaluation for that segment. The length of the presentation period was one minute. Afterwards, the rater had 25 seconds to provide a rating. The rater was able to change only the current segment rating. The rater was not permitted to go back to a previous segment to change a rating.

After the trainee had completed the ratings, a post-training questionnaire (Appendix L) was administered. This questionnaire assessed the trainee's ability to recall information learned in the training program, trainee motivation, personal self-efficacy, and the trainee's opinion of convenience to attend two hypothetical training scenarios.

RESULTS

Ability to Rate Drowsiness Level

The first step in analyzing the trainee rating data was to convert the ratings on the ratings scales to numerical values, as in the Wierwille and Ellsworth (1994) study. This task was accomplished by converting the rating scale to a 100-point scale and then measuring the location of each given rating. Zero was at the left end and 100 was at the right end.

Absolute error. Raw absolute error was obtained by subtracting the trainees' ratings from the mean experts' ratings obtained in the Wierwille and Ellsworth (1994) study. Differences between the lecture and computer-based multimedia groups were evaluated with a Training Type (Lecture and Multimedia) X Segment (24 video segments) mixed-factor analysis of variance with Training Type as the between subjects variable and Segment a within subject variable. The analysis was performed using SuperANOVA software. The sums of squares table generated by SuperANOVA is presented in Table 4.1.

Table 4.1. Results of absolute error ratings.

Type III Sums of Squares

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Training Type	1	867.000	867.000	6.212	0.0184		
Subject (Group)	30	4187.186	139.573				
Segments	23	15516.930	674.649	8.958	0.0001	0.0001	0.0001
Segment*Training	23	3074.216	133.662	1.775	0.0144	0.0700	0.0384
Segment*Subject	690	51967.630	75.315				

Dependent: Segments

As can be seen, the main effect of Training Type was significant ($F_{(1, 30)} = 6.212$, $p = 0.0184$), with the Lecture group performing with less absolute error. However, as shown in Table 4.2, the mean difference in absolute error between the two groups is only 2.125. This is a rather small difference when considering that the rating scale was a 100-point scale.

Table 4.2. Means table for the main effect of training type.

	Count	Mean	Std. Dev.	Std. Error
Multimedia	384	11.995	10.135	0.517
Lecture	384	9.870	9.615	0.491

The main effect of Segment was significant ($\underline{F}_{(23, 690)}$ = 8.958, $\underline{p}$ = 0.0001). Furthermore, there was a significant interaction of Segment x Training ($\underline{F}_{(23, 690)}$ = 1.775, $\underline{p}$ = 0.0144). On the repeated measures variable (Segments) Greenhouse Geiser and Hunyh-Feldt corrections were conducted. Both correction statistics showed agreement that the main effect of Segment was significant. The Greenhouse Geiser, a conservative estimation, did not show agreement that the Segment x Training interaction was significant ($\propto$ = 0.05). The Hunyh-Feldt, a liberal estimation, did show agreement that the Segment x Training interaction was significant.

For the main effect of Segment and the interaction of Segment x Training, Student Newman Kuels (SNKs) tests were conducted on the combined segment absolute error means, and the multimedia and lecture segment absolute error means separately. Table 4.3 shows results of the SNKs.

As can be seen, some of the videos were more difficult to rate accurately than others. The ratings for Segments 1, 4, 9, 14, and 23 had high absolute error means, while Segments 8, 15, and 17 had low absolute error means. The mean expert ratings for these video segments are shown in Table 4.4. The means correspond to the rating placement on the drowsiness continuum as a 100-point scale. By simply looking at the mean ratings, it would appear that the video segments that are more difficult to rate are those in the Slightly Drowsy to Moderately Drowsy range, while those segments at the extremes of the continuum, either Not Drowsy or Extremely Drowsy, are less difficult to rate. These generalizations probably occur because individuals in the Slightly to Moderately Drowsy range often exhibit few, if any, signs of drowsiness. The exception to this explanation is Segment 14, which has a low mean expert rating (meaning the individual in the video segment was close to Not Drowsy). However, when reviewing Segment 14, it was noticed that the individual in Segment 14 was an individual who the trainees had seen in the

Table 4.3. Trainee rating segment absolute error means for the multimedia and lecture trainee ratings combined, and for each trainee group individually.

Seg #	Combined Segment Absolute Error Means	Seg #	Multimedia Segment Absolute Error Means	Seg #	Lecture Segment Absolute Error Means
1	21.76 a	1	27.23	14	20.81 a
14	19.31 a b	3	19.79 a	6	16.63 a b
9	16.73 a b c	4	18.34 a b	1	16.29 a b c
23	15.28 a b c d	9	17.89 a b c	9	15.56 a b c d
4	15.18 a b c d e	14	17.81 a b c d	23	14.46 a b c d e
3	14.65 a b c d e f	23	16.11 a b c d e	20	12.13 a b c d e f
6	12.88 b c d e f g	13	15.23 a b c d e f	4	12.02 b c d e f g
22	12.66 b c d e f g	22	14.19 a b c d e f g	22	11.13 b c d e f g h
13	12.64 b c d e f g	2	12.48 a b c d e f g h	18	10.81 b c d e f g h
20	12.06 b c d e f g	7	12.17 a b c d e f g h	13	10.04 b c d e f g h
7	10.98 b c d e f g	20	12.00 a b c d e f g h	7	9.79 b c d e f g h
2	10.76 c d e f g	10	11.33 a b c d e f g h	3	9.50 b c d e f g h
18	9.84 c d e f g	16	10.06 b c d e f g h	5	9.42 b c d e f g h
11	8.97 c d e f g	11	10.00 b c d e f g h	2	9.04 b c d e f g h
10	8.64 c d e f g	12	9.31 b c d e f g h	11	7.94 b c d e f g h
12	8.63 c d e f g	24	9.27 b c d e f g h	12	7.94 b c d e f g h
5	8.38 c d e f g	6	9.13 b c d e f g h	21	7.81 b c d e f g h
21	8.00 c d e f g	18	8.88 b c d e f g h	19	6.54 d e f g h
16	7.80 c d e f g	21	8.19 d e f g h	8	6.00 d e f g h
24	7.40 c d e f g	19	7.46 e f g h	10	5.96 d e f g h
19	7.00 c d e f g	5	7.33 e f g h	16	5.54 d e f g h
8	4.97 f g	15	5.69 f g h	24	5.52 e f g h
17	4.25 f g	17	4.06 g h	17	4.44 f g h
15	3.63 g	8	3.94 h	15	1.56 h

Cell means without a common letter are significantly different ($p < 0.05$).

Table 4.4. Mean expert ratings for the video segments selected for qualitative evaluation.

Video Segment	Expert Mean Rating
1	12.83
4	46.67
9	14.17
14	8.67
23	27.83
8	6.00
15	100.00
17	95.17

training portion of the experiment. During training, a different video segment of the same individual was used to demonstrate the Slightly Drowsy level, which looked very similar to Segment 14. Indeed, the trainees in both training groups rated Segment 14 in the slightly drowsy range on the continuum.

Intrarater reliability. On the tape there were three segments that were repeated. Segments 3, 8, and 10 were repeated as segments 23, 18, and 19, respectively. The repetition of the segments gave three pairs of scores for each rater which were used to determine intrarater reliability. A correlation to compare the first and second viewing of the video segments for each trainee in each group revealed a correlation value of $r = 0.946$, $p < 0.05$ for the multimedia training group, and a correlation value of $r = 0.851$, $p < 0.05$ for the lecture training group. To determine if the correlations were significantly different, a *Fisher's z* transformation was conducted. The intrarater reliabilities for the two groups were found not to differ significantly ($z = 0.5653$, $p = 0.21$).

Interrater reliability. To determine interrater reliability, all repeated segments were deleted from the data. Only those segments that were not repeated were used in the statistical analysis. After deleting the repeated segments there remained 18 segments. A correlation to evaluate the consistency of each rater across segments revealed an average correlation value of $r = 0.841$ for the multimedia group and a correlation value of $r = 0.852$ for the lecture group. Since the correlation values were similar, no further analysis was conducted. The interrater reliabilities of the two groups were considered to be the same.

Ability to Recall Presented Information

The ability of the trainee to accurately recall critical information about the training program (other than how to rate the level of drowsiness, such as, ability to know when to intervene for a drowsy person) was examined. This was assessed by evaluating trainees' responses to questions 1 - 12 on the post-test questionnaire, shown in Appendix D. Twenty-five points were possible on the test. The multimedia group had a mean score of

17.6 and the lecture group had a mean score of 18.0. An unpaired t-test conducted using StatView revealed no significant difference.

Time to Complete the Training Session Including Practice

The time for the multimedia group to complete training, including practice and excluding remediation, ranged from approximately 65 minutes to 69 minutes, with one outlier of 72 minutes. The time for the lecture training group was 72 minutes. It should be noted that the lecture training group did stop the lecturer to ask some minor questions, such as who the people were in the video segments, and why the people in the video segments were drowsy. The lecturer stopped the training to explain the process of videotaping the individuals who were participants in a previous drowsy driver experiment. This may have added three or four minutes to the lecture time. Therefore, considering the few questions asked during the lecture training program, and the fact that the recorded times are so close, the time to complete training is considered to be the same for the two training methods.

Number of Remediation Sessions Necessary

Remediation was given if the results of the practice sessions were not satisfactory. This number was obtained by counting each remediation session that the trainees needed. This number was obtained for the lecture group as a whole, and for the multimedia group individually. For the multimedia group, all but two of the trainees had all three remediation sessions. One of the trainees had only one remediation session, and the other had only two remediation sessions. For the lecture training program, all three remediation sessions were given because the minimal 75% consensus was not obtained on three of the practice segments. Therefore, overall, three remediation sessions were necessary for both groups.

Trainee Motivation

Trainee motivation was determined for each trainee by his or her response to question 13 on the post-test questionnaire (Appendix D). Converting the scale into a 100-point scale (zero on the left and 100 on the right), both the multimedia group and the lecture group produced mean motivation ratings of the program of approximately 55. Therefore, the trainee motivation for the two groups was essentially the same.

Self-efficacy

Self-efficacy is defined as the belief in one's ability to perform a task and was determined for each trainee by his or her response to question 14 on the post-test questionnaire (Appendix D). Converting the scale into a 100-point scale, the multimedia group rated their self-efficacy to be approximately 55, and the lecture training group rated their self-efficacy to be approximately 59. An unpaired t-test did not reveal a significant difference between the means ($p = 0.36$).

Convenience to Attend Training

Convenience for the trainee to attend a training session was determined by each trainee's response to question 15 on the post-test questionnaire (Appendix D). All of the multimedia trainees thought that the multimedia training program would be more convenient to attend. Eleven of the 16 lecture trainees thought the multimedia training would be more convenient to attend, while 5 of the lecture trainees preferred to attend a lecture training program.

Trainee Responses to Question 16 on the Post-test Questionnaire (Trainee "Likes")

The following list shows the categories of responses given by the multimedia training group for the question, "What aspects of the training program did you like?"

- interactive program
- the different media brought together made the program more interesting
- computer was less intimidating than having an instructor check your practice
- the computer application is very user friendly, simple to use and understand
- the one on one aspect of the program
- the video examples were entertaining and helpful
- being able to compare the practice segments to the experts' rating
- even though the training was computerized, it had a very human feel
- the narrator was excellent and captivating, made the experience more "real"
- the narrator keeps the trainee's attention
- outline of important points was helpful and easy to understand
- very detailed, well developed, easy to learn
- good comparison of levels

- repetition of different mannerisms
- program was concise and to the point
- the sounds (*referring to the occasional audio feedback*)
- the comparison of different levels (*referring to the side by side video examples*)
- repeated information on video to give reasoning behind the experts' opinion
- learning how to separate drowsiness

The following list shows the categories of responses given by the lecture training group for the question, "What aspects of the training program did you like?"

- use of explanation, both written and overhead
- use of practice videos
- topic was interesting
- videos were entertaining
- instructor was relaxed and made class feel relaxed
- introduction to material I have not seen before
- learning new information
- everything was well presented and ran smoothly
- the practice segments
- side by side video examples
- very complete, not too complicated
- the many examples
- presentation was concise

To summarize what the trainees' liked about the training program, trainees in both groups found the videos to be entertaining. Also, trainees in both groups liked the narrator for the multimedia training program and the lecturer for the lecture training program, who was the same person. Furthermore, both groups thought the information presented was well organized and precise. Trainees in the multimedia group liked the interaction, and some multimedia trainees responded positively to using a computer.

<u>Trainee Responses to Question 17 on the Post-test Questionnaire (Trainee "Dislikes")</u>

The following list shows the categories of responses given by the multimedia training group for the question, "What aspects of the training program did you dislike? What aspects of the training program would you change if you could?"

- need more subjects (*different people in the video examples*)
- make the narrator in color
- more color
- an additional person for voice overs to give more balance

- seemed tedious and long
- background information in beginning seemed overly simplified
- some videos were too long...shorten the scenes as the program progresses
- need more practice at the end of each level
- show the differences of drowsiness for one person as time goes on
- hard to distinguish between middle levels on rating scale
- information was dry, makes me drowsy
- some of the video images were not clear

The following list shows the categories of responses given by the lecture training group for the question, "What aspects of the training program did you dislike? What aspects of the training program would you change if you could?"

- kind of dull
- add color
- seemed tedious and long
- more interactivity, like with a computer
- video is slow and fuzzy
- video segments too long
- reduce time of practice videos
- program makes viewer drowsy
- need more practice
- should see at least one person progress through all the stages
- the slightly and moderately drowsy levels could be combined

To summarize the trainees' dislikes, both groups found the drowsy driver videos to be too long in some instances and of poor quality. Several trainees suggested color video segments, including making the narrator video in color. Overall, suggestions were given to make the programs do more to keep the trainee's attention. For both the multimedia and the lecture groups, a suggestion was made to show one person at all levels of drowsiness (one person was shown at both the Not Drowsy and Extremely Drowsy stages). Also, it is interesting to note that several of the trainees in the lecture training group requested more interaction.

DISCUSSION

Evaluation of the self-paced computer-based multimedia training program and the group-paced lecture training program for the topic of drowsiness detection training revealed that a significant difference did exist between the two groups for absolute error, although the error difference was only approximately 2 points out of 100. A more relevant finding was the significant main effect of Segment, and the significant Training Type x Segment interaction revealing that, overall, the trainees in both groups had more difficulty rating the Slightly and Moderately drowsy levels than rating the other levels of drowsiness. This is not too surprising because individuals at the Slightly and Moderately drowsy levels may exhibit few if any signs of drowsiness. The issue was addressed in the training programs, and it was explained to the trainees that in an actual work setting the trainees should draw on past knowledge and experiences of their colleague to make an accurate assessment of the person's drowsiness level. This was considered an important aspect of the training program to detect and correct before the final version of the training program is released. Also, regarding the ability to rate the level of drowsiness, the correlations for intrarater and interrater reliability for both groups was high, and not significantly different.

The ability of the trainees to recall presented information was not significantly different; however, the test scores were not satisfactorily high. The multimedia group scored, on average, about 70%, and the lecture group scored, on average, about 72%. In particular, the question that trainees had difficulty with was question 3, "What are the five task components to help a colleague who is drowsy?" The correct answer was stated in the training program and, as each task component was discussed, the name of the task component was the title for the computer screen or slide. In retrospect, and by evaluating the trainees responses to that question, it became clear that the question itself was misleading to some trainees. Instead of focusing on the words "task components," they focused on the words "help a colleague" and gave examples of how to intervene to help a colleague.

The dependent variables of 1) time to complete the training session, 2) number of remediation sessions, 3) trainee motivation, 4) self-efficacy, and 5) convenience to attend training were either not significantly different or not different enough to analyze statistically. It should be mentioned that it would be preferable to have the trainee motivation ratings a bit higher; however, the level of motivation the trainees' felt could be impacted by the fact they did not recognize the ramifications of having crew members work while drowsy. The intended user group, flight crew personnel and air traffic controllers, may find the topic more motivating and personally relevant. Similarly, it would be preferable to have the self-efficacy rating higher; however, it is reasonable to believe that practice in an actual work setting would give workers more confidence in their ability to rate drowsiness.

Finally, the results of the dependent variables of the trainees' likes and dislikes are presented. The categories of comments are listed and summarized in the results section, several of which should be taken into account. It should be noted that the trainees liked the use of the videos and the practice sessions, and appreciated a relaxed presentation atmosphere. Regarding the trainees' dislikes, the comments regarding the quality and length of the videos are important. It is a possibility that these factors impacted the trainees' motivation level toward the programs. The trainees' likes and dislikes were seen as a valuable source for improvements to the final version of the training programs.

SUGGESTED CHANGES FOR THE DROWSINESS DETECTION TRAINING PROGRAMS BASED ON EMPIRICAL EVALUATION

Based on the results of the evaluation of the computer-based multimedia and self-paced lecture training programs, some suggestions for improvement were made. Overall, the two training programs were not significantly different from one another, except perhaps for the significant difference in ability to rate the level of drowsiness. However, as previously discussed, the difference is minimal when considering that the mean difference in absolute error between the groups was only about 2 points out of 100. Also, although there was a significant Training Type x Segment interaction, a more thorough evaluation shows that, overall, the trainees in both groups had problems with rating segments in the Slightly to Moderately drowsy range. This finding suggests that the training program should give more focus to these levels of drowsiness, especially since this is the range in which a worker would need to start monitoring the level of drowsiness more closely for changes. Furthermore, these are important levels to distinguish since, if the worker cannot distinguish these levels, the drowsy person may appear to quickly move into the Very Drowsy or Extremely Drowsy range without warning, leaving the crew with less options for intervention.

The results of the trainees' ability to recall presented information was not as high as expected, and there was a particular problem with some of the trainees recalling the five task components that need to be performed. However, the question used to obtain this information was ambiguous to some trainees. Nonetheless, this information was clarified by listing and numbering these task components, and more distinctly stating that another task component is being discussed. For example, instead of stating, "We are to our next task component...," the statement was changed to, "Next we come to task component three. Task component three involves determining the specific level of drowsiness." This was accompanied by appropriately redundant visual cues.

As is often the case, the best suggestions for change come from the trainees themselves. The trainees liked the use of the videos and the practice sessions, and appreciated a relaxed presentation atmosphere. These aspects of the training programs should be maintained in future revisions.

Several of the trainees stated that they got tired of seeing the same people over and over again in the video segments. However, part of the reason that the trainees saw the same people often was that the trainees had a post-training rating session which consisted of rating some of the same people. Therefore, the actual trainees would not see the same person as often since they would not go through a post-test rating period.

Several of the trainees suggested having the video segments of the drowsy individuals and of the narrator (in the case of the multimedia program) in color. This is not feasible for two reasons. First, the video of the drowsy individuals was shot with a low-light level camera which does not allow for color video. Second, the multimedia training program is not capable of playing back full screen, full frame rate video in color. In fact, the narrator video was shot in color, yet the computer for the multimedia training program played the video back in black and white to maximize frame rate. As equipment advancements are made, it should become possible to have the narrator video appear in color.

Some comments were also made about the quality of the video segments of the drowsy individuals. As mentioned, the drowsy driver video segments were obtained with a low-light level video camera. Although it is true that low-light level photography is not as polished in appearance as high-resolution color images, the video images taken were of extremely high quality considering that low-light level video imaging technology was used.

A suggestion was also made to make some of the videos shorter in length. While talking to a trainee after he had finished the multimedia training, it was noted that the segments of the very and extremely drowsy people were easy to rate in the first few seconds. Also, it is true that some of the video segments show people who exhibit almost no change throughout the segment. Such segments can be made shorter without losing critical information.

The trainees also stated that they would like to see one person through all stages of drowsiness. The original training program did have an example of the same person at the Not Drowsy stage and the Extremely Drowsy stage in a side by side video example. Nonetheless, an example of the same person at each level of drowsiness is a good method for reinforcing the concepts of the various levels of drowsiness, and of reinforcing the consideration of individual characteristics of the person.

The trainees also requested more practice videos, especially in the Slightly and Moderately drowsy range. This ties into the issue that almost all trainees used all three remediation segments. A suggested change to the training program is to eliminate the concept of remediation in both training programs. Instead, the trainees should be given 10 practice segments with the majority of them in the middle ranges of drowsiness where the trainees need the most help. Because the trainees enjoyed receiving specific feedback as to how the experts rated and why, this feature should be maintained in the programs.

There are a couple of changes that were made to the training programs that are not based on the empirical evaluation, but were made as a request by individuals from the supporting organization. It is noted that although some critical changes were made to the programs during a two-phased optimization process of each of the training programs, the changes listed below were not made before user testing since they were time intensive changes and not viewed as potentially interfering with the testing of the two training media. These suggested changes are as follows:

1. The use of the word "safe" and forms thereof will be removed from the training program.
2. Terminology of "quickly drop off" will be changed to "may move toward a greater level of drowsiness.
3. The word "colleague" will be made singular in all cases.
4. The words "he/she" will be eliminated as much as possible.

The revised outline of the training programs can be seen in Appendix M, and some example screen captures can be seen in Appendix N. The revised training programs are not likely to have any significant performance differences between them. This fact allows for the selection of the most cost-effective training medium for the *Drowsiness Detection Training Program* to be based solely upon issues of cost of dissemination.

CHAPTER 5

METHOD FOR THE FORMATIVE COST-EFFECTIVENESS APPROACH

EXPERIMENTAL DESIGN

For the formative cost-effectiveness model, data were collected for the identified variables during a traditional formative evaluation of the training programs (Chapter 4). Data were collected for the six identified training benefits and the two identified operational benefits of the Drowsiness Detection Training program. These results are summarized in Table 5.1

Table 5.1. Costs and benefits summarized for the formative evaluation approach.

Formative Costs and Benefits	**Values Assigned**	
	Lecture	**CBM**
Cost to Design and Develop:	$31,536.73	$68,475.42
Benefits:		
Trainee's time to complete the training session	no difference	
Number of remediation sessions necessary	no difference	
Developer's time to revise the training program*	70 hours	210 hours
Trainee's motivation	no difference	
Trainee's self-efficacy	no significant difference	
Trainee's convenience to attend training	most preferred to attend multimedia	
Trainee's ability to rate level of drowsiness	no practical difference	
Trainee's ability to recall information	no significant difference	
Predicted Cost to Disseminate		
350 trainees (1 training circuit)	$193,578.00	$214,004.00
700 trainees (2 training circuits)	$387,156.00	$390,439.00
1,400 trainees (4 training circuits)	$774,312.00	$743,309.00
2,100 trainees (6 training circuits)	$1,161,468.00	$1,096,179.00
3,500 trainees (10 training circuits)	$1,935,780.00	$1,801,919.00
7,000 trainees (20 training circuits)	$3,871,560.00	$3,566,269.00

* Included in the cost to design and develop.

To summarize the data, each variable is listed in terms of its significant or insignificant performance outcome. The design and development costs were based upon the designer's time recorded during the design and development process in hours

(Appendix O) and that information incorporated into a revised cost accounting form like that for the predicted costs. Table 5.2 shows the formative cost assessment.

Table 5.2. Formative costs for the design and development of the Drowsiness Detection Training Program.

Lecture Cost Category	Amount	Computer-based Multimedia	Amount
A. Salaries and Wages		A. Salaries and Wages	
1. Designer/developer, M.S. Salary 45,000 521 hours	$11,271.58	1. Designer/developer, M.S. Salary 45,000 887.5 hours	$19,201.05
2. P.I./Project Manager, Ph.D. Salary 60,000 39 hours	1,125.00	2. P.I./Project Manager, Ph.D. Salary 60,000 133 hours	3,836.55
3. Secretarial Support Salary 22,000 26 hours	275.00	3. Secretarial Support Salary 22,000 89 hours	941.35
B. Fringe (30% of Salaries)	3,801.47	B. Fringe (30% of Salaries)	7,193.69
C. Supplies	150.00	C. Supplies	150.00
D. Travel Expenses 2 day trips to Washington D.C.	900.00	D. Travel Expenses 2 day trips to Washington D.C.	900.00
E. Equipment Purchases (3,174/3)	1,058.00	E. Equipment Purchases	10,769.95
Total Direct Cost (sum of A through E)	18,581.05	Total Direct Cost (sum of A through E)	42,992.59
Overhead @ 60% (sum of A – D)	10,513.83	Overhead @ 60% (sum of A – D)	19,333.58
Equipment Overhead @ 10% (Item E)	105.80	Equipment Overhead @ 10% (Item E)	1,077.00
Sum of Overhead	10,619.63	Sum of Overhead	20,410.58
Sum of Direct and Overhead	29,200.68	Sum of Direct and Overhead	63,403.17
Fee @ 8%	2,336.05	Fee @ 8%	5,072.25
Grand Totals	$31,536.73	Grand Totals	$68,475.42

Because the extent of the dissemination of the Drowsiness Detection Training program had not been determined at the time of this writing, the actual scenario and conditions in which the training programs would be disseminated was determined based upon the same dissemination scenario used for the predictive cost-effectiveness approach (Appendix G). For the lecture training program, the scenario is that a trainer travels to 7 U.S. cities and instructs 50 individuals at each site. For the multimedia training program, the scenario is that a single multimedia worksite is set up in 7 U.S. cities at training facilities to accommodate 50 individuals at each site who come in for individual training. Costs were then determined for the incremental training of more individuals.

As Eicher (1980) notes, cost-effectiveness analysis only provides information that must be evaluated by the decision-maker; it does not provide objective and definite criteria of choice. Therefore, a decision of the most cost-effective training program was based on the results of the trainees' performance and response to the related questions on the post-test questionnaire.

When evaluated it was found that no significant or practical differences exist between the two training programs, and it was hypothesized that after the training programs were revised that the minimal differences would be even smaller. Without differences between the training programs the need to determine the importance of the training benefits and operational benefits is no longer relevant – both training programs have the same benefits. Considering that no practical differences exist between the training programs, the issue of which training program is the most cost effective must be based on the costs of the design, development, and dissemination of the training programs. A review of these costs is in order.

Regarding the cost to design and develop, the multimedia cost more than twice the lecture training program for a difference of $36,938.69. With this in mind, one must evaluate the multimedia training program to determine if that cost difference can be recouped during the dissemination. Excluding the trainees' costs, the multimedia training program costs almost twice that of the lecture training program for the first 350 trainees (50 trainees in 7 U. S. cities). However, for each increment of training, the multimedia training program costs approximately one-fourth that of the lecture training program. If the costs of the training programs are examined without the trainees' costs included, the costs to disseminate would be those shown in Table 5.3. It is evident that the large difference in cost for the multimedia training program can be recovered during dissemination.

Table 5.3. Costs to disseminate the Drowsiness Detection Training programs excluding trainees' expenses.

Number of Trainees		Cost of Computer-based Multimedia	Cost of Lecture
350	(1 circuit)	$42,119.00	$21,693.00
700	(2 circuits)	$46,669.00	$43,386.00
1,400	(4 circuits)	$55,769.00	$86,772.00
2,100	(6 circuits)	$64,869.00	$130,158.00
3,500	(10 circuits)	$83,069.00	$216,930.00
7,000	(20 circuits)	$128,569.00	$433,860.00

What gets diluted with the trainees' costs added is the significant savings of using the computer-based multimedia if many trainees are to be trained. The assumption is made that training is to be mandated and targeted at flight crew personnel and air traffic controllers. Since this is a large population, the cost differences become more significant. Furthermore, if refresher training is required periodically or the training program will be have a long lifespan, the multimedia option becomes more appealing.

However, there are other considerations that are not taken into account. Technology advancements is a serious concern for the multimedia training program. For instance, in the 18 months since this project started, the top-of-the-line computer used to design and develop the multimedia project was introduced into the market, dropped considerably in price, and is now no longer made. In fact, the method used to digitize, compress, and playback the video in the multimedia training program is obselete. The equipment necessary to disseminate the training program is currently available only through third party small name companies, and will likely not be available for new purchase in six months time. Under this prospect, if the computers for dissemination are not purchased in a short period of time, the video for the multimedia training program will need to be re-digitized and compressed for access by the training program software. However, a relevant note is that the latest generation of computers and video compression equipment is becoming more general and available for use with cross-platform computers and software. Nonetheless, this will of course involve added costs. The issue will be whether the added costs will be recovered during dissemination. If the training program does indeed need to reach a large number of trainees and has a moderate lifespan, it seems likely that these costs can be recovered using the multimedia training.

Another issue is the presentation style of the lecturer. It was stated previously that the generalizability of the study was dependent upon the degree to which the lecturer who presented the training program is representative of the lecturers available to disseminate the lecture training program. The quality of the presentation could affect training benefits such as motivation and potentially affect the level to which the trainees learn the presented material. If keeping a good public speaker on staff or if maintaining consistent quality between lecturers is a concern, especially considering the amount of travel that would be involved with such a position, then the multimedia approach would be a more favorable option.

Determining the most cost-effective training medium based upon these considerations, in light of the fact that the trainees performance in the lecture and multimedia training programs was similar, and with the knowledge that the computer-based multimedia is (currently) more cost-effective to disseminate, then the choice for the most cost-effective training medium is the computer-based multimedia.

CHAPTER 6

RESULTS AND DISCUSSION OF THE PREDICTIVE AND FORMATIVE APPROACH OUTCOMES

RESULTS

The outcome of the predictive approach to determining the most cost-effective training medium was based on the experts' opinion. All 14 of the experts chose the computer-based multimedia as the most cost-effective training medium. However, the experts' choice was based either soley or heavily on dollar amounts associated with design, development, and dissemination, with the data obtained through the validation process given little or no weight. All experts stated that it would not have been possible to use the information gathered through application of the Benefits Model to determine cost-effectiveness with any confidence.

The outcome of the formative evaluation approach to determining cost-effectiveness was determined by the outcome of the trainees' performance data and the cost to disseminate the training programs. The computer-based multimedia was chosen as the most cost-effective training medium.

DISCUSSION

It was expected that, from the results of this study, the use of variables for the *Benefits Model* would be better identified, classified, and quantified for the comparison of the two media – lecture and multimedia – for the training topic of drowsiness detection. Furthermore, it was expected that it would be determined whether the *Benefits Model* would serve as a validated, predictive cost-effectiveness model.

The *Benefits Model* Application Process

A few comments can be made regarding the identification of variables. While selecting the variables it became clear that not all variables that could impact the training program could be selected. Also, to compare the cost-effectiveness models, the variables that were selected had to be those that could be tested through formative evaluation in a reasonable amount of time. However, it does not seem plausible that a different set of variables would have resulted in a significantly different outcome for the comparison of the models. The more obvious shortcomings of the predictive use of Kearsely and Compton's *Benefits Model* were observed during the classification and quantification processes.

As reported in Chapter 3, the classification process was very vague for most subjects. Although the concept is to determine a causal relationship between variables to quantify the worth of a benefit to a training program, the actual task of classification is very far removed from the concept. Ironically, this gap is even more pronounced for the experts who do not have the same intimacy with the project and the desires of the funding agency as the designer. The experts in this study never seemed to connect the classification task with the goal of the model.

The experts suggested more definitions and more examples to aid with the classification task. Perhaps a better approach would have been to call all the experts together for a meeting in which they were more fully informed of the training program, the background of the work, and why the particular variables were identified. Also, another aspect of informing the experts might have been to explain the *Benefits Model* in detail at the start of the project. Of course, the problem with this refined approach is the time and effort it would take to bring together 14 people for such an information session. The time lost while simply organizing such a gathering would likely be seen as critical time for any business.

The quantification process was also difficult for most of the experts. The problem that the experts had with using the concept of correlating the variables can be clarified in a future study. However, that was a relatively minor problem. The larger issue for most of the experts was the constraint of selecting either a positive or negative relationship between two variables. The idea for a rating scale was mentioned by the experts repeatedly. Also, some experts mentioned that it would have been easier to combine the classification and quantification processes. An alternate approach to this study could be to explain the model thoroughly and leave the experts to classify and quantify the variables as each person saw fit. Perhaps a marker board and pen would be well suited to making many changes in the relationships and quantifications until each expert was happy with the end result. Likewise, this approach could be used for the experts as a group to help brainstorming and new ways of considering the aspects of each variables impact on the training program.

Other Considerations for Future Research

One aspect of this comparison of a predictive and formative cost-effectiveness model is the degree to which the process of comparison depended upon the outcome of the formative evaluation of the Drowsiness Detection Training program. One concern is that the trainees used for the stduy were not flight crew personnel and air traffic controllers, but junior and senior level undergraduate students who volunteered for the study. Although an effort was made to match the subject group as closely as possible to the actual group, the possibility may exist that some differences in the population would cause a difference in performance between the two training programs. Another issue is the experimenter's interpretation of the subjects' data. The only significant difference that did exist was in the ability to rate the level of drowsiness; however, this was a very small difference of only 2 points of 100 possible. If another researcher chose to interpret that data differently, it could be decided that the lecture training program met an operational benefit that the multimedia training program did not. In this case, the lecture training program would then be determined to be the most cost-effective by default, regardless of any other considerations.

Another concern has to do with the experts themselves. The volunteers chosen were those who had completed a training program design course at Virginia Polytechnic Institute and State University. To someone who has been working in the field of training for many years, these graduate students would not likely be considered experts. This contention would, of course, invalidate the comparison of the cost-effectiveness models. In defense of this type of arguement, it can be said that if the model depends on the availability of many seasoned training experts, it is not a practical model for application.

A third consideration for the validity of this study is the experimenter's interpretation of the model. The model was chosen for elaboration because, in theory, it incorporated aspects of a training program that are rarely considered for cost-effectiveness comparison, namely, the idea of the quantification of training benefits. However, the model itself is not well documented nor clearly explained, requiring the interpretation of the model based on the authors' explanation and examples. Their approach did leave some aspects of the model open for interpretation. One aspect that was unclear was the method in which the authors relate all variables back to the training medium. The example given on page 12 shows how each variable relates to the training medium of multimedia. This appears relatively straightforward until considered with the concept that each variable is to be related to its impact on every other variable. To perform this task, the expert must relate the variables to each other considering the training medium. The example on page 12 disregards relating the variables to each other, in which case the concept of determining a causal relationship between variables in order to quantify the worth of a benefit to a training program is lost.

A final consideration is whether using such a model is worth the effort. Clearly, the model as applied in this study is not worth the effort. The data gleened from the experts as applied to the model were nondiscript at best. However, this is not to say that more research and refinement of the model will not improve the information from the model. If refinements of the model does result in a more useful output, the time involved may be significantly less costly than developing two separate training system prototypes and testing them through formative evaluation. Furthermore, if one considers that formative evaluation is still not an industry standard, the cost of summative evaluation makes such a model look even more appealing.

Future Research Implications

Several issues of this study were mentioned for consideration. Future research into the area of cost-effectiveness models will need to address these issues. Once a predictive cost-effectiveness model has been validated for a drowsiness detection training program, the doorway will be open to several future studies. First, it would be a contribution to investigate the most cost-effective way to design each training medium. This would entail the identification, classification, and quantification of variables that are specific to each medium, regardless of the training topic. Such research would provide a platform for training program analysts to choose appropriate media. From that platform the analyst could identify the benefits that are particular to a specific topic, then have a complete list to perform the classification and quantification of variables in a cost-effectiveness model.

Second, future studies should evaluate the impact of the assumption made in this study that the lecturer for the lecture training program is representative of the lecturer population. The impact of this assumption is that the generalizability of cost effectiveness is restricted to the level to which the one lecturer represents the lecturer population for the drowsiness detection training topic. The problem with this assumption is that all lecturers are not equally as interesting, as motivating, or simply as skilled at getting the material across to the learners.

It would also be very worthwhile to validate the -1 to +1 value scheme used in the Kearsley and Compton (1981) model as it was applied to the training media. The value scheme applied in this study was not empirically evaluated beforehand. Without proper validation, one cannot determine to what extent it is affecting the outcome of the cost-effectiveness results.

Finally, another line of research that would be worthwhile would be the validation of a weighting scheme for the B_j benefits; in other words, the benefit being quantified. This is not to be confused with the assignment of values to the B_i benefits as they apply to the training medium used. The premise here is that not all benefits of the training program are likely to be considered equally important. A starting point to validate a weighting study would be to examine the work of Edwards and Newman (1982). They proposed three weighting schemes for the evaluation of attributes in social programs: equal or unit weighting, weights from ranks, and ratio weighting. Such a weighting scheme applied to Kearsley and Compton's model may further strengthen the model and enhance the predictive cost-effectiveness model accuracy.

REFERENCES

Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In I. L. Goldstein (Ed.), Training and Development in Organizations (pp. 32). San Francisco, CA: Jossey-Bass Publishers.

Cohen, S. I. (1985). A cost-benefit analysis of industrial training. Economics of Education Review, 4(4), 327-339.

Cullen, J. G., Sawsin, S. A., Sisson, G. R., and Swanson, R. A. (1978). Cost effectiveness: A model for assessing the training investment. Training and Development Journal, 1, 24-29.

Edwards, W. and Newman, J. R. (1982). Multiattribute evaluation. Beverly Hills, CA: SAGE Publications.

Eicher, J. (1980). Results and problems: Some thoughts on the economic analysis of new educational media. In The Economics of New Educational Media: Volume 2, Cost and Effectiveness (pp. 9-15). France: The UNESCO Press.

Finkle, C. L. (1987). The true cost of a training program. Training and Development Journal, 9, 74-78.

Gordon. S. E. (1994). Systematic training program design: Maximizing effectiveness and minimizing liability. Englewood, Cliffs, NJ: Prentice Hall.

Hawthorne, E. M. (1987). Evaluating employee training programs: A research-based guide for human resources managers. New York: Quorum Books.

Kearsley, G. (1983). Computer-based training: a guide to selection and implementation. Reading, MA: Addison-Wesley Publishing Company.

Kearsley, G. and Compton, T. (1981). Assessing costs, benefits and productivity in training systems. Training and Development Journal, 1, 52-61.

Mirabel, T. E. (1978). Forecasting future training costs. Training and Development Journal, 7, 78-87.

Mohr, L. B. (1988). Impact analysis for program evaluation. Homewood, IL: Dorsey Press.

Neale, V. L. and Wierwille, W. W. (1995a, March). Research on vehicle-based driver status/performance monitoring: Observer training assessment for a drowsiness detection training program: Knowledge, skills, and training objectives. (ISE Tech. Report No. 95-01). Blacksburg, VA: Virginia Polytechnic Institute and State University, Vehicle Analysis and Simulation Laboratory.

Neale, V. L. and Wierwille, W. W. (1995b, December). Research on vehicle-based driver status/performance monitoring: Part II. Empirical evaluation of lecture and computer-based multimedia format drowsiness detection training programs. (ISE Tech. Report No. 95-24). Blacksburg, VA: Virginia Polytechnic Institute and State University, Vehicle Analysis and Simulation Laboratory.

Reynolds, A. and Anderson, R. H. (1992). Selecting and developing media for instruction (3rd ed.). New York: Van Nostrand Reinhold.

Rosenthal, S. M. and Mezoff, B. (1980). Improving the cost/benefit of management training. Training and Development Journal, 12, 102-106.

Salas, E., Burgess, K. A., and Cannon-Bowers, J. A. (1995). Training effectiveness techniques. In Weimer (Ed.), Research Techniques in Human Engineering, (pp. 439-471). Englewood, Cliffs, NJ: Prentice Hall.

Steinmetz, C. S. (1967). The evolution of training. In R. L. Craig and L. R. Bittel (Eds.), Training and Development Handbook (pp. 1-15). New York: McGraw-Hill Book Company.

Tannenbaum, S. I., Cannon-Bowers, J. A., Salas, E., and Mathieu, J. E. (1993). Factors that influence training effectiveness: A conceptual model and longitudinal analysis. (Tech. Report No. 93-011). Orlando, FL: Naval Training Systems Center: Human, Systems Integration Division.

Wierwille. W. W. and Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. Accident Analysis and Prevention, 26(4), 571-578.

APPENDICES

Appendix A:

Expert Informed Consent Form.

Expert's Informed Consent Form

You are being asked to volunteer to be a subject in a research project. The purpose of the research is to determine the most cost-effective method for teaching individuals how to determine the level of drowsiness of colleagues during safety critical crew operations. Your part in this research, if you choose to participate, will be to give your opinion of the relationship between the variables that make up the cost-effectiveness model being used to analyze two training programs.

This is a three part study. The first session will consist of answering questions in a yes or no format about the impact of each training variable on another training variable. This will take about one hour. The second session will consist of answering questions about the impact of each training variable on another training variable in a numeric format considering the training media. This will take about one hour. The third session will consist of determining which of two training programs you consider to be the most cost-effective based on the summarized cost information and the information on the relationship of the training variables. This will take approximately 10 minutes. You are being asked to complete the first session today, the second session in one week, and the third session in about one month.

For each of the first two sessions, all the questions you answer will be in a HyperCard Player software program. You will be asked to sit in a quiet office space at a computer workstation with the program running and answer the questions on each card by clicking on one of the possible answers. Any instructions or definitions you need will be provided in the program.

There are no known risks to you in this experiment, and discomforts are believed to be minimal: You may incur some slight discomfort associated with sitting in one seat for a period of time. However, you are encouraged to get up from your seat to stretch and move about if you would like. The benefit to you as a participant is the opportunity to contribute to education in the area of training program development and issues of cost-effectiveness of training media. You should understand that no promise or guarantee of benefits has been made to encourage you to participate in this study.

Your anonymity will be maintained. A number will be used to replace your name in order to assure that your name will not be associated with your data.

You will be paid $6.00 for the first session, $6.00 for the second session, and $3.00 for the third session for a total of $15.00. Payment will be made at the end of each session.

The principle investigator, Dr. Walter W. Wierwille (231-7952), and the researcher Vicki L. Neale (231-9084), will answer any questions you may have about your participation. You should not sign this informed consent form until you are satisfied that you understand all of the previous descriptions and conditions. You should further be aware that you may contact Dr. Ernest Stout, Chairman of the University's Institutional Review Board, if you have any questions or concerns about this experiment. His phone numbers are 231-9359 and 231-6077.

You should know that at any time you are free to withdraw from participation in this research program for any reason without penalty.

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University.

I have read and understand this Informed Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

______________________________________ Date:______________

Participant's Signature:

______________________________________ Date:______________

Experimenter's Signature:

Appendix B:

Expert Instructions Included in the Session 1 HyperCard Player Program.

Expert Instructions for the Classification of Training Variables

The cost-effectiveness of two training media is being compared. For the purposes of this study, a cost-effectiveness analysis will be defined as:

> The input into the training program based on the
> training system parameters (constraints of the training program)
> as compared to the output from the training program based on the
> operational benefits (goals of the program)and the
> training benefits (an artifact of the training medium chosen).

To do this cost-effectiveness analysis, the causal link, or relationship, between training variables that affect the cost-effectiveness of each program must be compared.

A lecture training program and a computer-based multimedia training program are being compared. These training programs are designed to teach individuals how to rate the level of drowsiness of a colleague during crew operations.

The lecture training program will be conducted in a group-paced format, and the computer-based multimedia program will be conducted in a self-paced format. The training programs consist of an information session broken into five sub-categories, a practice session of rating the drowsiness level of individuals in video segments, and if necessary, a remediation session which consists of more practice at rating video segments.

It is worth noting that both training programs are designed such that the trainees will never be aware they are receiving remediation. Whether or not the trainees receive remedial training is dependent upon their response to the first practice sessions. For the lecture training program, the lecturer asks the trainees to provide their rating was by a show of hands. From this information, the lecturer determines if the trainees need remedial sessions. For the computer-based multimedia training program, the program is designed with an algorithm that gives the trainee remedial practice if the responses to previous ratings are not in an acceptable range. After completion of the training program and the practice sessions, participants will be tested for their ability to rate the level of drowsiness of individuals in video segments.

You are being asked to give your opinion about the relationship of each training variable that makes up the cost-effectiveness model to each other training variable. This is referred to as the classification of the training variables.

There are 24 cards in this HyperCard stack with several relationships on each card. Each card will refer to either a lecture training program or a multimedia training program. For each variable, click on every variable listed that impacts it. Until you go to the next card you can change your responses as much as necessary; however, once you go to the next card, you cannot go back.

The variables that you will be comparing are defined on a sheet of paper next to the computer. Refer to the definitions as often as necessary.

Appendix C:

Hardcopy Version of HyperCard Player Program for Session 1.

Instructions for the Classification of Training Variables
(Scroll down to read all the instructions).

The cost-effectiveness of two training media is being compared. For the purposes of this study, a cost-effectiveness analysis will be defined as:

The input into the training program based on the training system parameters (constraints of the training program) as compared to the output from the training program based on the operational benefits (goals of the program)and the training benefits (an artifact of the training medium chosen).

To do this cost-effectiveness analysis, the causal link, or relationship, between training variables that affect the cost-effectiveness of each program must be compared.

Go Next

This set of questions refers to the lecture training program previously described.

Go Next

Will the **financial resources** available for the **lecture** training program affect the: (check all that apply)

- ☐ human resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **human resources** available for the **lecture** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **equipment and tools** available for the **lecture** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **time resources** available for the **lecture** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ equipment and tools
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **number of trainees being trained** with the **lecture** training program affect the: (check all that apply)

- ☐ **financial resources**
- ☐ **human resources**
- ☐ **equipment and tools**
- ☐ **time resources**
- ☐ **number of locations at which trainees can be taught**
- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **number of locations at which trainees are taught** with the **lecture** training program affect the: (check all that apply)

- ☐ **financial resources**
- ☐ **human resources**
- ☐ **equipment and tools**
- ☐ **time resources**
- ☐ **number of trainees who can be trained**
- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's time to complete** the **lecture** training program affect the: (check all that apply)

- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **number of remedition sessions necessary** with the **lecture** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **developer's time to revise** the **lecture** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's motivation** with the **lecture** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **developer's time to revise the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's self-efficacy** with the **lecture** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's convenience to attend** the **lecture** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

This next set of questions refers to the computer-based multimedia training program previously described.

Go Next

Will the **financial resources** available for the **multimedia** training program affect the: (check all that apply)

- ☐ human resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **human resources** available for the **multimedia** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **equipment and tools** available for the **multimedia** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **time resources** available for the **multimedia** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ equipment and tools
- ☐ number of trainees who can be trained
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **number of trainees being** trained with the **multimedia** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of locations at which trainees can be taught
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **number of locations at which trainees are taught** with the **multimedia** training program affect the: (check all that apply)

- ☐ financial resources
- ☐ human resources
- ☐ equipment and tools
- ☐ time resources
- ☐ number of trainees who can be trained
- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **trainee's time to complete the training program** with the **multimedia** training program affect the: (check all that apply)

- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's convenience to attend the training program
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

Will the **number of remedition sessions necessary** with the **multimedia** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **developer's time to revise** the **multimedia** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **trainee's motivation**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's motivation** with the **multimedia** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's self-efficacy**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the **trainee's self-efficacy** with the **multimedia** training program affect the: (check all that apply)

- ☐ **trainee's time to complete the training program**
- ☐ **number of remediation sessions necessary**
- ☐ **developer's time to revise the training program**
- ☐ **trainee's motivation**
- ☐ **trainee's convenience to attend the training program**
- ☐ **trainee's ability to rate the level of drowsiness**
- ☐ **trainee's ability to recall presented information**

Go Next

Will the trainee's convenience to attend the multimedia training program affect the: (check all that apply)

- ☐ trainee's time to complete the training program
- ☐ number of remediation sessions necessary
- ☐ developer's time to revise the training program
- ☐ trainee's motivation
- ☐ trainee's self-efficacy
- ☐ trainee's ability to rate the level of drowsiness
- ☐ trainee's ability to recall presented information

Go Next

This completes Session 1 of this study. Thank you for participating.

Appendix D:

Variables Defined in the Session 1 (Classification) HyperCard Player Program.

Financial resources - The monetary resources available for the design, development, and dissemination of a training program.

Human resources - The amount of person hours available for the design, development, and dissemination of a training program.

Equipment and tools - The hardware, software, paper, etc. necessary for the design, development, and dissemination of a training program.

Time resources - The amount of time, in hours, that is available for the design, development, and dissemination of a training program.

Number of trainees that can be trained - The training program will have many people that require training (number in the hundreds).

Number of locations at which trainees can be taught - The people that will participate in the training program live in several cities across the United States.

Trainee's time to complete the training session - The time for a trainee to complete the training session, the practice trials, and any remediation that may be necessary.

Number of remediation sessions necessary - The number of extra sessions the trainee may need after the regular practice sessions.

Developer's time to revise the training program - The amount of developer's hours necessary to revise the training program after formative evaluation.

Trainee's motivation - The level to which a trainee wants to learn the presented material.

Trainee's self-efficacy - The trainee's belief in his or her ability to perform a specific task.

Trainee's convenience of attending training - Considering the training media, how convenient the trainee will find attending the training session.

Trainee's ability to rate level of drowsiness - The level of mastery of a trainee on the training topic.

Trainee's ability to recall information - The ability of the trainee to recall information taught in the training program.

Appendix E:

Hardcopy Verson of HyperCard Player Program for Session 2

Instructions for the Quantification of Causal Relationships Between Variables (Scroll down to read all the instructions).

The cost-effectiveness of two training media is being compared. For the purposes of this study, a cost-effectiveness analysis will be defined as:

The input into the training program based on the training system parameters (constraints of the training program) as compared to the output from the training program based on the operational benefits (goals of the program)and the training benefits (an artifact of the training medium chosen).

To do this cost-effectiveness analysis, the causal link, or relationship, between training variables that affect the cost-effectiveness of each program must be compared.

Go Next

This set of questions refers to the lecture training program previously described.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **financial resources** available affect:

- the human resources
- equipment and tools
- number of locations at which the trainees can be taught

The lecture training program will require
an average amount of financial resources for design
a low amount of financial resources for development, and
a high amount of financial resources for dissemination.

Please determine how each relationship is impacted by the financial resources available considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Financial resources ○ positively +1 ○ negatively -1 affect human resources available.

Financial resources ○ positively +1 ○ negatively -1 affect equipment and tools necessary.

Financial resources ○ positively +1 ○ negatively -1 affect the number of locations at which trainees can be taught.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **human resources** available affect:

- time resources
- the number of trainees that can be trained
- number of locations at which the trainees can be taught

The lecture training program will require
an average amount of human resources for design
a low amount of human resources for development, and
a high amount of human resources for dissemination.

Please determine how each relationship is impacted by human resources available considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Human resources ○ positively +1 ○ negatively -1 affect time resources available.

Human resources ○ positively +1 ○ negatively -1 affect the number of trainees that can be trained.

Human resources ○ positively +1 ○ negatively -1 affect the number of locations at which trainees can be taught.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **equipment and tools** necessary affect:
– the developer's time to revise the training program

The lecture training program will require
an low amount of equipment and tools for design
a low amount of equipment and tools for development, and
a low amount of equipment and tools for dissemination.

Please determine how each relationship is impacted by the equipment and tools necessary considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Equipment and tools ○ positively +1 ○ negatively -1 affect the developer's time to revise the training program.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **time resources** available affect:
– the number of trainees that can be trained
– the trainee's time to complete the training program
– the developer's time to revise the training program revision

The lecture training program will require
an average amount of time resources for design
a low amount of time resources for development, and
a low amount of time resources for dissemination.

Please determine how each relationship is impacted by time resources available considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Time resources ○ positively +1 ○ negatively -1 affect the number of trainees that can be trained.

Time resources ○ positively +1 ○ negatively -1 affect the trainee's time to complete the training program.

Time resources ○ positively +1 ○ negatively -1 affect the developer's time to revise the training program.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of trainees that can be trained** affect:

- human resources
- equipment and tools
- time resources

The lecture training program will be available to many trainees (number in the hundreds).

Please determine how each relationship is impacted by the number of trainees that can be trained when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of trainees who will be trained ○ positively +1 ○ negatively -1 affect the human resources available.

The number of trainees who will be trained ○ positively +1 ○ negatively -1 affect the equipment and tools necessary.

The number of trainees who will be trained ○ positively +1 ○ negatively -1 affect the time resources available.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of locations at which trainees will be taught** affect:

- financial resources
- human resources
- equipment and tools
- time resources
- the number of trainees that will be trained
- the trainee's convenience to attend training

The lecture training program will be taught in several cities across the United States.

Please determine how each relationship is impacted by the number of locations at which trainees will be taught when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the financial resources available.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the human resources available.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the equipment and tools necessary.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the time resources available.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the number of trainees who will be trained.

The number of locations at which trainees will be taught ○ positively +1 ○ negatively -1 affect the trainee's convenience to attend training.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's time to complete the training program** affect:

– trainee's motivation
– trainee's self-efficacy

It is estimated to take two hours to complete the lecture training program.

Please determine how each relationship is impacted by the trainee's time to complete the training program when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's time to complete the training program will ○ positively +1 ○ negatively -1 affect the trainee's motivation.

The trainee's time to complete the training program will ○ positively +1 ○ negatively +1 affect the trainee's self-efficacy.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of remediation sessions necessary** affect:

– the trainee's time to complete the training program
– the trainee's motivation

The lecture training program includes up to three remediation sessions. The lecturer gives remediation to the class as a whole based upon the group response to each of five previous practice segments. The trainees will not know the extra sessions are remedial training; instead, they will simply think it is another practice session.

Please determine how each relationship is impacted by the number of remediation sessions necessary when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of remediation sessions necessary will ○ positively +1 ○ negatively -1 affect the trainee's time to complete the training program.

The number of remediation sessions necessary will ○ positively +1 ○ negatively +1 affect the trainee's motivation.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's motivation** will affect:

– the number of remediation sessions necessary
– trainee's self-efficacy
– the trainee's ability to rate the level of drowsiness
– the trainee's ability to recall presented information

The lecture training program may be more or less motivating for the trainees than the multimedia training program based upon several factors.

Please determine how each relationship is impacted by the trainee's motivation when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the number of remediation sessions necessary.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's self-efficacy.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's ability to rate the level of drowsiness.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's ability to recall presented information.

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's self-efficacy** will affect:

– the number of remediation sessions necessary
– the trainee's motivation
– the trainee's ability to rate the level of drowsiness

The lecture training program may create more or less self-efficacy in a trainee than the multimedia training program based upon several factors.

Please determine how each relationship is impacted by the trainee's self-efficacy when considering that a lecture training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the number of remediation sessions necessary.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the trainee's motivation.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the trainee's ability to rate the level of drowsiness.

Go Next

This set of questions refers to the computer-based multimedia training program previously described.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **financial resources** available affect:

- the human resources
- equipment and tools
- time resources
- the developer's time to revise the training program

The multimedia training program will require
an average amount of financial resources for design
a high amount of financial resources for development, and
a high amount of financial resources for dissemination.

Please determine how each relationship is impacted by the financial resources available considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Financial resources ○ positively +1 ○ negatively -1 affect human resources available.

Financial resources ○ positively +1 ○ negatively -1 affect equipment and tools necessary.

Financial resources ○ positively +1 ○ negatively -1 affect the time resources available.

Financial resources ○ positively +1 ○ negatively -1 affect the developer's time to revise the training program.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **human resources** available affect:

– time resources
– the developer's time to revise the training program

The multimedia training program will require
an average amount of human resources for design
a high amount of human resources for development, and
a high amount of human resources for dissemination.

Please determine how each relationship is impacted by human resources available considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Human resources ○ **positively +1** ○ **negatively -1** affect time resources available.

Human resources ○ **positively +1** ○ **negatively -1** affect the developer's time to revise the training program.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **equipment and tools** necessary affect:

– time resources
– number of trainees that will be trained
– the number of locations at which trainees can be taught
– the developer's time to revise the training program

The multimedia training program will require
an average amount of equipment and tools for design
a high amount of equipment and tools for development, and
a high amount of equipment and tools for dissemination.

Please determine how each relationship is impacted by the equipment and tools necessary considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

Equipment and tools ○ **positively +1** ○ **negatively -1** affect the time resources available.

Equipment and tools ○ **positively +1** ○ **negatively -1** affect the number of trainees that will be trained.

Equipment and tools ○ **positively +1** ○ **negatively -1** affect the number of locations at which trainees can be taught.

Equipment and tools
affect the developer's ○ **positively +1** ○ **negatively -1**

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of trainees that can be trained** affect:

– equipment and tools

The multimedia training program will be available to many trainees (number in the hundreds).

Please determine how each relationship is impacted by the number of trainees who will be trained when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of trainees who will be trained ○ **positively +1** ○ **negatively -1** affect the equipment and tools necessary.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of locations at which trainees will be taught** affect:

– equipment and tools
– time resources
– the trainee's convenience to attend training

The multimedia training program will be taught in several cities across the United States.

Please determine how each relationship is impacted by the number of locations at which trainees will be taught when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of locations at which trainees will be taught affect the equipment and tools necessary. ○ **positively +1** ○ **negatively -1**

The number of locations at which trainees will be taught affect the number of trainees who will be trained. ○ **positively +1** ○ **negatively -1**

The number of locations at which trainees will be taught affect the trainees convenience to attend training. ○ **positively +1** ○ **negatively -1**

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's time to complete the training program** affect:

– the number of remediation sessions necessary
– trainee's motivation-

It is estimated to take two hours to complete the multimedia training program.

Please determine how each relationship is impacted by the trainee's time to complete the training program when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's time to complete the training program will affect the number of remediation sessions necessary. ○ positively +1 ○ negatively -1

The trainee's time to complete the training program will affect the trainee's motivation. ○ positively +1 ○ negatively +1

Go Next

For a lecture training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **number of remediation sessions necessary** affect:

– the trainee's time to complete the training program
– the trainee's motivation
– the trainee's self-efficacy

The multimedia training program includes up to three remediation sessions. Remediation is given by a computer algorithm based upon the trainee's response to each of five previous practice segments. The trainees will not know the extra sessions are remedial training; instead, they will simply think it is another practice session.

Please determine how each relationship is impacted by the number of remediation sessions necessary when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The number of remediation sessions necessary will trainee's time to complete the training program. ○ positively +1 ○ negatively -1 affect the

The number of remediation sessions necessary will trainee's motivation. ○ positively +1 ○ negatively +1 affect the

The number of remediation sessions necessary will affect the trainee's self-efficacy. ○ positively +1 ○ negatively +1

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's motivation** will affect:

– the time to complete the training program
– the number of remediation sessions necessary
– trainee's self-efficacy
– the trainee's ability to rate the level of drowsiness
– the trainee's ability to recall presented information

The multimedia training program may be more or less motivating for the trainees than the lecture training program based upon several factors.

Please determine how each relationship is impacted by the number of locations at which trainees will be taught when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the time to complete the training program.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the number of remediation sessions necessary.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's self-efficacy.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's ability to rate the level of drowsiness.

The trainee's motivation will ○ positively +1 ○ negatively -1 affect the trainee's ability to recall presented information.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's self-efficacy** will affect:

– the trainee's time to complete the training program
– the number of remediation sessions necessary
– the trainee's motivation

The multimedia training program may create more or less self-efficacy in a trainee than the lecture training program based upon several factors.

Please determine how each relationship is impacted by the number of trainees who will be trained when considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the trainee's time to complete the training program.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the number of remediation sessions necessary.

The trainee's self-efficacy will ○ positively +1 ○ negatively -1 affect the trainee's motivation.

Go Next

For a multimedia training program that is designed to teach individuals how to determine the level of drowsiness of colleagues, expert opinion has determined that the **trainee's convenience to attend training** will affect:

– the trainee's motivation

The multimedia training program will be available to a trainee at a time the trainee has scheduled in a relatively local Department of Motor Vehicles (DMV) or flight school.

Please determine how each relationship is impacted by the trainee's convenience to attend training considering that a multimedia training program will be used. Select your answer by checking on only one of the two radio buttons in each sentence.

The trainee's convenience to attend trainning will ○ **positively +1** ○ **negatively -1** affect the trainee's motivation.

Go Next

This concludes Session 2.
Thank you for participating.

Appendix F:

Expert Instructions Included in the Session 2 (Quantification) HyperCard Player Program.

Expert Instructions for the Quantification of Training Variables

The cost-effectiveness of two training media is being compared. For the purposes of this study, a cost-effectiveness analysis will be defined as:

> The input into the training program based on the
> training system parameters (constraints of the training program)
> as compared to the output from the training program based on the
> operational benefits (goals of the program)and the
> training benefits (an artifact of the training medium chosen).

To do this cost-effectiveness analysis, the causal link, or relationship, between training variables that affect the cost-effectiveness of each program must be compared.

A lecture training program and a computer-based multimedia training program are being developed. These training programs are designed to teach individuals how to rate the level of drowsiness of a colleague during crew operations.

The lecture training program will be conducted in a group-paced format, and the computer-based multimedia program will be conducted in a self-paced format. The training programs consist of an information session broken into five sub-categories, a practice session of rating the drowsiness level of individuals in video segments, and if necessary, a remediation session which consists of more practice at rating video segments.

It is worth noting that both training programs are designed such that the trainees will never be aware they are receiving remediation. Whether or not the trainees receive remedial training is dependent upon their response to the first practice sessions. For the lecture training program, the lecturer asks the trainees to tell what their rating was by a show of hands. From this information, the lecturer determines if the trainees need remedial sessions. For the computer-based multimedia training program, the program is designed with an algorithm that gives the trainee remedial practice if the responses to previous ratings are not in an acceptable range. After completion of the training program and the practice sessions, participants will be tested for their ability to rate the level of drowsiness of individuals in video segments.

Your last session consisted of the classification of the training variables. Now, you are being asked to give your opinion about the value of the relationship of each variable to each other variable. This is referred to as the quantification of the training variables. You may occasionally be asked to quantify a relationship between variables that you did not pair as impacting each other during the classification stage. This means simply that a significant number of the experts agreed that a relationship between the variables did exist. In this situation, you are asked to quantify the relationship with the assumption that the significant outcome of the classification session was valid.

Each HyperCard in this stack has a question set on it, and there are a total of 14 question sets. Each set has from one to 13 questions in it. Please read each question very carefully and select the appropriate completion to the sentence by clicking on one of the two choices available for each sentence. Until you go to the next card you can change your responses as much as necessary; however, once you go to the next card, you cannot go back.

Appendix G:

Cost Accounting Worksheets for the Dissemination of the Lecture and Computer-based Multimedia Training Programs.

Training Program Dissemination Costs

The following dissemination scenarios are hypothetical. Several assumptions were made to develop reasonable costs: Since this program is to be disseminated by a federal government agency it was assumed that Washington, D. C. would be the headquarters for the dissemination of the training program. Seven cities were selected for training. These cities are:

- Washington, D. C.
- Cedar Rapids
- Oklahoma City
- Tampa
- Los Angeles
- Portland
- Salt Lake City

It is assumed the 50 people will go to each of the seven sites for training for a total of 350 trainees. Trainees are flight crew personnel and air traffic controllers.

Trainees Time and Travel Expenses (considered equal for each training program)

<u>Trainees Traveling by Automobile</u>	
Three-fourths day pay (average $55,000/year)	$158.65
Fringe @ 30%	$47.60
Auto per diem for 50 mile average each way ($0.30/mile)	$30.00
Total per trainee	$236.25
Total for 20 Trainees Traveling by Automobile	$4,725.00
<u>Trainees Traveling by Commuter Air</u>	
One day pay (average $55,000/year)	$211.54
Fringe @ 30%	$63.46
Air travel (average flight of $350.00)	$350.00
Taxi from airport to training site (average $18.00 each way)	$36.00
Total per trainee	$661.00
Total for 30 Trainees Traveling by Commuter Air	$19,830.00
Total Trainee Expense Per Site	$24,555.00
Total Trainee Expense for All Sites	**$171,885.00**

Multimedia Training Program

Scenario: Computer-based multimedia training systems will be set up in the seven cities listed above. All computer hardware will first be shipped to headquarters. At headquarters a 7100 series AV card and a Video Spigot Power AV card will be installed into each computer. Also, the training program will be loaded onto each external drive using the original development computer system. Then one complete system will be shipped to each of the seven training sites. An individual at the training site will configure the equipment following the manufacturer's instructions. A short, separate document will be included to explain general information regarding the training program. It is assumed that 50 people will come in for training at each training site.

<u>Hardware for each site</u>	
[1]Power Macintosh 7100/80	$1,399.00
[1]7100 series AV card	$499.00
[2]Quantum external 2.2 GB drive	$675.00
[2]Keyboard	$69.00
[2]Display	$1,844.00
[3]Video Spigot Power AV card	$599.00
Shipment to each site	$30.00
Total Hardware for each Training Site	$5,115.00
Total Hardware for All Sites	$35,805.00
Shipment of equipment directly to Headquarters (special bulk rate)	$100.00
* Time of person to install AV boards and load training program (8 hours at $20.00/hour)	$160.00
* Time of person to schedule trainees and handle travel reimbursements ($12.50/hour for 10 hours/week for 4 weeks at each site)	$3,500.00
* Time of person to set up hardware (8 hours at $20.00/hour at each site)	$1,120.00
Fringe for personnel time @ 30% (starred items)	$1,434.00
Total	**$42,119.00**
Trainees' Time	$171,885.00
Total + Trainee's Time	**$214,004.00**

[1] Prices quoted by Shreve Systems as of 3/11/96.

[2] Prices quoted by GTSI, as of 3/6/96. Rates apply to a Government Services Agency.

[3] Price quoted by Club Mac as of 3/11/96.

Lecture Training Program

Scenario: A lecture training program will be taught in the seven cities listed previously. Four of the training sites will be government facilities to be used at no cost. Three of the sites will be conference hotels with appropriate facilities. One lecturer will teach at each of the seven cities. The lecturer will dedicate five weeks of time to the preparation and training of the classes. The lecturer's salary is $45,000 per year. Two same-day lectures will be held at each site with twenty-five trainees in attendance for each class, for a total of 50 trainees at each site.

(Week one) First half of training circuit
Lecture 1 Washington, D. C. (commercial flight to Cedar Rapids)
Lecture 2 Cedar Rapids (commercial flight to Oklahoma City)
Lecture 3 Oklahoma City (commercial flight to Tampa)
Lecture 4 Tampa (commercial flight back to DC)

(Week two) Second half of training circuit (leaving from DC)
Lecture 5 Los Angeles (commercial flight to Portland)
Lecture 6 Portland (commercial flight to Salt Lake City)
Lecture 7 Salt Lake City (commercial flight back to DC)

The equipment list for the lecture training program is as follows:

large screen television
video-cassette player
overhead projector
screen
lecturer's overheads
lecturer's notes
trainee's handouts

It is assumed that the training site will have the television, video-cassette player, overhead projector, and screen. The lecturer will bring the overheads, notes, and the original of the trainee handouts. The trainee handouts will be copied in each city to reduce the carry-on items for the lecturer.

<u>Lecturer's Time and Travel Expenses</u>

Lecturer's Salary (five weeks at $45,000/year) $4,326.92
(Time includes scheduling flights, hotels, and classrooms.)
Fringe @ 30% $1,298.08

Lecturer commercial air travel
First training circuit $2,000.00
Second training circuit (leaving from DC) $2,200.00

Lecturer per diem rate
Washington, D. C not included for headquarters training site

Cedar Rapids
$94.00 lodging/night
$44.00 meals/day
for two days and two nights $276.00

Oklahoma City
$94.00 lodging/night
$44.00 meals/day
for two days and two night $276.00

Tampa
$94.00 lodging/night
$44.00 meals/day
for two days and one night $182.00

Los Angeles
$112.00 lodging/night
$48.00 meals/day
for two days and two night $320.00

Portland
$94.00 lodging/night
$44.00 meals/day
for two days and two night $276.00

Salt Lake City
$87.00 lodging/night
$48.00 meals/day
for two days and one night $183.00

Taxi service for employment relevant travel ($40.00/day for 12 days).................$480.00

<u>Other Costs</u>

Hotel conference room charge and equipment rental for three sites.................. $3,750.00

Coffee and juice service $2.50 per trainee for each site ($125.00/site)................$875.00

* Time of person to schedule trainees and handle travel reimbursements
($12.50/hour for 10 hours/week for 4 weeks at each site)........................... $3,500.00

Fringe for personnel time @ 30% (starred items).................................... $1,050.00

Copying expenses for trainee handout
20 sheets of paper at $0.10 each for 50 people for each site............................$700.00

Total $21,693.00
Trainees' Time $171,885.00
Total + Trainees' Time..$193,578.00

Dissemination Costs for Incremental Training

The following dissemination costs consider the financial impact of a second increment of training. It is assumed that another 50 people will be trained at each of the same seven training sites.

Computer-based multimedia training program:

* Time of person at each training site to schedule trainees and handle travel reimbursements ($12.50/hour for 10 hours/week for 4 weeks times 7 sites)........................ $3,500.00

Fringe for personnel time @ 30% (starred items).................................... $1,050.00

Total.. $4,550.00
Trainees' Time $171,885.00
Total + Trainee's Time.. $176,435.00

Lecture training program

The cost for a second increment of training would be the same as the first increment.

Total.. $21,693.00
Trainees' Time $171,885.00
Total + Trainees' Time.. $193,578.00

Number of Trainees		Cost of Computer-based Multimedia	Cost of Lecture
350	(1 circuit)	$214,004.00	$193,578.00
700	(2 circuits)	$390,439.00	$387,156.00
1,400	(4 circuits)	$743,309.00	$774,312.00
2,100	(6 circuits)	$1,096,179.00	$1,161,468.00
3,500	(10 circuits)	$1,801,919.00	$1,935,780.00
7,000	(20 circuits)	$3,566,269.00	$3,871,560.00

Appendix H:

Expert Post-experiment Questionnaire.

Post-Experiment Questionnaire

General Questions

1. When you decided which training program was the most cost-effective, did you base your decision on 1) the costs, 2) the predicted benefits, 3) the costs and the benefits? How did you weight the importance of the costs when determining the most cost-effective training program? How did you weight the importance of the benefits when determining the most cost-effective training program? If you did not have the dollar values, could you have used the experts' data to make a decision of cost-effectiveness with confidence?

The **identification** of variables.

There were 14 variables identified that you later classified and quantified. They were:

financial resources	number of remediation sessions necessary
human resources	developer's time to revise the training program
equipment and tools	trainee's motivation
time resources	trainee's self-efficacy
number of trainees that can be trained	trainee's convenience to attend training
num. of locations at which trainees can be taught	trainee's ability to rate the level of drowsiness
trainee's time to complete the training session	trainee's ability to recall information

2. Considering that determining cost-effectiveness was the goal, what other variables do you think should have been included in the list?

3. Considering that determining cost-effectiveness was the goal, what variables do you think were least appropriate to represent the two training programs? Why?

The **classification** of variables.
The task of classifying the variables included deciding whether each variable impacted each other variable.

4. How valid or relevant did you consider this task? How would you do the task differently to make it more valid or relevant?

5. How do you think the task of classifying variables could have been made more clear?

The **quantification** of variables.
The task of quantifying the variables included deciding how related or "classified" variables positively or negatively impacted each other.

6. How valid or relevant did you consider this task? How would you do the task differently to make it more valid or relevant?

7. How do you think the task of quantifying variables could have been made more clear?

8. What is your overall impression with this method of determining the most cost-effective training program? Would you use this method yourself? Are there any other comments you would like to make regarding this method?

Appendix I:

Trainees Informed Consent Form.

Trainee Informed Consent Form

You are being asked to volunteer to be a subject in a research project. The purpose of the research is to determine the most cost-effective method for teaching individuals how to determine the level of drowsiness of colleagues during safety critical crew operations. Your part in this research, if you choose to participate, will be to participate in the training program as it would be administered to its intended users, flight crew personnel and air traffic controllers.

You are being asked to perform the following tasks:

1. Complete a pre-test questionnaire to collect demographic information about you.
2. Participate in a training tutorial on the topic of recognizing and rating the drowsiness level of individuals.
3. Watch video segments in which you will rate the level of drowsiness of the person in the video.
4. Fill out a post-test questionnaire to answer questions about what you learned and your opinion of the training program.

There are no known risks or discomforts associated with participation in this experiment. The benefits to you as a participant in this study are:

1. The opportunity to contribute to education in the area of training program development and issues of cost-effectiveness of training media.
2. The opportunity to learn how to identify the level of drowsiness of individuals on a drowsiness continuum.

You should understand that no promise or guarantee of benefits has been made to encourage you to participate in this study.

Your anonymity will be maintained. A number will be used to replace your name in order to assure that your name will not be associated with your data.

The experimental session will last approximately three hours. You will be paid $6.00 per hour.

The principle investigator, Dr. Walter W. Wierwille (231-7952), and the researcher Vicki L. Neale (231-9084), will answer any questions you may have about your participation. You should not sign this informed consent form until you are satisfied that you understand all of the previous descriptions and conditions. You should further be aware that you may contact Dr. Ernest Stout, Chairman of the University's Institutional Review Board, if you have any questions or concerns about this experiment. His phone numbers are 231-9359 and 231-6077.

You should know that at any time you are free to withdraw from participation in this research program for any reason without penalty.

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University, by the Department of Industrial and Systems Engineering for the Human Factors Option.

I have read and understand this Informed Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

_______________________________________ Date:___________________
Participant's Signature:

_______________________________________ Date:___________________
Experimenter's Signature:

Appendix J:

Trainee Introduction to the Study.

Introduction to the Study for the Drowsiness Detection Training Program

The purpose of the research is to determine the most cost-effective method for teaching individuals how to determine the level of drowsiness of colleagues during safety critical crew operations. Your task is to act as a trainee for this training program as though you were a flight crew member or air traffic controller.

You are being asked to perform the following tasks:

1. Complete a pre-test questionnaire to collect demographic information about you.
2. Participate in a training tutorial on the topic of recognizing and rating the drowsiness level of individuals.
3. Watch video segments in which you will rate the level of drowsiness of the person in the video.
4. Fill out a post-test questionnaire to answer questions about what you learned and your opinion of the training program.

The experimental session will take approximately three hours. The possible risks and discomforts to you are negligible.

The principle investigator, Dr. Walter W. Wierwille (231-7952), and the researcher Vicki L. Neale (231-9084), will answer any questions you may have about your participation. You should not sign this informed consent form until you are satisfied that you understand all of the previous descriptions and conditions. You should further be aware that you may contact Dr. Ernest Stout, Chairman of the University's Institutional Review Board, if you have any questions or concerns about this experiment. Her phone numbers are 231-9359 and 231-6077.

You will be paid $6.00 per hour for you participation.

You should know that at any time you are free to withdraw from participation in this research program without penalty for any reason.

Appendix K:

Trainee Pre-test Questionnaire.

Trainee Pre-test Questionnaire

Please respond to the following questions:

1. Gender: Male Female

2. Age: __________

3. What year are you in college? Fresh Soph Junior Senior Grad

4. List the computer systems that you use regularly and how many times per week you use them.

____

____

____

____

4. Have you ever participated in any research that has been conducted in the Vehicle Analysis and Simulation Lab at Virginia Tech? No Yes If yes, please elaborate.

____

____

5. Have you ever completed ground school? No Yes If yes, please elaborate.

____

____

6. Do you have any experience as a pilot? No Yes If yes, please elaborate.

____

____

7. Have you ever worked for an airline? No Yes If yes, please elaborate.

____

____

8. Have you ever participated in any sleep deprivation or drowsiness detection research studies? No Yes If yes, please elaborate.

____

____

____

Appendix L:

Post-test Questionnaire.

Post-Test Questionnaire

1. Fill in the blank. Fatigue and drowsiness contribute to as many as ____ % of accidents in aviation and air traffic control.

2. What are the five general levels of drowsiness?

   1.

   2.

   3.

   4.

   5.

3. What are the five task components to help a colleague who is drowsy?

   1.

   2.

   3.

   4.

   5.

4. Why is monitoring the level of drowsiness of colleagues important?

5. How do you monitor colleagues?

6. How often should you monitor colleagues?

7. How should you rate an individual who is fighting drowsiness?

8. When an individual is varying in and out of drowsiness, what tends to be the trend in that person's drowsiness level?

9. Mark the range that is the criterion level for intervention on the scale below.

Not Drowsy	Slightly Drowsy	Moderately Drowsy	Very Drowsy	Extremely Drowsy

10. What is considered appropriate intervention depends on what two factors?

    1.

    2.

11. Name four methods to help someone who is drowsy.

    1.

    2.

    3.

    4.

12. If you recognize that your colleague is successfully fighting drowsiness, what should you do?

13. Please make your rating on the scale below. How motivated were you to learn the drowsiness detection material. Stated another way, to what level did you to want learn about drowsiness detection training?

Not Motivating	Slightly Motivating	Moderately Motivating	Very Motivating	Extremely Motivating

14. Please make your rating on the scale below. How able do you think you are at rating the level of drowsiness of someone in a working situation?

Not Able	Slightly Able	Moderately Able	Very Able	Extremely Able

15. The training program that you participated in is intended for flight crew personnel and air traffic controllers. In the future, they may be able to participate in this training program in one of two ways. One scenario would be that lecture seminars would be conducted in major U.S. cities and the trainees would go to these seminars at a scheduled day and time and hear an instructor present the material with other classroom participants. A second scenario is that computers would be located in flight schools in major U.S. cities and the trainee could go to these flight schools at a time he or she has scheduled and then independently participate in a computer-based multimedia training program on the subject. Of these two scenarios, which do you think would be more convenient to attend, the lecture training program or the multimedia training program? Check one of the following:

☐ The lecture training program would be more convenient to attend.

☐ The computer-based multimedia training program would be more convenient to attend.

16. What aspects of the training program did you like?

17. What aspects of the training program did you dislike? What aspects of the training program would you change if you could?

Appendix M:

Outline of the Lecture and Computer-based Multimedia Drowsiness Detection Training Programs

Outline of the Lecture and Computer-based Multimedia Training Programs

I. Introduction

A. Title Screen
B. Background Information
  1. Statistics related to accidents
  2. How statistics relate to the trainee
C. The purpose of the training program and overview of participant's role.
D. Notes on some of the video segments that will be used in the training program.
  1. Some of the video being used was taken with a low-light level camera
    a. some areas of the face look more or less bright which is an effect of the infra-red light being picked up from the person's skin
  2. The video images taken are close up images
    a. the image of the person's face may be slightly distorted compared to the person's actual features
  3. The people in the video images will not look exactly like your work colleagues will look.
    a. Focus on the concepts that will be presented in the training program.
    b. Think of your work situation and draw on past experience and knowledge of your colleagues.

II. Presentation of the Task Components

A. There are five general levels of drowsiness
  1. The levels of drowsiness on a continuum

Not Drowsy	**Slightly Drowsy**	**Moderately Drowsy**	**Very Drowsy**	**Extremely Drowsy**

  2. The rating scale shown is a continuous rating scale since people do not fall into one category at one point then instantly ratchet to another level of drowsiness.
B. Monitoring Colleague
  1. Why monitoring colleagues for level of drowsiness is important
  2. How often to monitor colleagues
    a. depends on how fast the colleagues level of drowsiness declines
  3. How to monitor colleagues
    a. glances of a few seconds in order to recognize the features and mannerisms of the colleague.
C. Determine if the colleague exhibits signs of drowsiness
  1. What the "not drowsy" (alert) mannerisms are. (*Use video of two peole who are alert and point out the following*)
    a. normal facial tone
    b. normal fast eye blinks/purposeful eye closures
    c. short ordinary eye glances
    d. exhibits occasional body movements or gestures

2. If the mannerisms are not those of a person who is alert, person may be drowsy.
3. What the general mannerisms are for each level of drowsiness
   a. slightly drowsy (*Use video of two people who are slightly drowsy and point out the following*)
      - facial tone begins to decrease
      - rubbing the face or eyes
      - scratching
      - facial contortions
      - moving restlessly in the seat
   b. moderately drowsy (*Use video of two people who are moderately drowsy and point out the following*)
      - some or all of the signs of slight drowsiness plus...
      - person becoming more subdued
      - slower eye lid closures
      - glassy-eyed appearance
      - staring at a fixed position
   c. Note on the mannerisms of slightly and moderately drowsy: The problem that not all people exhibit mannerisms at the slightly drowsy to moderately drowsy range.
      - difficult stages to assess
      - it is helpful to pull from past experiences and knowledge of the person
   d. very drowsy (*Use video of two people who are very drowsy and point out the following*)
      - some or all of the signs of slight or moderate drowsiness plus...
      - eyelid closures of 2 to 3 seconds or longer occurring
      - eyelid closures accompanied by a rolling upward or a sideways movement of the eyes themselves
      - not focusing the eyes properly
      - exhibit a cross-eyed (lack of proper vergence) look
      - facial tone has probably decreased
      - lack of apparent activity
      - large isolated (or punctuating) movements
      - reorienting the head from a leaning or tilting position
   e. extremely drowsy (*Use video of two people who are extremely drowsy and point out the following*)
      - person exhibits some or all of the signs of slight, moderate, or very drowsy plus...
      - falling asleep
      - exhibiting prolonged eyelid closures (4 sec or more)
      - exhibiting prolonged periods of lack of activity

- exhibiting large punctuated movements as transition in and out of intervals of dozing

4. A person may use methods to successfully fight drowsiness (*Show two video examples of people fighting drowsiness with an accompanying rating scale*).
   a. Point out the persons mannerisms.
5. Side by side video
   a. not drowsy and slightly drowsy
   b. not drowsy and moderately drowsy
   c. not drowsy and very drowsy
   d. not drowsy and extremely drowsy
   e. slightly drowsy and moderately drowsy
   f. moderately drowsy and very drowsy
   g. very drowsy and extremely drowsy
6. One person at each general level of drowsiness.

D. Determine the level of drowsiness within the general level (the specific level of drowsiness)
   1. Describe that once the general level of drowsiness is assigned, it is necessary to make a more accurate assessment since there is a range to each general level of drowsiness (*show rating scale*).
   2. (*Shown as an example*) Determine how well the mannerisms of the person match the description of the general level of drowsiness
      a. Does the person exhibit few, some, or all of the signs of drowsiness for the a particular level of drowsiness?
         - if the person exhibits few signs for the general level of drowsiness, rate the person at the low end of the general level
         - if the person exhibits several signs for the general level of drowsiness, rate the person near the middle range of the general level of drowsiness
         - if the person exhibits all signs for the general level of drowsiness, rate the person at the high end of the general level of drowsiness.
   3. Show where the person's level of drowsiness would fall on the rating scale.
   4. Give 7 more examples.

E. Determine if intervention is necessary
   1. The criterion level for intervention is in the obviously slightly drowsy up to the moderately drowsy range.
      a. *Show area on the rating scale.*

F. Determine appropriate intervention.
   1. What may be considered appropriate intervention depends on the dynamics of the work situation and management.
   2. What is the working situation?
      a. working as a team of two vs. working as a team of three or more
      b. can the drowsy person be spared for some amount of time?
   3. What does the person's level of drowsiness demand?
      a. caffeine
      b. stretch and move about
      c. ten minutes break
      d. 30 minute break to nap
      e. relieved of duty
   4. What to do when the colleague cannot be spared.
      a. caffeine

    b. conversation
    c. stretch and move about
5. What to do if the person is able to successfully fight drowsiness.
    a. closely monitor
    b. inform the person that you realize he/she is fighting drowsiness and ask if the person needs assistance or would like to take a break, etc.

III. Review

A. Why this skill is necessary
B. How to monitor
C. What are the general levels of drowsiness and how does one specify the level of drowsiness.
D. When and how to intervene

IV. Practice

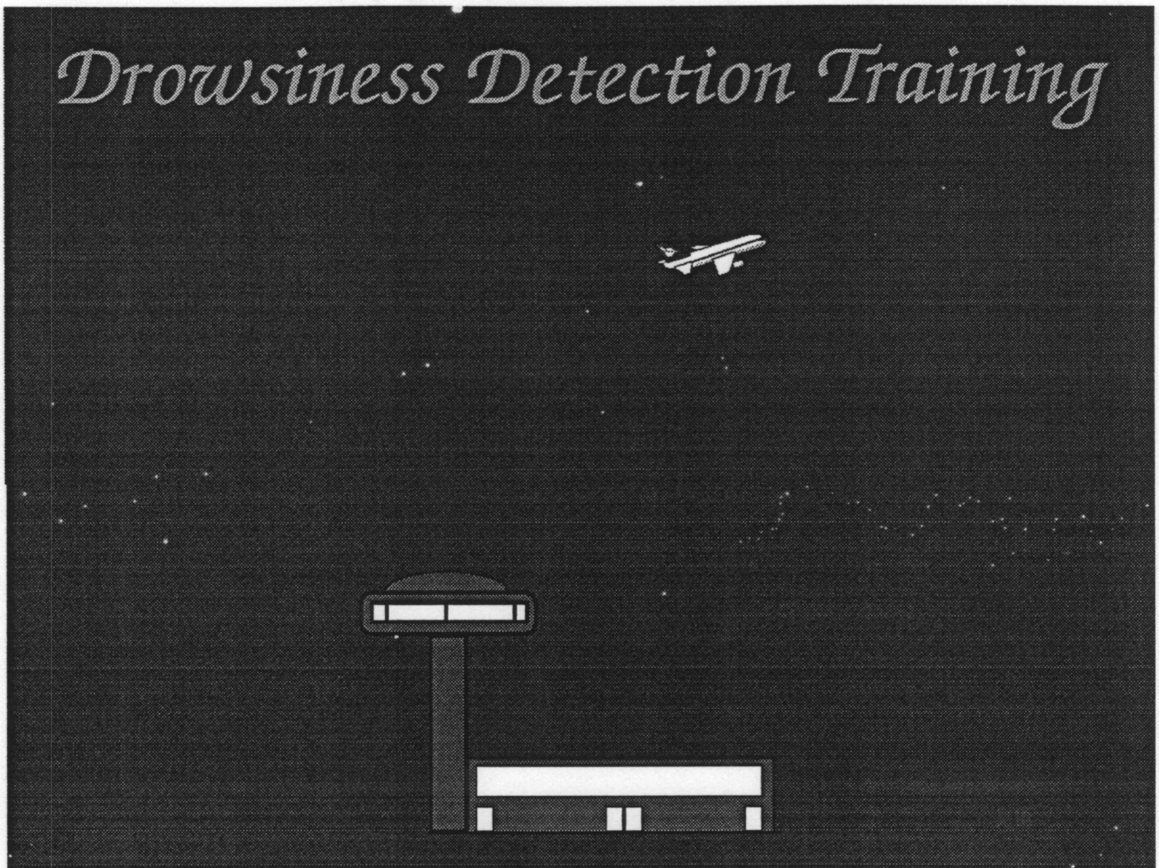
A. Ten video clips for trainee to judge the level of drowsiness on rating scale.
B. Feedback given after each rating regarding the correct response.

V. Credits

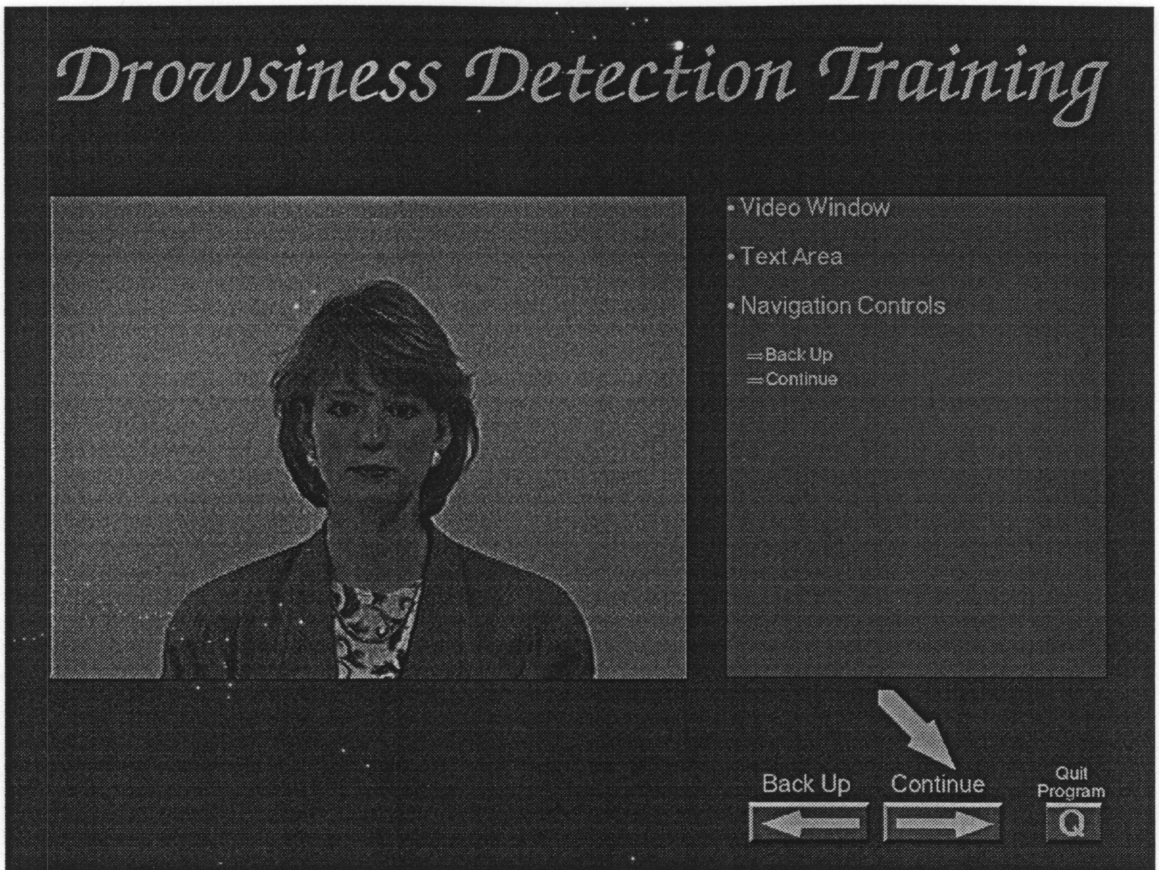
Appendix N:

Screen Captures of the Computer-based Multimedia Drowsiness Detection Training Program

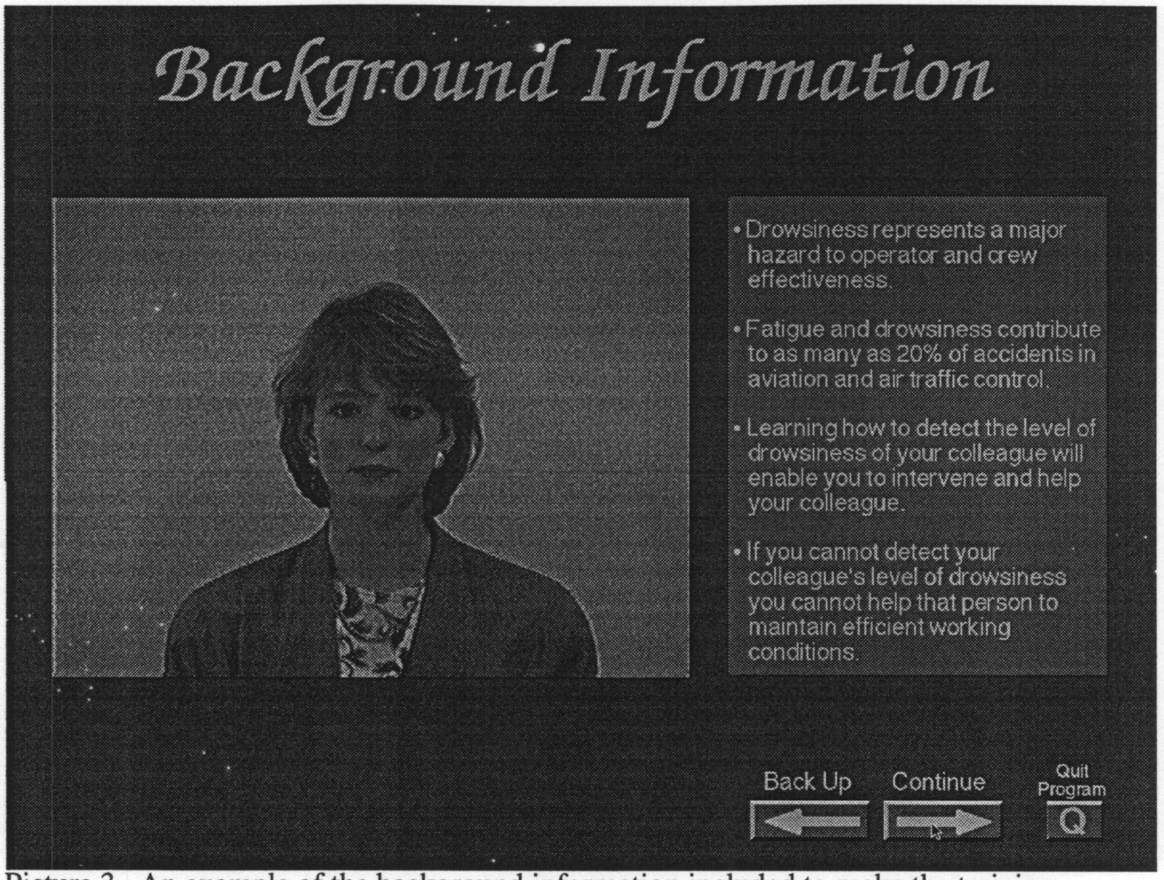
Screen Captures from the Computer-based Multimedia Drowsiness Detection Training Program.



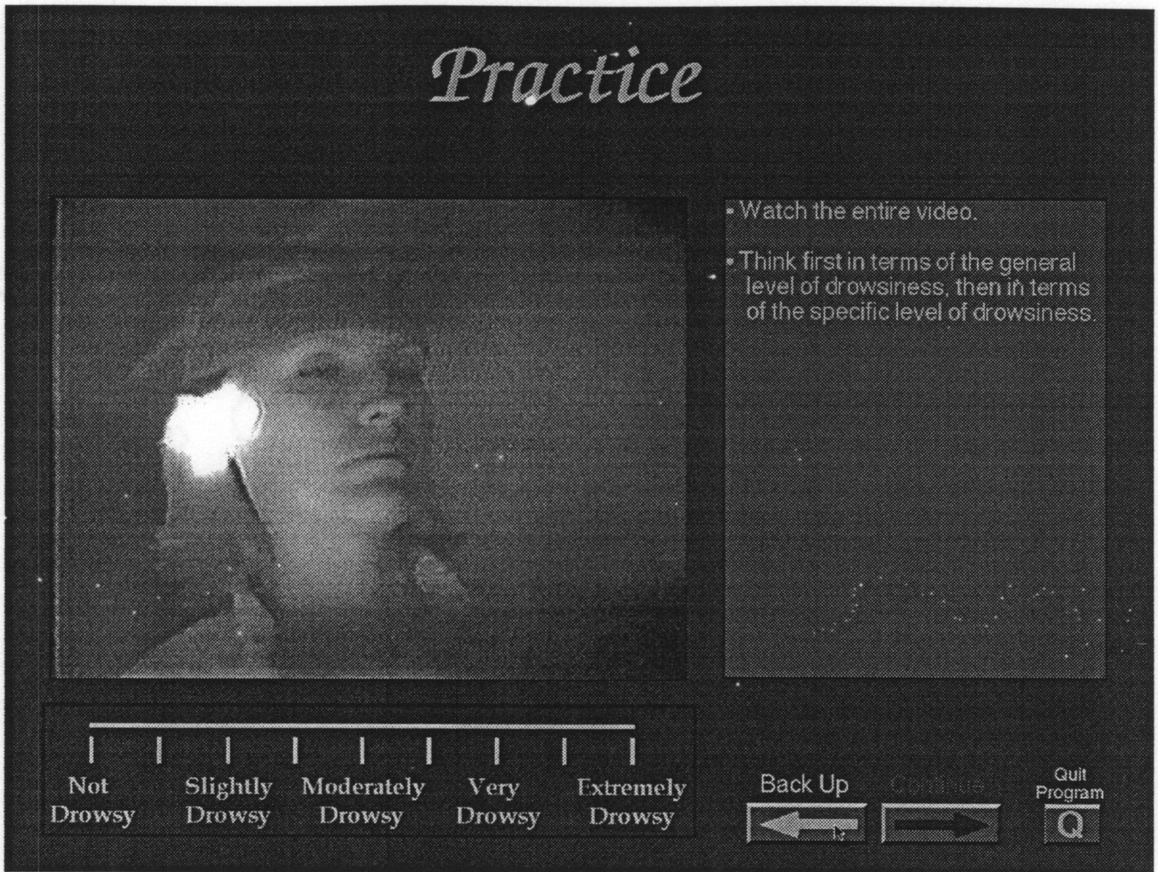
Picture 1. The opening screen of the Drowsiness Detection Training Program. The program opens with music and an animation of an airplane at take off.



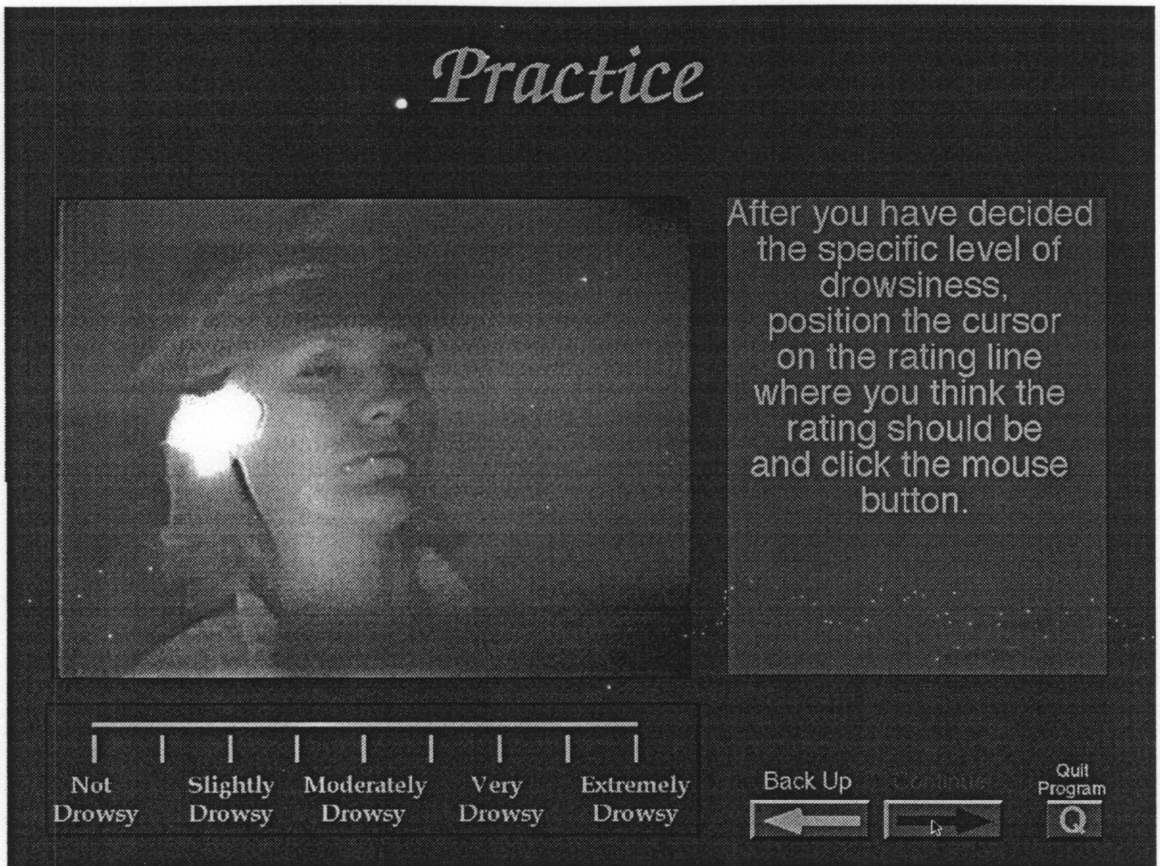
Picture 2. The first narrator screen. The narrator describes the computer interface features to the user.



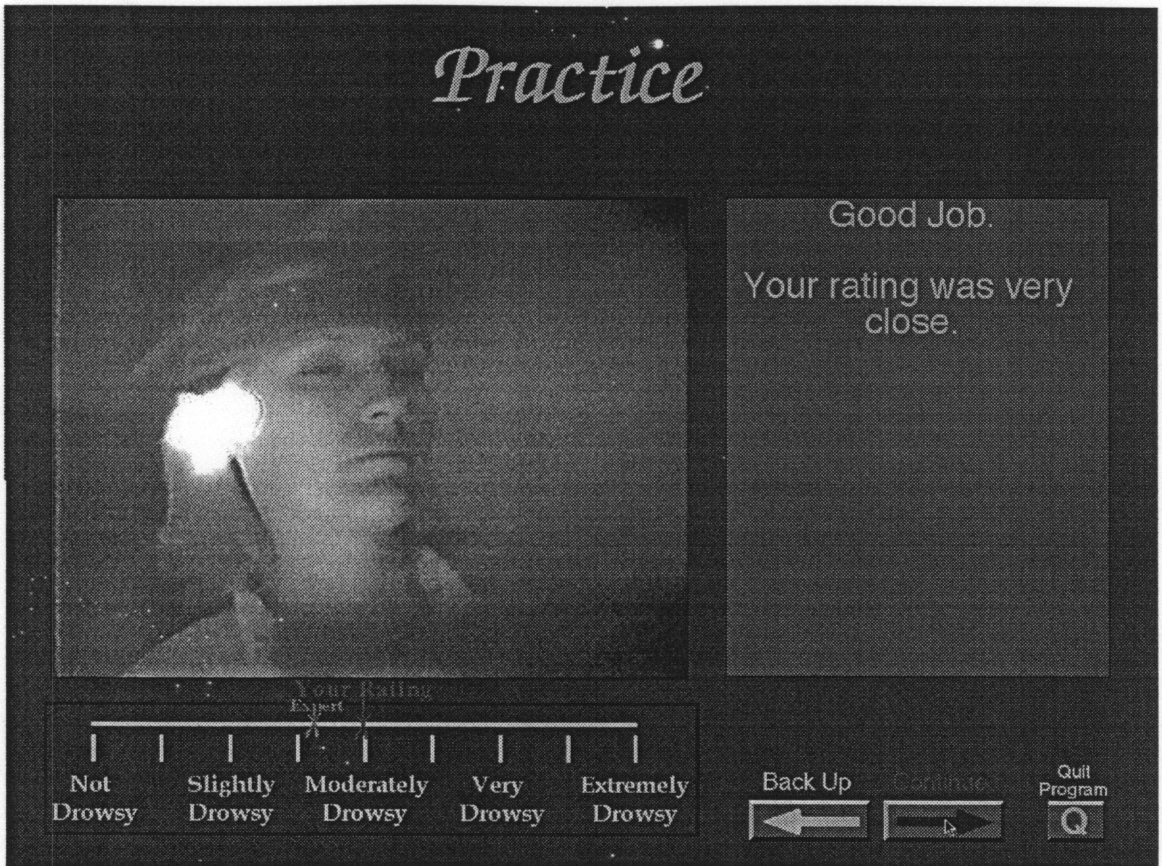
Picture 3. An example of the background information included to make the training program relevant to the trainee.



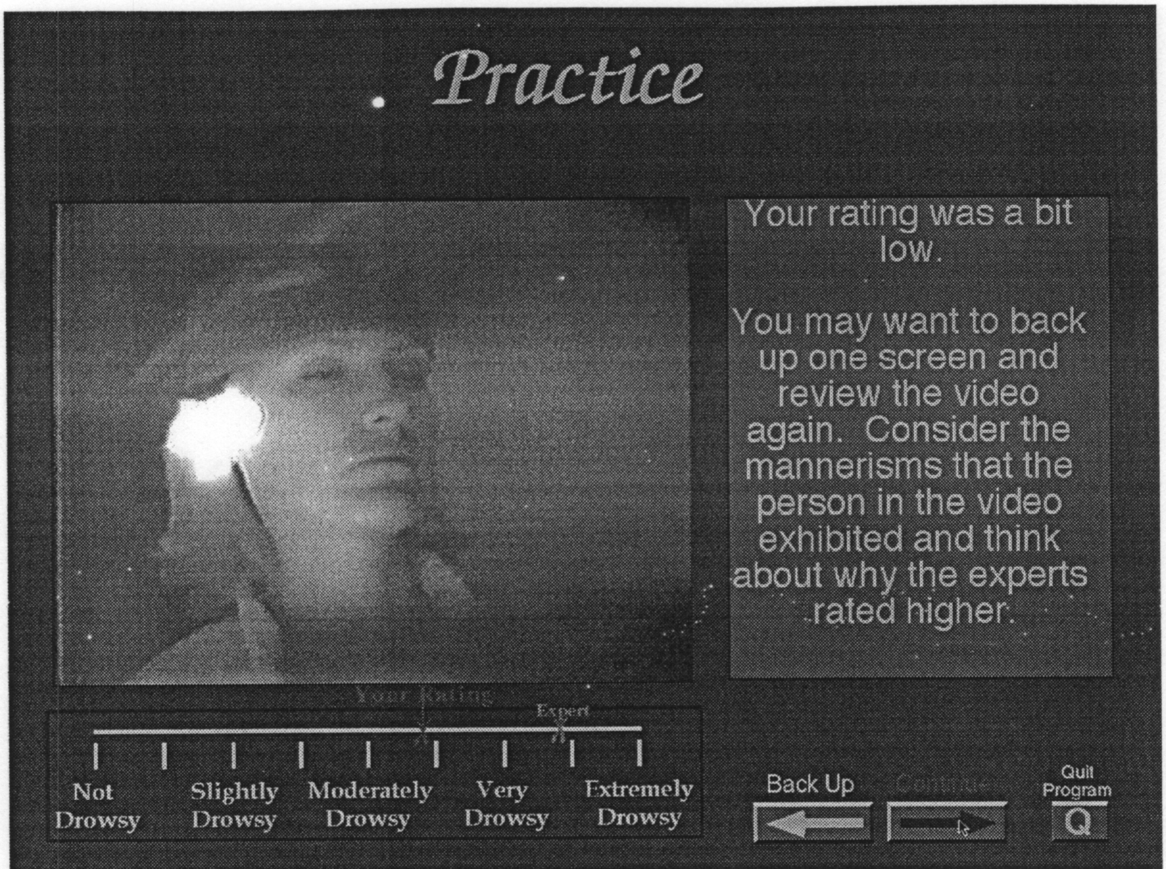
Picture 4. An example practice segment. Each practice segment begins with a video segment for the trainee to watch.



Picture 5. Once the practice video segment stops, the trainee is cued to enter a rating.



Picture 6. Once the trainee enters a rating, the expert rating appears and feedback is given with text and an audio explanation.



Picture 7. An example of the text feedback that appears if the trainee's rating is out of an acceptable range of performance.

Appendix O:

Cost Accounting Worksheet for the Design and Development of the Lecture and Computer-based Multimedia Training Programs.

Cost Accounting Worksheet for the Design and Development of the Lecture and Computer-based Multimedia Training Programs.

	Creation of Lecture and Computer-Based Multimedia (CBM)	**TIME**
	Drowsiness Detection Training Programs	**in hours**
	Selecting and Purchasing Computer Equipment	
	Determining Necessary System Functionality	14
	Finding Equipment in Catalogs/comparison shopping	10
	Confirming list with expert	2
	Putting together list, making revision, getting approval	10
	Completing requisitions and orders	4
M	**Total Hours Selecting and Pruchasing Comp. Equip.**	**40**
	Front-end Analysis	
	Needs Assessment	
	Organizational Analysis	5
	Task Analysis	40
	Trainee Analysis	10
	Training Need and Resource Analysis	5
	Functional Specifications	
	Post-Training Goals and Objectives	10
	System Performance Requirements	5
	System Constraints	5
	Generating Front-end Analysis Document	25
	Presentation to COTR	
	Presentation Preparation	5
	Day trip to Washington DC	10
M, L	**Total Hours on Front-end Analysis**	**120**
	Design and Development Phase	
	General Design of Storyboards (Applies to both lecture and CBM)	
	Outline - 1st Draft	8

	Outline - 2nd Draft	2
	Storyboard 1st Draft	30
	Storyboard 2nd Draft	11
	Storyboard 3rd Draft	
	Rework narrative to match video that was captured	10
	Storyboard 4th Draft	
	Final draft before optimization	4
M, L	**Total Hours Storyboard Design**	**65**
	Drowsy Driver Video Segments	
	The Videotaped Drowsy Driver Segments	
L	Select Video Segments	20
	Digitize Video Segments	20
L	Edit Segments (including Side by Side Segments)	85
M	**Total Hours Drowsy Driver Video Segments**	**125**
	Lecture Training Design and Development	
	Tailor Storyboard to Lecture Training	20
	Optimization Process	
	First Optimization Process	
	Presentation time to four individuals	3
	Preparation for lecture	10
	Revision time on storyboard	11
	Second Optimization Process	
	Presentation time	3
	Revision time	3
	Creation of Videotape	
	Version 1	2
	Version 2	4
	Version 3	9
L	**Total Hours Lecture Design and Development**	**65**
	CBM Design and Development	
	Tailor Storyboard to CBM Training	30
	Digitize Voice Only Narrative	8
	Videotape Narrator	

	Making Cue Slides in PowerPoint	3
	Setup	6
	Taping	4
	Digitizing	8
	Programming	100
	Optimization Process	
	First Optimization Process	
	Presentation Time	3
	Revision Time	25
	Second Optimization Process	
	Presentation Time	3
	Revision Time	5
M	**Total Hours CBM Design and Dev**	**195**
M, L	Generate Design Document (Complete Storyboards)	30
	Total Hours Design and Dev of Lecture and CBM	**480**
	Test and Evaluation Phase	
	a Prepare ICF and other forms	10
	b Filing IRB Forms	2
	c Arrange Payment of Trainees	2
	d Develop and Post Flyers	3.5
	e Schedule Subjects	10
	(Run Multimedia Trainees)	
	f Time with subjects – 16 subjects x 2.5 hours each	40
	(Run Lecture Trainees)	
	g Time with subjects in classroom	2.5
	h Prepare for session including practice and materials	10
	i Reduce Data	15
	j Analyze Data	30
L	**Total Hours T&E for Lecture (a+b+c+d+e+g+h+i+j)**	**31**
M	**Total Hours T&E for CBM (a+b+c+d+e+f+i+j)**	**67.5**
M, L	Write Test and Evaluation Phase Report	20
M, L	Presentation to COTR in Washington DC	15

	Total Hours Test and Evaluation for Lecture and CMB	**160**
	Revision of Training Programs	
M	Revision of Drowsy Driver Video Segments	100
L	Revision of Lecture Training Program	30
M	Revision of CBM Training Program	70
M, L	Generate Revised Storyboard Document	40
	Total Hours Revision of Training Programs	**240**
	Total Hours Devoted to Drowsiness Detection Training Programs	**1040**
	Total Hours Devoted to Lecture (Items marked with L)	**521**
	Total Hours Devoted to CBM (Items marked with M)	**887.5**

VITA OF VICKI LEWIS NEALE

Doctor of Philosophy – Virginia Polytechnic Institute and State University
Industrial and Systems Engineering - Human Factors Option, May 1996
Dissertation Title: A Comparison Between Predictive and Formative Cost-effectiveness Evaluation Techniques for the Assessment of Lecture and Computer-based Multimedia.

Master of Science – University of Idaho
Psychology - Human Factors Option, May 1994
Thesis Title: An Evaluation of Training Technique as a Means of Influencing Risk Perception, Safety Attitudes and Knowledge Retention Among Loggers.

Bachelor of Science – University of Idaho
Psychology - Human Factors Option, May 1990

WORK EXPERIENCE

08/94 - 04/96 **Research Assistant**:
Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering, Vehicle Analysis and Simulation Laboratory.
- Funded by the Federal Aviation Administration and the National Highway Transportation & Safety Administration
- Designed and developed both a lecture and computer-based multimedia drowsiness detection training programs to teach individuals how to discriminate the levels of drowsiness of co-workers.
- Designed and administrated the test and evaluation plan of the lecture and computer-based multimedia drowsiness detection training systems.
- Investigated the issues of cost-effectiveness of training media.
- Investigated the usability of a cost-effectiveness assessment method.

08/93 - 05/94 **Teaching Assistant**:
Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering.
- Assisted students in the basic principles of Work Physiology, including principles of anatomy and ergonomics.
- Assisted students in the basic principles of Human Factors.
- Developed testing materials.
- Conducted workshops.
- Conducted lectures periodically.

01/93 - 06/93 **Human-Computer Interface Designer**:
Virginia Polytechnic Institute and State University, Continuing Education Center.
- Performed needs analysis via surveying businesses and industries throughout Virginia.
- Designed and developed a computer system (Partnership in Progress) for business and industry to use in accessing university services and electronic databases.

05/92 - 10/92 **System Designer/Human-Computer Interface Designer**:
University of Idaho, Department of Civil Engineering.
- Designed a "laptop" computer system, including the interface and keyboard layout, for the Traffic Data Input Program (TDIP).
- Designed the TDIP human-computer interface and the traffic pattern data collection method.

08/91 - 10/92 **Research Assistant**:
University of Idaho, Departments of Psychology, and Wildlife and Forestry.
- Funded through the Occupational Safety and Health Administration.
- Performed field observation and interviews of timber harvesters.
- Designed and developed an interactive computer-based multimedia safety training program for timber harvesting landing operations.

08/91 - 12/91 **Human-Computer Interface Designer**:
University of Idaho, Department of Psychology.
- Designed and developed a brochure and on-line training program to educate faculty in the uses of multimedia as a teaching aid.

06/91 - 08/91 **Technical Assistant**:
Boeing Commercial Airplane Group, Flight Deck Research.
- Created a "living" guideline document of human-computer interaction principles.
- Developed a test and evaluation plan to investigate data link communications between flight crews and air traffic controllers.

01/91 - 05/91 **Research Assistant**:
University of Idaho, Department of Psychology.
- Developed usability analysis simulator test plan for Hughes Aircraft to evaluate the attentional demand of TRAVTEK navigation and information system.

08/90 - 12/90 **Teaching Assistant**:
University of Idaho, Department of Psychology.
- Aided students in acquiring basic concepts of Introduction to Psychology.
- Conducted class lectures periodically.

01/90 - 04/91 **Co-principle Investigator**:
University of Idaho, Department of Psychology.
- Designed and evaluated Hypertext instructional systems using AI knowledge engineering.

08/89 - 12/89 **Research Assistant**:
University of Idaho, Department of Psychology.
- Funding provided by General Motors.
- Performed human factors test and evaluation of the Highway Driver's Assistant (HDA) and the TRAVTEK Navigation and Information System, including tests of the visual attentional demand of the dynamic screens.

PUBLICATIONS AND TECHNICAL REPORTS

Dingus, T. A. and Lewis, V. R. (1991, Sept.). *Data link communications: A test and evaluation plan of presentation method.* Submitted to Boeing Commercial Airplane Group.

Gill, R. and Lewis, V. (1992). *Towards improved college teaching: A preliminary report.* White Paper to the University of Idaho, Moscow, ID.

Gordon, S. E. and Lewis, V. (1990). Knowledge engineering for hypertext instructional systems. *Proceedings of the 34th Annual Meeting of the Human Factors Society* (pp. 1412 - 1416). Santa Monica, CA: Human Factors Society.

Gordon, S. E. and Lewis, V. (1992). Enhancing hypertext documents to support learning from text. *Journal of the Society for Technical Communication, 6*(2), (p. 305-308).

Neale, D. C., Lewis, V. R., and Petitt, C. (1994). A systems approach for incorporating human factors in the development of a computer-based multimedia vacation planning system. *Proceedings of the 2nd Annual Mid-Atlantic Conference of the Human Factors and Ergonomics Society* (pp. 70-74). Santa Monica, CA: Human Factors and Ergonomics Society.

Neale, V. L. and Dingus, T. A. (1996). A comparison of landing site safety training techniques for loggers. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society.* Submitted for Review.

Neale, V. L. and Wierwille, W. W. (1995, March). *Research on vehicle-based driver status/performance monitoring: Observer training assessment for a drowsiness detection training program: Knowledge, skills, and training objectives.* (ISE Tech. Report No. 95-01). Blacksburg, VA: Virginia Polytechnic Institute and State University, Vehicle Analysis and Simulation Laboratory.

Neale, V. L. and Wierwille, W. W. (1995, June). *Research on vehicle-based driver status/performance monitoring: Training materials for a drowsiness detection training program in lecture and computer-based multimedia formats.* (ISE Tech. Report No. 95-05). Blacksburg, VA: Virginia Polytechnic Institute and State University, Vehicle Analysis and Simulation Laboratory.

Neale, V. L. and Wierwille, W. W. (1995, December). *Research on vehicle-based driver status/performance monitoring: Part II. Empirical evaluation of lecture and computer-based multimedia format drowsiness detection training programs.* (ISE Tech. Report No. 95-24). Blacksburg, VA: Virginia Polytechnic Institute and State University, Vehicle Analysis and Simulation Laboratory.

PROFESSIONAL ORGANIZATIONS AND HONORS

Workshops Co-Chair: Human Factors and Ergonomics Society (HFES) 39th Annual Meeting, San Diego, CA

Co-Coordinator: HFES 3rd Annual Mid-Atlantic Conference, Virginia Polytechnic Institute and State University

Manager: HFES Student Council Mail List (hfessc@discus.ise.vt.edu)

President: HFES, Virginia Polytechnic Institute and State University Student Chapter - 1994

Student Member: HFES

Treasurer: HFES, University of Idaho Student Chapter, 1990 - 1992

Life Member: Psi Chi - National Honor Society in Psychology.

Dean's List: University of Idaho: All years in attendance.