The Design, Development, and Evaluation of a Prototype Training Course Life Cycle Cost Estimating Tool

by

Gary L. Macomber

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 Engineering and Operations Research

APPROVED:

Robert C. Williges, Chairman

Harold A. Kurstedt, Jr.

Dennis L. Price

J. William Schmidt, Jr.

James M. Young

May, 1988

Blacksburg, Virginia

Design, Development, and Evaluation of a Prototype Training Course Cost Estimating Tool

by

Gary L. Macomber

Committee Chairman: Robert C. Williges
Industrial Engineering and Operations Research

(ABSTRACT)

A prototype cost estimating tool was built using an iterative design approach to help training managers assess the impact of changes in the training environment on course life cycle costs. In the first of two evaluations, training managers evaluated the prototype tools during a demonstration for both functionality and face validity in their environment. They also evaluated the usefulness of different levels of graphics and detail in a guide designed to assist the user with the tool. Feedback from the first evaluation was used to revise the prototype tool, and the guide usefulness ratings were used to select the guide for use in the second evaluation. The impact of the guide on performance with the tool was analyzed in a second evaluation. The second evaluation analyzed the difference in performance of two groups. The first group only had the users manual to help them solve the problems. The second group had the users manual and a performance aid to help them. Subjects' performance was compared on the number of spreadsheets used, number of errors made, and the amount of time they took to complete the tasks. The two groups did not significantly differ on these measures.

ACKNOWLEDGEMENTS

The USAF Institute of Technology, Civilian Schools Program supported this research. The cooperation of the USAF Air Training Command Tecnical Training Center personnel at Goodfellow, Lowry, Keesler, and Lackland Air Force Bases was critical to completing this effort.

Several individuals deserve recognition for their contributions and assistance. Dr. Robert Williges provided invaluable guidance, encouragement, and insight. Dr. James Young set up the path for data collection and always found time in his schedule to help. Dr. Dennis Price was always willing to ensure this research covered all the areas necessary. Dr. Harold Kurstedt and Dr. William Schmidt provided the extra insight and background this research needed. deserves special recognition for her efforts. She stepped in and provided much needed assistance and guidance. Without the contributions of all these individuals this research could not have been done.

Finally, I wish to thank my wife . Her support, encouragement, and confidence in me sustained me throughout this program.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

The Introduction chapter consists of four sections. The first section defines the problem and the operational setting forming the basis for this study. The second section describes the fundamental research question this study addresses. The third section discusses the assumptions of the study. The fourth section discusses the constraints on this study.

OPERATIONAL SETTING AND PROBLEM

The setting for this dissertation was the USAF Air Training Command Technical Training Centers. Specifically, the training managers were used as subjects and aspects of their job formed the basis for the decision support tool. These individuals oversee several different training courses. Each training manager might oversee five to fifteen or more courses, depending on the training center and the number of trainees each course trained per year. The training managers are responsible for forecasting future requirements and assessing the impact of policy decisions. They are more concerned with planning and programming for the future than they are with the day-to-day operations of their courses.

The problem this dissertation addresses is lack of a methodology to help training managers compare course alternatives based on projected course life cycle cost. The training managers know the trainee load and the course material to be taught. The alternatives deal with the resources required to teach the course (instructional media, personnel, and time, for example). These comparisons occur primarily at four different times: budget preparation, budget defense, course initiation, and course revision.

The first two times training managers compare training system costs are during the preparation and defense of the annual budget. The training manager must show the effect of changing the resources available to a course. Examples of this are the impact on the total number trained when manpower is reduced, the impact of shortening (or lengthening) a course, and the impact of changing the student load. These changes affect the resource requirements for a course and the total cost of training.

The second two times training managers compare training system costs is during course initiation and revision. The current USAF training environment requires training managers to consider alternative instructional delivery modes, such as computer-assisted instruction and interactive videodisc. In some cases, the alternative modes may be more cost effective than lecture based instruction.

Traditionally, training managers consider only initial course costs. Those are equipment and facilities. The true cost of training is much more than that. Someone must develop and maintain the courseware. Classrooms must be maintained. Also, instructors and trainees cost money. Not only do the trainees and instructors draw salaries and benefits directly, but they require support facilities and personnel. The support personnel include both those at a particular training center and at command levels above the training center. The command levels referred to are those impacting the training process in a particular course. Adding these factors gives the training manager a different perspective on the cost of training.

Orlansky (1985) discussed the cost effectiveness of flight simulators, computer based instruction (CBI), and maintenance simulators. He found training managers should amortize flight simulator acquisition costs in two years and maintenance simulators acquisition costs in four years. Due to a lack of cost data, he did not report an amortization period for CBI. The lack of cost data was due in part to the wide range in quality, quantity, and the small amount of CBI usage. Orlansky (1985) based the comparisons of trainee performance on aircraft and maintenance simulators with actual equipment on the premise the trainees were equally qualified after training. This means he assumed there were no differences in task performance between trainees trained on simulators, actual aircraft training, conventional instruction, or using actual maintenance equipment. Orlansky (1985) based his comparisons on mixed media training systems. For example, in the flying simulator comparisons, a portion of the flying hours were replaced with simulator hours. In the maintenance simulator comparisons, a portion of the actual maintenance equipment use time was replaced with maintenance simulator time.

Advanced Technology (1985) reviewed the current and projected Air Force training environment. The report confirmed, for the Air Force, the DoD-wide results of the Defense Science Board Summer 1983 Study. They confirmed that the Air Force did not employ training technology efficiently. They found the training managers had no tools or guidelines to help them compare the effects of using different training technologies. This void, they proposed, made it difficult for the training managers to defend the use of advanced training technologies. To fill this need they called for the development of a set of general training decision aids, including cost models.

Orlansky (1985) and Advanced Technology (1985) stated the same conclusion. The training community needed two things to help make decisions. First, a method to compile and relate training costs to their courses. Second, a decision support tool to help compare the costs of using different training technologies in a particular course. The prototype Course Cost Estimating Tool (CCET) developed as part of this dissertation was aimed at this requirement.

FUNDAMENTAL RESEARCH QUESTION

The fundamental research question was:

> Can the usefulness of a course cost estimating tool and the effect of performance aids on that usefulness be determined within the USAF training environment?

This study used selected human factors laboratory methods in the USAF Air Training Command environment to answer the research question. The study employed several different evaluation methodologies in two formative evaluations of two different versions of a prototype decision support tool. The two formative evaluations were done within an iterative design framework where the results of the first evaluation were used to modify the Course Cost Estimation Tool (CCET). Given a course cost estimating tool did not exist, one was designed and developed before the research question was addressed. To assist with using the tool, performance aids were developed.

ASSUMPTIONS

There were three groups of assumptions in this study. The first group consisted of general assumptions. The second group consisted of assumptions about the training courses. The third group consisted of assumptions about the subjective rating scales used in this study.

The first group consisted of three general assumptions. The first was a deterministic model could adequately model a training course and its costs. The second was the training managers would have access to the necessary information. Third, the sample of training managers used as subjects was representative of all USAF training managers.

The second group consisted of three assumptions about the training courses. The model assumed trainees are tested through end-of-block of instruction tests. Also, the model assumed recycled trainees restarted the block of instruction they failed (not the last lesson taken, only the material failed, or restarting the entire course). The model assumed the training system scheduled classes based on maintaining equal class sizes and a stable number of trainees going through the course across the year.

The third group of assumptions consisted of assumptions about the subjective rating scales used. For the subjective rating scales, the end points were assumed to be polar pairs on the dimension under consideration. In the first evaluation the dimensions were understanding the tool, accuracy in finding information, and effect on the time to solve course cost estimation problems. In the second evaluation, all the ratings in the spreadsheet rating task dealt with the usefulness of the spreadsheet in the rater's work environment.

CONSTRAINTS

The training center environment placed some severe constraints on this study. It constrained the number of people available for experimental sessions. Also, the sessions could not interfere with the working of the training center. This meant the subjects could participate only if they still met their current job demands. It also limited the length of time each person was available to test the prototypes. This second constraint basically limited each session to an hour or less. Both constraints restricted the amount and variety of data gathered.

The locations where data were collected placed additional constraints on the study. Four of the six USAF Air Training Command Technical Training Centers were visited during this effort. The travel and per diem costs associated with going to these sites restricted the length of each visit and number of visits possible.

REVIEW OF THE LITERATURE

There are several different areas involved in developing and testing a management tool such as the CCET. The areas can be broken done into decision support tools, training and training system models, iterative design methods, human computer interface design methods, experimental design models, and specific evaluation methods. The decision support tool literature provides background on the type and format of information needed for making various types of decisions. The training and training model literature defines the environment, the critical issues, and discusses earlier modeling efforts. The iterative design method literature describes methods to incorporate user information into the design process. The human computer interface design methods literature proposes ways to develop and test the CCET. The experimental design model literature provides information about the different evaluation structure alternatives. The last area is a look at the use of bipolar adjectives and the Questionnaire of User Interface Satisfaction (QUIS). This reviews the literature on two of the evaluation methods used in this study.

The Review of the Literature chapter consists of six sections. The first one discusses decision support systems. The second section reviews training and relates it to manufacturing systems. The third section describes training models. The fourth section describes the iterative design process. The fifth section reviews human computer interface design principles. The last section describes the types of experimental designs used in this study.

DECISION SUPPORT SYSTEMS

This section consists of three parts. The first part discusses the decision maker's environment. The second part discusses types of decisions and decision frameworks. The third part discusses decision support tool development and evaluation.

DECISION MAKERS' ENVIRONMENT

Kurstedt (1986) proposed a model of the decision maker's environment consisting of three parts: what is used to manage, who manages, and what is managed. He proposes that the manager uses the "what is used to manage" part to maintain balance in the system. The "what is used to manage" part includes all of the tools, aids, procedures, plans, and policies the manager uses to manage the domain of responsibility. The entire system is the manager's domain of responsibility. All the tools a manager uses in making decisions form a decision support system. A decision support tool focuses on a specific decision and data to help the manager maintain the balance between the three parts.

The "who manages" part is the person responsible for the specific domain under consideration. Kurstedt (1986) asserts everyone manages something, so, everyone has a domain of responsibility and a management system. A student manages assignments, classes, and participation in other activities; in short, a student manages time usage. The student also manages money and supplies. The student may use a date book, some time estimation heuristics, experience with similar activities, and a priority system to manage time. The tools used to manage time do not necessarily work in managing money and supplies. A corporate officer also manages time, money, and supplies. However, the corporate officer manages them on a much larger scale; and the effects of making a wrong decision may have greater effect on society.

Applying Kurstedt's (1986) Management System Model to this dissertation defines the context within which it operates. The domain of responsibility is course-level technical training. Course-level refers to the management level directly responsible for teaching. The "who manages" is a training manager or similar person charged with estimating the effects of changing course configurations on the cost of training. The "what is managed" is a technical training course or a set of courses. A course provides initial or upgrade skill training within a particular job (for example, plumbing or heavy equipment operator). The "what is used to manage" are all the existing reports and background information on the course configuration and media cost. The prototype decision support tool developed as part of this dissertation was designed to help the training manager compare estimated training course costs across different course configurations.

DECISION TYPES

Singhal (1986) presents an in-depth review of several different frameworks for describing decision types. This study reviewed four of those he discussed here: Simon (1960), Anthony (1965), Gorry and Morton (1975), and Kurstedt (1986). These were chosen to trace the evolution of the framework used here. The Kurstedt (1986) framework was chosen based on its completeness and ability to describe a variety of different decision types and integrate the role of decision aids with the decision types. Other models considered included: Simon (1960), Forrester (1961), Blumenthal (1969), Macintosh (1981), Peterson (1977), and Mintzberg (1980). These are the same models considered by Singhal (1986) in analyzing a more complete pool of decision frameworks.

Simon (1960) proposed there are two types of decisions: programmed and non-programmed. Programmed decisions are routine, repetitive decisions. An example would be selecting the next part in assembling a model car. The builder knows the part's size and shape and the directions spell out the assembly procedure. Non-programmed decisions are novel, one of a kind decisions, usually without complete information. For example, designing the advanced tactical fighter aircraft, at best the specifications are qualitative or provide a quantitative performance range. Few specific, detailed procedures are available. This framework tells us a decision support tool for programmed decisions may not work for unprogrammed decisions.

Anthony (1965) broke decisions down by management level. He proposed three levels: operational control, management control, and strategic planning. Operational control refers to the level of management activity characterized by specific rules and repetitive decisions with little uncertainty about the accuracy of the information used. Management control refers to the level of management activity characterized by guidelines and the level of uncertainty is higher. The strategic planning level refers to the long range planning aspect of management activity. No firm rules or guidelines generally exist or several different sets may have application. The level of uncertainty about the information is high due to the lack of structure (boundary constraints) placed on the decisions. These levels refer to the level of activity, not necessarily to the level of the information used to decide or the level of information in the reports generated by a management information system.

Gorry and Morton (1971) combined the basic frameworks of Simon (1960) and Anthony (1965) to produce a matrix of management level by decision structuredness. Also, they changed Simon's term "programmed" to "structured" and added a semi-structured category. To Anthony's framework, they changed "management" to "tactical." The matrix design implies the two factors (Simon's and Anthony's frameworks) are orthogonal. This may not be true. Also, Anthony's activity levels could be confused with organization levels. This would be a mistake. When the foreman on the shop floor prepares a five year (or less) forecast for a work area, the foreman is doing strategic level planning. Similarly, the chief executive assigning work to the support staff is doing operational level management.

Kurstedt (1986) proposes a decision framework based on management pursuits. The pursuits differ in the amount of uncertainty or lack of definition associated with the task. The framework consists of five management pursuits: process, program, project, problem, and perplexity. The pursuits range from having well defined beginning and ending conditions (process) to having unclear or undefined beginning and ending conditions. This framework expands Simon's two categories and defines each level by the amount of information available about the starting and ending conditions.

Kurstedt (1986) also proposes a framework for management endeavors consisting of four levels, from strategic to clerical. Anthony's (1965) framework forms the basis of this framework. By adding the clerical level, Kurstedt's (1986) framework can account for all of a manager's activities. Unlike the other levels, clerical is a "doing" instead of a "managing" endeavor. The three managing endeavors (strategic, tactical, and operational) differ in the scope and precision of the information and decisions. As in Anthony (1965), managers at each level in the organization perform more than one type of endeavor.

Using the Kurstedt (1986) framework for endeavors, the decision support tool was designed to aid the training managers with tactical-level endeavors. Examples of tactical-level endeavors are choosing a new training system from several alternatives or assessing the projected effects of environmental changes on the present one. It was not designed to help the training manager with the day-to-day working of the course (operational-level endeavors). Also, it requires the training manager provide more detailed information than the strategic level endeavors have available.

For the pursuit level, the training systems have been defined and the cost factors selected. The instructional quality comparison of course alternatives is beyond the scope of this decision support tool. The means of comparison is estimated costs; so, a quantitative description of the end (course alternatives) is known. The starting quantitative specifications of the tool are also known. This reduces the five possible levels of pursuits to two, project and process. The decision support tool itself is a process having known start and end points with specified intermediate steps. Kurstedt proposes managing pursuits by driving them to the next simpler level. In this case, the decision support tool drove a project level pursuit to the process level. Once a training system is fielded and validated it becomes a process-level pursuit. In the planning stages of a training course, a critical task is to compare different training systems. The training system is the process that changes an untrained into a trained person.

DECISION SUPPORT TOOL DEVELOPMENT AND EVALUATION

Kurstedt (1985), and Murdick and Munson (1986) present two different general methods for developing decision support tools. Kurstedt (1985) emphasizes evolutionary design and life-cycle management. Murdick and Munson (1986) emphasizes one time development followed by long term maintenance. Although there are many common steps between both development methods, the underlying philosophy is different. Murdick and Munson (1986) champion the linear programming method. Kurstedt (1985) champions the iterative design and implementation method.

Murdick and Munson (1986) break the management information system (MIS) development process down into four steps. The first step involves planning for the MIS and its inclusion into the existing management systems. The second step is the conceptual development of the MIS. This involves analyzing the current system for good and bad points, identifying waste and deficiencies, and comparing various means of making the necessary improvements. The third step, detailed design, involves acquiring, building, and documenting the new system. The fourth step, implementation, involves installing the system, training the users, and maintaining the system.

Kurstedt (1985) advocates the need for all the steps described by Murdick and Munson (1986). The difference is after the system is implemented. Kurstedt (1985) views the management system as a dynamic, changing environment. This means the role of the MIS and the decision support tools changes and the environment in which they exist changes. This means that system maintenance is more than just maintaining the physical system; it includes maintaining the conceptual relationships between the system, its environment, and its uses. Schmidt (undated), referring to simulation models, points out the fragile relationship between a model and the real world system it represents. Any change in the real world (such as better information, a shift in a variable, or a change in the relationships of variables) impacts the validity of the model. This enlarges the definition of a "broken" system. It emphasizes the need for recurring and detailed reviews of any management tool by the designer and the user. For the CCET this means there is a continuing need to reevaluate the system. The users should check CCET out in their specific environment prior to initial use and on a regular basis.

TRAINING SYSTEMS

This sections consists of two parts. The first part discusses training and the training environment. The second part relates the training environment to the manufacturing environment analyzed by industrial engineers.

OVERVIEW OF TRAINING

Goldstein (1986) defines training as "the systematic acquisition of skills, rules, concepts, or attitudes that result in improved performance in another environment." This definition encompasses a wide range of activities, including things normally called education. This comes from the use of the training in another environment. The Air Force would further expand this definition to include those skills, rules, concepts, or attitudes learned on the job. This includes the on-the-job training programs used extensively in government and industry as an alternative to formal training. The key factor in all this is the requirement for a systematic approach.

Goldstein (1986) breaks the training system down into three phases: needs assessment, training development, and evaluation. Each phase consists of specific tasks. The final result of the procedure is a validated course.

The needs assessment phase consists of four different tasks: organizational analysis, task analysis, person analysis, and the comparison of training system alternatives. These tasks lead to the definition of the objectives that need training and an efficient training system. The training system efficiency is based on the cost and effectiveness of the training system alternatives.

The training development phase combining the training objectives from the needs assessment phase and the instructional technology used to build the course materials. The first step is to define the objectives by the type of objective, its relationships to other objectives, and learning theories. This information guides the course material design. Another step is defining the required training environment to support the training system design.

The evaluation phase consists of two parts, formative and summative evaluations. The formative evaluation proves that the system trains what it is supposed to. The summative evaluation proves that the training system meets the mission goals for training. This involves testing retention and the adequacy of the training to meet the job needs of the user.

Goldstein (1986) primarily was dealing with training courses developed and implemented within the organization. Air Force training covers many different types of training.

The basic types of training are: resident technical training, on-the-job training, field training detachment training, mobile training team training, and contractor training. The length of a training course ranges from a few hours to several months. Also, some courses are arranged in sequences, with entry into a course requiring the successful completion of one or more precursor courses. Course content ranges from training the trainer to training interpersonal skills to training a set of technical skills on a piece of equipment. And finally, the courses are arranged in a career progression sequence. Against this backdrop, the trainers design, develop, evaluate a dynamic set of courses, and manage the flow of trainees through the courses.

COMPARISON OF TRAINING AND MANUFACTURING SYSTEMS

It is interesting to note the similarities between how Goldstein (1986) discusses trainees and how Banks and Fabrycky (1987) discuss the products of a manufacturing system. Both references talk about the initial and final conditions, the flow, and the rules regulating flow through the systems. Industrial Engineering (IE) professionals have spent much time studying manufacturing systems. These efforts have resulted in a large body of knowledge in the efficient production of goods. A training system is also concerned with efficient production. Instead of widgets, the training system produces trained personnel.

A training system follows the multi-item multi-source with remanufacture type of system defined by Banks and Fabrycky (1987). A single course is a single-item multi-source with remanufacture type of system. The multiple sources represent the various sources of trainees entering the course (Reserves, Guard, other services, Basic Training, other training courses). The single item is the trained individual and/or class. Remanufacture refers to the recycle (washback) process. Table 1 is a comparison of training and industrial engineering terms.

Greene (1974), and Banks and Fabrycky (1987) state there is no single model of an industrial system covering all aspects of the process from planning through product distribution. Instead they both cite several different types of models, each tailored to a particular phase of the industrial process. These include ordering models, inventory models, and production models. Each of these models use different methods for tracking and predicting the process.

Just as in IE production, training production revolves around the flow of materials (trainees) through the system and the availability of resources to manufacture (train) the products. The variables affecting the flow are similar.

TABLE 1

A Comparison of Training and Manufacturing System Terms

Training	Industrial Engineering
instructor's salaries	direct labor cost
admin salaries	indirect labor cost
student salaries	inventory carrying costs
classrooms, simulators, class materials	direct materials
non-classroom facilities and materials	indirect materials
class size	lot size
evaluation system	quality control system
class frequency	reorder frequency
course length	process duration
onboard strength	in-process inventory

Desired yearly production quantity and the length of the production process provide an unconstrained figure for number of production runs and lot/class size based on one-lot-at-a-time production. Constraints on class frequency act like constraints on inventory size in determining the flow through the course. Constraints on class size act like order quantity and process flow constraints in determining the number of production runs required to meet the yearly production goals.

TRAINING MODELS

Many different training models have been developed. Some were developed before the widespread use of computers (Edney, 1972; Braby, Henry, Parrish, and Swope, 1975). Some were developed as training quality models (Advanced Technology, 1988). Others were developed to estimate costs (Orlansky and String, 1979). Most of the military models were developed for training on a particular piece of equipment or weapon system. Few take a more general look at the training environment; restricting the applicability of the existing models in new settings.

Stichna, Knerr, and Goldberg (1984) reviewed Army training models. These models were primarily task performance models simulating the steps followed in completing a task. They did not find any models of a whole training course. They referred to whole course models as general training system models. Although they did not produce a model to test, they called for developing the general training system models and integrating them with cost information to provide training managers an "essential set of tools." They say the flexibility of using simulations will allow the training managers to answer questions ranging from choosing between training course alternatives to the impact on an existing system of doubling the student load. The simulation can also include stochastic variables giving the training manager a range of estimated results. This provides the training manager with a means to assess the precision and accuracy parts of the estimate.

Advanced Technology (1984) reviewed Air Force training methods and training technologies, both current and predicted, to the year 2000. They cited the need for a set of decision aids for training managers. They recommended developing a multi-attribute utility model (MAUM) to help choose among training course alternatives. A MAUM combining the quantitative and qualitative aspects of the training system would help in selecting among alternatives if a common basis of comparison can be determined. A severe handicap of the MAUM approach is an inability to handle stochastic variables. The MAUM would provide a very precise estimate, but would not specify confidence limits of the estimate. Also, it would not assess the impact of violating the assumptions and conditions of the model.

Hawley and Frederickson (1983) reported on their efforts to develop a training decision support system. They tried to develop a set of tools to help the training designer with everything from job analysis to conducting trade-off analyses of the training system alternatives. Working on an Apple II, they started to develop their system. The limited capabilities of their hardware and software prevented them from developing the data bases necessary to do the analyses.

Knapp and Orlansky (1983) looked at the variables associated with acquiring training and training systems from the perspective of new systems acquisition. They provided a break down of training costs based on the weapon systems acquisition cost structures. They started with the research phase, proceeded to the design and initial development phases. Their final step was the production of a training system. A more common approach is to take off-the-shelf products and put them into a training program and not incur the time and money costs of training device research. This reduces the time to get a training system going and the upfront costs incurred. Still, Knapp and Orlansky (1983) do provide a detailed taxonomy and framework to structure training cost variables. But, they did not develop a training course model to use their variables and structure to compute training system (or course) costs.

Braby, Henry, Parrish, and Swope (1975) developed a technique for choosing cost-effective instructional delivery systems. Initially their technique was only a paper and pencil implementation, later it was converted to a computerized format. This technique involved three steps. The first step was the classification and organization of the learning objectives into groups according to the type of learning strategy needed. The second step involved assigning training media to the groups. Third, cost projections were made based on course delivery systems derived from the media choices. The three-step technique lists 55 different characteristics of training media, and 91 different possible media. The technique provided comprehensive guidance and help to the training developer through the course definition and trade-off analysis efforts. The very comprehensive nature of the technique made it cumbersome to use for quick "what if" analyses or for making general cost-effectiveness comparisons. It required 36 different pieces of information before it could develop a course cost estimate. Many of the 36 pieces of information included intermediate calculations. Some, like "average cost of developing one hour of instruction" were based on a series of other processes before the needed information was calculated. To make this system easier to use, Hay (1986) updated and implemented the system in a micro-computer environment.

Edney (1972) also took a general view of the training system and developed a paper and pencil cost estimation system based on the training alternatives available then. The revolution in training technologies since 1972 limits the exact use of the method described, but not the general methodology proposed. Edney (1972) proposed a detailed, mathematical model of a training system requiring the user to select the descriptor from a set of options that most closely matched the user's training system. The training system model was based on three parts: the student, the environment, and the instructional content. By analyzing each of these parts using Edney's method, the training manager would construct training system alternatives. Then the training manager would conduct trade-off analyses to determine the least costly alternative. Unlike Braby, Henry, Parrish, and Swope (1975), no updated version of Edney (1972) was found.

ITERATIVE DESIGN

This section consists of three parts. The first part discusses general iterative design procedures. The second part discusses formative evaluation. The third part discusses critical incident methods.

GENERAL PROCEDURES

Figure 1 shows the iterative design process proposed by Williges, Williges, and Elkerton (1987). It is broken down into three stages: initial design, formative evaluation, and summative evaluation. It is similar to the training design, development, and evaluation stages proposed by Goldstein (1986). The primary difference between the two methods is the stress Williges, Williges, and Elkerton (1987) put on formative evaluation. The target systems of the two different sets of design stages is different: computing and training systems. Unlike the design method proposed by Murdick and Munson (1986), Williges et al.(1987) explicitly provided forward and backward links between the stages. For this dissertation several new links are proposed to complete the iterations of the design. The first new link takes the output of the formal experimentation and links it to the design objectives. This would answer Singhal's (1986) objection to using the design objectives as evaluation

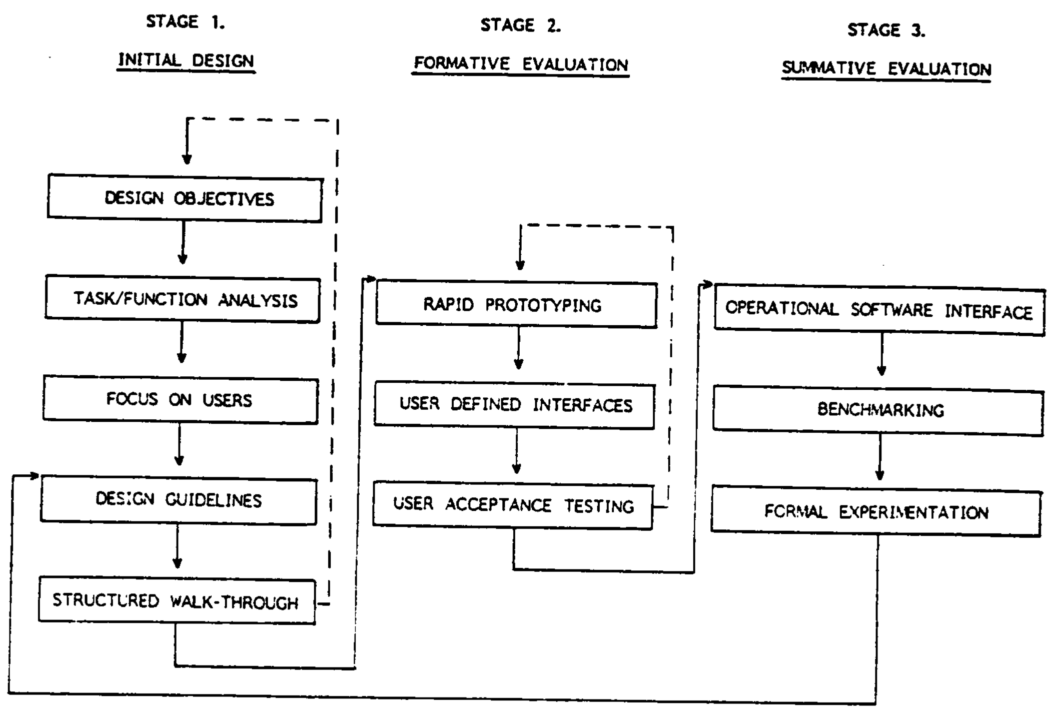


Figure 1. Iterative Design Procedure
(Williges, Williges, and Elkerton; 1987)

criteria. Singhal's objection was that the design criteria do not reflect the current use of the system. Also, it would maintain the documentation trail of the system. A second new link is between the user acceptance testing results and the design objectives. This link provides feedback from the final step of formal evaluation and the system specifications. This keeps the design specifications current. Another benefit of the new links is they provide the formal structure for maintaining and updating the system.

The initial design stage of Williges et al. (1987) involved defining the design goals, requirements, and configuration of the system. To do this, the designer collects information from the users and design guidelines. These design guidelines came from formal experimentation on earlier versions of the current system and from general design guidelines (for example, Smith and Mosier, 1987). This stage ended with the designer and the user conducting a structured walk-through of the system. The walk-through used data flows, structured English, or some other means to help the designer and the user understand each other and the system.

Following the structured walk-through, either the formative evaluation stage starts or the design process goes back to the design objectives to fix the problems discovered during the walk-through. Assuming the process has moved to the formative evaluation stage the first step was building a prototype. The prototype and later versions of it serve as a model of the interface between the user and the system. These models are useful in discovering the information format requirements of the user and checking the design objectives and task/function analysis.

The final stage was summative evaluation. This stage involves building and testing the operational system. Williges, Williges, and Elkerton (1987) proposed two different tests. The first test involved benchmark testing. This test set the performance characteristics of the system. The second test involved formal experimentation. The experiments answered two different questions. First, did the system meet the users' requirements based on the design objectives? Second, did the design objectives meet the users' requirements? The formal experimentation also provided a vehicle to analyze and improve design guidelines. To do this build different versions of the system using different values of a design guideline. Then conduct an experiment to analyze the effect of different levels of the guideline on system usability.

FORMATIVE EVALUATION

Dick and Carey (1978), while describing a method to design and develop instruction, proposed using formative evaluations to test out new instruction before finalizing and distributing it. They suggested three different types of formative evaluations for instructional systems: one-to-one, small group, and field evaluation. Goldstein (1986) defined formative evaluation as an evaluation used to test if a system is operating as planned or if changes need to be made. This dissertation did not develop classroom instruction, it developed a prototype decision support tool for individual training managers, so only the one-to-one type of evaluations were done. In field evaluations, Dick and Carey stressed the importance of making the materials and the situation as much like the operational world as possible. The dissertation collected all four types of data they suggest: test data, comments, attitude questionnaires (rating scales), and the reactions of subject-matter experts. These data were plugged back into the design process to revise the prototype.

CRITICAL INCIDENT TECHNIQUE

The critical incident technique gathers information about a system by asking people familiar with the system to report unusual occurrences they have experienced while using the system. These occurrences (incidents) are analyzed to determine the relationships between them and their impact on system performance. This information can guide the revision of the system or be used as a summative evaluation.

Fitts and Jones (1947) gathered 460 incidents on the use of airplane controls. They collected only negative incident data (control errors). They found the incidents clustered into six categories. Based on these categories they made recommendations for airplane control design, control layout, and training. Their work formed the basic procedure for the critical incident technique.

Andersson and Nilsson (1964) collected 1,847 incidents while analyzing the training and job requirements of store managers. They gathered the incidents from 410 people. They enlarged the type of incidents collected to include both positive and negative incidents. They grouped the incidents into three areas and 17 categories. They used this information to reanalyze the incidents to determine when all of the different behaviors were obtained. They did this by breaking the subject data pool into 20 groups, each group had the same relative proportion of each type of subject as the whole subject pool. They added groups to the analysis one at a time. They incremented the number of behaviors recorded by the number of new behaviors in each group. They got to 75% of the final sample size before they accounted for 95% of the behaviors. This shows strong support for the need for a large sample. They also found they got twice as many incidents when they interviewed the subjects than when the subjects filled out a questionnaire.

Both of the studies using the critical incident technique mentioned above were analyzing systems the subjects were familiar with, not new systems. Del Galdo, Williges, Williges, and Wixon (1986) modified the procedure and applied it to the analysis of a new system. They were analyzing on-line and hard copy help materials for an on-line conferencing utility. They used an on-line questionnaire (using a different CRT and keyboard) to gather the critical incident information. The questionnaire asked the user to input the type of materials used during the incident, classify the incident as a success or failure, describe the incident, and rate the criticality of the incident. They got 95 useful incidents from 16 subjects. The subjects used the utility for up to an hour. The incidents were used to make recommendations for revising the software based on frequency and severity of the incidents. They gave 31 recommendations to the software engineers. Following the revisions, a second evaluation was done to test the effects of the changes on the system. They found a significant increase in the number of positive incidents reported and decrease in the number of negative incidents.

From these studies the critical incident technique has evolved from a design guide to a robust design and development evaluation system. The critical incident method requires the user to get involved in the system evaluation. The emphasis on user involvement provides both direct design feedback and increases user ownership in the system. In this dissertation the critical incident method was used to test the first prototype of the decision support tool.

HUMAN-COMPUTER INTERFACE DESIGN

Several different sets of interface design guidelines have been proposed. Smith and Mosier (1986); and Hendricks, Kilduff, Brooks, Marshak, and Doyle (1982) represent two of these attempts. These are sets of guidelines not standards. Smith (1986) looked at the amount of knowledge available to guide a system designer in the hardware and software realms. He concluded that sufficient knowledge had been gathered to support standards for hardware design, but not for human-computer interface design. He pointed out Smith and Mosier's (1986) set of guidelines had holes in it. The holes dealt with translating the guidelines to specific design rules to help the designer. Smith (1986) pointed out that we know what the threshold for the brightness of a character on a screen is or can find out through a standard test, but we do not know what exactly constitutes a "cluttered" screen outside of a particular application. This means the software designers' job is to take the guidelines and integrate them into the interface based on the environmental context.

To help in selecting appropriate guidelines Williges, Williges, and Elkerton (1987) defined seven principles of human-computer interface design. Not all systems are equally capable of implementing the principles, particularly the mental workload, memory, and individualization principles. Another principle, consistency, may involve trade-offs. For example, the consistent user interface on the Apple MacIntosh computer across a wide range of software products contrasts sharply with the chaotic interface scene found across MS-DOS computer software. A designer would have little trouble determining what parts of an interface to keep consistent in the MacIntosh. In MS-DOS, however, the designer would have a hard time selecting something to be consistent with, not to mention choosing particular attributes of the interface to keep consistent.

The compatibility, structure, and feedback principles proposed by Williges, Williges, and Elkerton (1987) are related. Compatibility means making the system speak the users' language. Structure refers to helping the user understand the underlying framework of the system. Feedback helps the user understand the results of particular actions. If you do not present the structure and the feedback in a manner compatible with the users' world, then they cannot use it. This principle is very important when the designer's environment is not the same as the user's.

EXPERIMENTAL DESIGN

This dissertation used two different experimental designs. The first evaluation was a case study of training managers and their information requirements for estimating the costs of training. The second evaluation combined the case study design with a randomized control group post-test only design (Isaac and Michael, 1985) to test the performance of a control and experimental group of training managers.

The overall experimental design of the dissertation follows the case study method described by Isaac and Michael (1985). They cite this method as useful for studying the background, status, and interactions of a particular group of people. In this study the particular group of people were the training managers. Of interest were the interactions of the training managers with other organizations that require information derived from the cost factors. To provide the training manager with the necessary information on the cost factors, some knowledge of those factors and the relationships between factors was necessary.

Isaac and Michael (1985) discussed the strengths and weaknesses of the case study approach in a general application. They cited two strengths of the case study approach: it may uncover important variables and/or interactions for further study, and it may provide anecdotal support for statistical findings. In the context of this study the first strength means it might provide information on requirements and/or cost factors that need more work before they could be implemented. The second strength means the case study method might provide anecdotal information on the practical significance of the prototype and the documentation that may not show up in statistical significance due to the small subject sample available.

They cite two weaknesses as well. First, the narrow focus limits representativeness. Second, case studies are subject to the experimenter's and the subject's subjective biases.

Isaac and Michael (1985) point out the use of the randomized control group post-test design does allow greater validity than the use of quasi-experimental designs. The assumption that the two groups were equivalent on the variables of interest before the experiment limits the generalizability of the design. This means the prior experience of the two groups balances out. With large numbers of subjects the assumption of equal experience can be proved statistically. With small numbers of subjects the assumption is more problematic. Isaac and Michael (1985) recommend using pre-test post-test designs when small numbers of subjects are used. The session time constraint ruled out using a pre-test to establish the subject's performance level before the performance evaluation.

EVALUATION METHODS

This section discusses the literature on two specific evaluation methods. The first method is the use of bipolar questionnaires. The second method is the use of the Questionnaire for User Interface Satisfaction (QUIS) developed by Shneiderman (1987).

BIPOLAR ADJECTIVE QUESTIONNAIRES

Nunnally (1967) and Edwards (1957) provide some guidelines and cautions in selecting bipolar adjectives. One method they suggested is to add the prefix "un" or "in" to an adjective (for example: important, unimportant). A problem with adding a prefix is the adjective pair may not fully describe the range of the scale the experimenter is interested in. For example, using pleasant as one end point, add "un" and get unpleasant as the other end point. "Pleasant-unpleasant" may or may not be the same as "pleasant-repugnant". Using "approach-avoid" probably would result in different ratings than using "approachable-unapproachable". The difference is on the negative end. Unpleasant usually is not as bad as repugnant. Deciding not to add "un" or "in" as a prefix is not without risks either. It opens up alternative interpretations of the scale. For example, a subject may not rate "important-trivial" the same as "important-unimportant".

This research used two different bipolar adjective questionnaires. The first one was used to test the usefulness of the performance aids. The second one was used to test the usefulness of the spreadsheets in the second version of CCET. Bipolar adjectives were selected to span the widest range possible.

Appendix A contains the bipolar adjective questionnaire used to test the performance aids. Three different bipolar pairs were used. For each pair, one adjective phrase was "very useful". "Not useful" was considered for the other adjective phrase. It was not used because it only marked the neutral point on the scale. "Hinders" and the key descriptor of the evaluation aspect from the sentence were used. They were chosen to show a negative impact on the usefulness of the prototype.

In the evaluation of the second version of CCET, each spreadsheet was tested on its usefulness in the subjects' work setting. "Very useful" was selected as one adjective. "Unneeded" was chosen as the other adjective to determine how necessary the subjects felt each CCET spreadsheet would be in their job. In this evaluation whether the spreadsheet was useful or not was the primary concern. "Unneeded" was chosen to emphasize the usefulness of the spreadsheet to the user.

<u>QUIS</u>

Shneiderman (1987) presented a long and short form of a questionnaire for users to rate interactive computer systems subjectively. This questionnaire formed the seed for QUIS. It consisted of 49 questions in seven different categories. The first seven questions asked for information about the subjects' computer experience and experience with the system being rated. The other 42 questions were answered by selecting a value on a bipolar adjective scale. The negative pole of the scale was always on the left side of the scale. Norman, Chin, and Shneiderman (1987) discussed some of the steps that resulted in the development of QUIS from the questionnaire in Shneiderman (1987). This study used version 5.0 of QUIS.

QUIS asks the subject to rate many different aspects of the hardware and software systems being rated. The areas rated are: type of system, experience, screen, terminology and system information, learning, and system capabilities. Appendix B contains a copy of QUIS. Although the questionnaire covers areas not considered in the prototype tool (hardware, help, advanced features), QUIS was used in its entirety.

FIRST EVALUATION

In the first set of evaluations, data were gathered on the subjects' opinion of the CCET and four performance aids. The evaluations were based on a demonstration of the first version of the CCET. This chapter consists of six sections. The first section describes the hypotheses evaluated in the first evaluation. The second section describes the tests of the hypotheses. The third section describes the methods and procedures used. The fourth section presents the results of the evaluations. The fifth section discusses the results of the first evaluation. The sixth section presents the conclusions from the First Evaluation.

HYPOTHESES

The first two hypotheses for the First Evaluation was concerned with the users' perceptions of the CCET.

1. The training managers will find the CCET to be useful for making training course cost estimates in their environment.

2. The training managers will find the CCET appropriate for making training course cost estimates in their environment.

The third hypothesis concerned the perceived effectiveness of four performance aids on using CCET.

3. The training managers will rate the four performance aids differently on how they effect accuracy and time to make a course cost estimate, and in their effect on the training managers' understanding of the CCET.

The second hypothesis was analyzed in three parts. The three parts were:

a. Training managers will rate the four performance aids differently in their effect on user accuracy in using the CCET,

b. Training managers will rate the four performance aids differently in their effect on the amount of time it takes a user to get a cost estimate with the CCET, and

c. Training managers will rate the four performance aids differently in their effect on the user's understanding of the CCET.

The four performance aids were developed to help the users with CCET. The four performance aids differed in format and level of detail. The two formats were outline and flow chart. Each of the formats were used for a high and low level of detail performance aid.

TESTS OF THE HYPOTHESES

For the first evaluation, the training managers watched a demonstration of CCET individually, making notes about what they liked and did not like. The experimental design was a case study of the training managers and their requirements for course cost information.

In the First Evaluation the usefulness of the CCET was tested by analyzing the comments from the critical incident rating task. The type, number, rating, frequency, and content of the comments were all used to analyze the usefulness of the prototype.

The data from evaluation of the performance aids were analyzed using a Friedman Two-Way Analysis of Variance and a pair-wise multiple comparisons test based on Friedman Ranked Sums. The two factors were subject and performance aid. Each subject rated the four performance aids on three questions. Each question was evaluated separately. The non-parametric statistics were chosen because the sample size was small and the data came from rating scales. These statistics do not assume the data come from a normal distribution, are measured on an interval scale, and have equal variances between the groups of data by performance aid. The Friedman Two-Way Analysis of Variance tested for an overall significant difference between the performance aids on a particular question across the subjects. The selection of the materials for the performance evaluation part of the second evaluation was based on the following two rules.

1. If a significant difference is found between one or more pairs of the performance aids, then the pair with the largest difference in Friedman Ranked Sum will be used in the second evaluation. One group will get the user's manual and the least preferred of the pair. The other group will get the user's manual and the most preferred of the pair of performance aids.

2. If no significant difference is found between the groups, then the most preferred performance aid and the user's manual will be the materials for the evaluation. One group will get the user's manual only. The other will get the user's manual and the most preferred performance aid.

METHODS AND PROCEDURES

The Methods and Procedures section consists of two parts. The first part describes the development of the first version of the CCET. The second part describes methods and procedures of the first evaluation.

DEVELOPMENT OF FIRST VERSION OF CCET

This part consists of six areas. The first three areas describe the first version of CCET. The second three areas describe the algorithms used to estimate costs for courseware, classrooms, and personnel.

Description of CCET

The first version of CCET consisted of eight different spreadsheets. Appendix C contains the screen prints from this version of the prototype. The user entered course information on the Input Spreadsheet. The other seven spreadsheets provided various levels of detail for the cost estimates. The other seven spreadsheets were broken down into five categories based on what estimates they provided: course processing, courseware, classroom, personnel, and life cycle.

The course processing category consisted of one spreadsheet, which processed information from the CCET's Input Spreadsheet and estimated various course parameters. It estimated length of course, number of classes per year based on class size, and number of classes per year based of facilities as intermediate calculations for this spreadsheet. The Course Processing Spreadsheet also generated estimates for use on other spreadsheets. These estimates were total trainee days, number of instructors, and number of classes per shift.

The courseware category consisted of two spreadsheets: one estimated the cost of developing courseware, the other estimated the cost of maintaining courseware. Both of these spreadsheets used the number of hours of instruction delivered by each type of media the user entered on the Input Spreadsheet. The spreadsheets used this data to estimate the cost of courseware development and courseware maintenance.

The classroom category had two spreadsheets: one for classroom acquisition, the other for classroom maintenance. Ten different types of classrooms were used to develop the cost estimates. A particular course might have all or just a few of the classroom types available. The CCET had reference costs for lecture (including equipment for films and tapes), computer assisted instruction (three different classroom configurations), interactive videodisc (three different classroom configurations), and three equipment laboratories. The cost information for the three equipment laboratories used in these spreadsheets came from the Input Spreadsheet. This was due to the wide range of operational equipment used in USAF training and the wide range of costs involved. For the other classroom types, the CCET had cost estimates included in the model.

The personnel category consisted of one spreadsheet. This spreadsheet calculated the cost of instructors and trainees. The instructor cost was based on an annual cost (including base support costs) for each instructor and the number of instructors required for the course. The trainee cost was based on the number of trainee days per year and an estimate of the cost per trainee day. The cost estimates included more than salary and benefits for the instructor and the trainee. Also, the CCET used a factor for indirect costs associated with both instructors and trainees when computing personnel cost. The indirect costs include the cost of other training support activities on the base and a part of the cost of management activities at higher headquarters.

The life cycle cost category consisted of one spreadsheet. This spreadsheet got the totals from the other cost estimating spreadsheets and placed them in a life cycle cost framework. The first year costs included courseware development and the classroom acquisition costs. These were considered one time expenditures. During years two through six, the course was assumed to be in full operation. CCET assumed the course incurred courseware maintenance, classroom maintenance, and personnel costs for years two through six.

Variables

The variables used in the CCET were broken down into three sets: input, output, and internal. There were 23 input variables and 12 output variables. The internal variables had default values coded into the CCET spreadsheets. Table 2 shows the input variables used in the prototype. If the course required specific equipment, the user could modify the values of some of the internal variables, such as computer terminal costs. An example of specific equipment requirement is requiring emanation shielded terminals instead of regular computer terminals.

The input variable set was designed to reflect information a training manager would have access to. The set was kept small to minimize the amount of information needed from the user. The user was not required to input values for all of the input variables, some were optional. For example, the user could input the number of instructors or allow the CCET to derive it.

The output variable set was designed to provide the training manager with the information needed to answer questions about training course costs. The output variable set consisted of the cost of developing courseware, the cost of acquiring classrooms and equipment, the personnel cost (both student and instructor), the cost of maintaining the courseware, the cost of maintaining the classrooms, the total cost per year of the life cycle, and the total life cycle cost. The variable set was derived from interviews with several training managers. Data from the

TABLE 2

Input Variables

Variable Group	Variable
Course Information	
	Trained Personnel Requirement Attrition Rate Maximum Class Size Maximum Number of Classrooms Average Percent of Course Material Revised in a Year Number of Instructors (optional)
Course/Media Allocation	
	Number of Hours of Instruction in Each of Nine Media Categories by the Number of Instructors Required
Classroom Configuration	
	Configuration of Computer Assisted Instruction Classrooms Configuration of Interactive Video Disc Classrooms
Classroom Cost	
	Cost of Equipment Laboratories
Miscellaneous	
	Course Model Number Inflation Rate

questionnaires provided validation of the output variable set.

Structure of CCET

A spreadsheet format was chosen as the structure of CCET for several reasons. First, it was a simple structure for both the experimenter and user to understand. Second, it facilitated modifying CCET to reflect the information gained in the evaluations. Third, it would operate in a microcomputer environment. This allowed the experimenter to conduct the study in the training managers' environment and not impact local computing systems. Also, it removed the necessity to rehost the prototype at each training center and eliminated a potential source of errors. Finally, the spreadsheet format provided near-instantaneous system response time.

CCET consisted of one spreadsheet page divided into separate spreadsheets. Each fit on one screen, except one where the user had to scroll to the left to cover the whole spreadsheet (Course Life Cycle Cost Spreadsheet). The spreadsheets were placed down the diagonal of the spreadsheet page with no sharing of columns or rows between the spreadsheets. This meant columns or rows could be added within a spreadsheet to meet the users' requirements when needed without affecting the other spreadsheets.

Courseware Cost Algorithm

The courseware cost algorithm consists of two parts: courseware development and courseware maintenance. Cost per hour of delivered instruction was computed for each training media option for both courseware development and maintenance. This involved dividing the development effort into the types of people needed: instructional designers, subject matter experts, media personnel, programmers, and typists. Each of these types of people may contribute a part to developing a course and the size of the contribution varies with the media type. Appendix D shows the courseware development breakdown structures. Training managers and training developers from the private sector reviewed the cost figures to determine the realism of the types of people and the costs for each media type. CCET generates representative cost figures, not exact ones. That is, the cost numbers used are estimates based on general experience, not on a particular development effort. Exact numbers depend on many more variables than the courseware development breakdown covered. Two examples of cost variables that CCET did not consider are the number of times the material is revised before it is accepted and the specific labor times and rates involved. The CCET uses average cost figures to derive its estimates.

Appendix D shows the cost breakdown algorithm of courseware development for the first version of CCET. Once the decision was made to compute the cost of developing courseware by using a fixed cost of development per hour of delivery, the next step was to break down the development process. The process was broken down into three stages: the contribution to the development effort of different personnel categories, hours of development per hour delivered, and the cost of the personnel categories.

Where possible, CCET used actual labor categories and cost values, allowing a more precise estimate of costs. The choice of three categories for computer-assisted instruction and interactive video disc simplified the way CCET considered the material taught, the teaching strategy used, and the hardware and software systems used.

Classroom Cost Algorithm

To provide a range of classroom environments, two categories were developed: system-specified and user-specified. The system-specified classrooms covered those classrooms with similar equipment such as lecture, computer-assisted instruction, and interactive videodisc. The user-specified classrooms covered the special equipment classrooms (laboratories) that contain equipment specific to a specific course or family of courses. CCET required the user to provide the acquisition and maintenance costs for the equipment laboratories.

Based on the categories described, the classroom costs for system-specified classrooms were constructed by partitioning each classroom type into its components: room, desks, and other equipment. A cost was assigned to each component and the component costs summed to get a representative classroom cost. Costs were assigned for both the acquisition and maintenance of each classroom component. In the version of CCET used in the First Evaluation only the cost per classroom was provided as part of the tool. Background figures and calculations (type, amount, and cost of equipment) were provided in a handout.

The configurations for the computer-assisted instruction and interactive videodisc classrooms were based on the three configurations proposed in Eagle Technology (1987). These are: a classroom configuration with no data transfer capability beyond the classroom; networked classrooms with an on-site central computer; and networked classrooms with an off-site central computer.

The CCET derived the number of classrooms of each type from the course-media profile and the course flow data (trained personnel requirement and attrition rate). Then the CCET used the per-classroom cost estimates to develop the cost estimate for each classroom type. The figures for classroom acquisition were summed to provide the total estimated classroom acquisition cost. The maintenance costs were summed and used for the classroom maintenance cost for Year 2 in the life cycle. This figure also served as the basis for the discounted figures used for years 3-6.

Personnel Cost Algorithm

A wide variety of personnel provide and take USAF technical training. Usually, the students and instructors are USAF enlisted personnel. But, USAF officers, USAF civilians, personnel from the other branches of DoD, foreign military members, and others instruct and are trained. This leads to a large variety in trainee personnel costs. Some charged to Air Training Command, some charged to other Air Force components, and some charged to other agencies. There is usually less variety in the types of instructors. Instructors are generally Air Force enlisted, officer, or civilian members. Just using the four different pay schedules for the basic categories (USAF officer, USAF enlisted, wage grade civilian, and general schedule civilian) complicates calculation of an average instructor cost. And the pay schedules are not necessarily the most meaningful source of information. The true cost of training includes the total costs of personnel (retirement, education plan, leave), base support (housing, personnel, finance, medical facility), and upper management (Headquarters Air Training Command and some portion of the DoD-level agencies concerned with training).

For the first prototype, the complex situation described was greatly simplified. First, the personnel costs were broken down into two categories: trainee and instructor. Then a total cost for each category was estimated. For the trainees, the total cost was based on the number of trainee days or student man-days. For the instructors, the total cost was based on the number of instructors per year. Both estimates were based on junior enlisted personnel salaries. However, CCET permitted the user to modify the cost estimates to fit specific conditions.

The different time periods used for the cost estimates reflect different assignment philosophies. Trainees are assigned to a base for the course plus some in and out processing time. Therefore, savings can be achieved in a training course by reducing the number of student man-days required per year. This can be done by shortening the course, reducing attrition, or both. Instructors are usually assigned to a base for a three or four year period, most of which they will spend as instructors. Savings can be achieved by reducing the number of instructors required by a course.

PROCEDURE FOR THE FIRST EVALUATION

This section consists of three parts. The first part describes the subjects used. The second part describes the procedure used. The third part discusses the data analysis procedures used.

Subjects

Eight training managers were used for the first evaluation. Three of the managers were from one training center and five were from another. They had between one and fifteen years of experience as training managers and all had over six years experience in USAF training.

Procedure

The materials for the first evaluation consisted of the informed consent form, the critical incident questionnaire, and the performance aids of the prototype (Appendix E). Each subject received the same materials. Also, both the exper nenter and the subject had paper for note taking. The subj(ts' notes formed the basic data for their critical inci ent questionnaire comments.

The procedure for the First Evaluation followed the case study method. The session started with an overview of the project, its goals and possible applications. Next, each subject was given an informed consent form. As the subject read the form, the experimenter answered any questions about the experiment. Once the subject understood the background, the goals of the session, and signed the informed consent form, the session began.

After the subject finished the informed consent form, the experimenter demonstrated the system on two course-cost-estimation tasks. The tasks were designed to show the subject each of the spreadsheets in the tool. After the two tasks were demonstrated, the experimenter asked the subject if he or she had any questions. Six subjects asked to see the tool work using some of their own data. The experimenter put their data into the tool and showed the subject how the tool processed the data. After the tool was demonstrated, the subject completed the Evaluation of the Cost Estimating Tool questionnaire. This asked the subject to write down those features they liked about the system, those features they did not like, and those features they would like to see in the next version of the tool.

Data Analysis

The first hypothesis was tested using the critical incident method. The critical incident method of evaluating CCET and its results did not permit analysis for statistical significance. The data were used to describe the subjects' reactions to CCET and to guide revisions of the first prototype version of CCET.

The data from the evaluation of the performance aids were analyzed to determine the most and least preferred performance aid. This information was used to test the second hypothesis. The questionnaire data was in the form of ratings. The ratings were on a seven point scale with _1_ indicating "very useful" and _7_ indicating "hinders use". The subjects rated each of the performance aids on three different questions. The first question concerned the usefulness of the performance aid in helping the user make accurate estimates of training course costs with CCET (hypothesis two, part one). The second question concerned the usefulness of the performance aid in helping the user make quick estimates of training course costs using CCET (hypothesis two, part two). The third question concerned the usefulness of the performance aid in helping the user understand the CCET (hypothesis two, part three).

The ordinal nature of the data called for the use of non-parametric statistics to analyze the results. A Friedman Two-Way Analysis of Variance was chosen because of its ability to handle the small number of subjects and the range of possible ratings. The independent variable was "subject" and the dependant variable was "performance aid." A separate analysis was done for each question and the choice of the performance aid was based on the decision rules shown on page 12.

RESULTS

This section consists of three parts. The first part presents the results from the critical incident questionnaire. The second part presents the results of the debriefing of the subjects after the demonstration and the critical incident questionnaire. These two parts relate to the first hypothesis. The third part presents the results of the evaluation of the performance aids. This relates to the second hypothesis (parts a, b, and c).

CRITICAL INCIDENT TECHNIQUE RESULTS

Appendix F contains the results from the critical incident evaluation of the first prototype version of CCET. Each subject made both positive and negative comments about CCET. After making the comments, each subject rated their own comments on how important the comment was. The rating scale went from 1 (absolutely critical) to 10 (nice to have but not essential). At both training centers, the subjects reported that the CCET was useful and appropriate for estimating the cost of training courses. This anecdotal evidence supports the first and second hypotheses. Several training managers asked to see what the prototype would do with their actual data. In each case, where a comparison between their actual data and the estimates from CCET was done, the training managers said the estimates matched what they would expect based on their experience.

Positive comments focused on the first prototype version of CCET's capabilities. Training managers rated two positive comments less than 5 (the critical to the mission half of the rating scale) and five comments higher than 5 (nice to have but not critical). The subjects did not rate two comments. Three of the nine comments concerned the ease of operation and simplicity of the CCET.

Negative comments focused on necessary changes to the prototype. They rated four comments less than 5 (critical), and three higher than 5 (nice to have but not critical). Three comments were not rated by the subjects. Rated comments reflected the perceived need for the training managers to tailor the values to their particular situations. Unrated comments dealt with adding or enhancing variables to account for the full range of technical training.

DEBRIEFING RESULTS

Subjects made five additional suggestions during the demonstration and the debriefing. The experimenter recorded these comments. First, the subjects wanted to compare different configurations of the same course to see which was the least costly. Second, they wanted to see how the classroom costs were computed. Third, they suggested the CCET should be modified to accommodate courses requiring more than one instructor at a time. Fourth, rather than the one classroom, one instructor rule used in the initial CCET; they suggested the InterService Training Review Organization formula. And fifth, they wanted to be able to input the number of instructors available per course.

EVALUATION OF PERFORMANCE AIDS RESULTS

Table 3 provides a summary of the results of the Friedman's Two-Way Analysis of Variance done on the three questions. No significant difference was found between the ratings of the performance aids on any of the questions. Therefore, the results failed to support any of the three parts of the second hypothesis. The performance aids were not rated differently on their effect on user accuracy, amount of time to develop a cost estimate, or understanding of the CCET.

TABLE 3

Friedman Two-Way Analysis of Variance Comparison of the Four Different Performance Aids

Question	Model Chart	Detailed Chart	Model Outline	Detailed Outline	p>
	Performance Aid Rank Sums				
Understanding	14.0	11.5	19.0	15.5	.35
Accuracy	13.5	12.5	17.5	16.5	.65
Speed	15.0	14.0	14.0	17.0	.90
Overall Sum	42.5	38.0	50.5	49.0	

Notes:

1. p> is the probability of the differences between these rank sums occurring by chance.

2. larger rank sum indicates the more preferred performance aid.

DISCUSSION

The first evaluation is discussed in two parts. The first part relates to the first hypothesis. It discusses the critical incident evaluation and the demonstration of CCET. The second part relates to the other three hypotheses and discusses the evaluation of the performance aids.

CRITICAL INCIDENT TASK RESULTS

The results from the critical incident evaluation produced no firm conclusion about the usability and appropriateness of the CCET (hypothesis 1: training managers will find the tool usable and appropriate). However, the content of the comments supported the need for a cost estimating tool. The comments were nearly evenly split between positive and negative. The negative comments rated the most critical dealt with changes to the existing prototype, not discounting the utility or the general implementation. The small number of comments and lack of a specific question on the usefulness of the prototype weaken the interpretation of the data.

Other uses of the critical incident evaluation techniques have involved hundreds of subjects and hundreds of comments. In this case there were six subjects and nineteen comments. Also, the amount of time the subjects saw the prototype was short, under one hour. Finally, the subjects did not actually use the prototype in their own job or to complete any standard tasks (as in the second evaluation).

These factors limited the amount of information obtained and the amount of analysis that could be done on the information obtained. However, the factors did not necessarily impact the validity of the information or usefulness of the information in building the second prototype.

PERFORMANCE AIDS EVALUATION

The analysis of the results from the evaluation of the performance aids in the First Evaluation produced no significant overall difference among the subjects' preferences across the performance aids. So, the most preferred performance aid was used in the Second Evaluation. Several factors (individually or in combination) could account for this result. Two factors of the possible factors are sample size and difference between the ranked sums.

First, the number of subjects involved in the evaluation was small. The few subjects used in the First Evaluation means the differences between the ratings of the spreadsheets must be large and the rating range for each spreadsheet small for the differences between the spreadsheets to be significant.

Second, the distribution of the rank sums could impact the overall differences. The rank sums for "understanding" were the only ones where there was at least a one unit difference between each of the rank sums. This coupled with the small sample size could explain the lack of a significant difference in preference between the performance aids.

CONCLUSIONS

The results from the first evaluation supported the first hypothesis and the value of the critical incident method in evaluating the prototype. It also showed the need for longer evaluations. The first evaluation used a prototype to support the structured walk-through. This allowed the user to evaluate more than the control flow and data descriptions.

The results from the subjective evaluation of the performance aids did not support any of the three parts of the second hypothesis. The performance aids were not rated significantly differently on any of the three variables used.

SECOND EVALUATION

In the second set of evaluations, data were gathered on the subjects' performance using the CCET and their opinion of the CCET. The evaluations were based on using the second version of the CCET. This chapter consists of six sections. The first section describes the hypotheses evaluated in the second evaluation. The second section describes the tests of the hypotheses. The third section describes the methods and procedures used. The fourth section presents the results of the evaluations. The fifth section discusses the results of the first evaluation. The sixth section presents the conclusions from the Second Evaluation.

Hypotheses

The hypotheses for the Second Evaluation consisted of two groups. The first group consisted of four hypotheses and compared the performance of the experimental and control groups in answering questions based on three different scenarios. The second group consisted of two hypotheses and involved the subjects' responding to two questionnaires evaluating CCET.

The performance of the subjects was analyzed with three specific hypotheses based on the comparison between an experimental and a control group in performing three sets of tasks. The experimental group had a users' manual and another performance aid to help them in completing the tasks. The control group only had the users' manual to help them. Of the four hypotheses in the first group the first three were:

1. The experimental group will make fewer errors than the control group,

2. The experimental group will use fewer spreadsheets to find the required information than the control group, and

3. The experimental group will take longer to complete the tasks than the control group.

Three performance variables (number of errors, number of spreadsheets used, and time to complete the tasks) were selected because they were the performance counterparts to the three questions on which the subjects in the first evaluation rated the performance aids. The number of spreadsheets used by a subject was used as an indication of the subject's understanding of the CCET (subjectively evaluated in the first evaluation). The other two, errors and time to solve problems, were measured by the performance of the subjects instead of being subjectively rated (as in the first evaluation).

The first two performance hypotheses focused on the greater amount of information available to the experimental group. By using the charts with the manual the subjects could trace the flow of information and find the location of the answers to the questions in the scenarios. Tracing the information flow allowed the experimental group to explore CCET without going to spreadsheets that did not have the necessary information. This should result in the subjects using fewer spreadsheets and making fewer errors. The third hypothesis was based on greater time requirements to look through the charts and then using the navigation information to manipulate the spreadsheets. This should increase the overall time to complete the tasks.

The fourth hypothesis dealt with the experimental group only. It concerned their usage of the charts across the session.

4. The experimental group will use the charts more often on the questions in the first scenario than on the questions on the second scenario.

The fourth hypothesis concerned the ease of use of the CCET. If the training managers had a hard time learning the titles used, and the relationships between information and its location in the tool in CCET; then the users would continue to rely on the performance aid. On the other hand, if the tool was easy to understand, then the users would need the performance aid to help understand the structure initially and then be able to use the CCET without the performance aid.

The subjective evaluation of the second version of the CCET was broken down into two parts. The first part involved how the subjects rated the individual spreadsheets. The second part involved the subjects' ratings of the tool's usefulness. These two different parts provide a piecewise and overall evaluation of the CCET and its usefulness to training managers.

The first part of the subjective evaluation, the rating of the individual spreadsheets, consisted of one hypothesis:

5. The experimental and control groups will rate the spreadsheets differently.

The second part of the subjective evaluation consisted of one hypothesis:

6. The users will find the CCET easy to use and appropriate for their environment.

This hypothesis involved the users rating the CCET on its overall usefulness. The information on this question was gathered from the comments made about the spreadsheets and the subjects' responses to the Questionnaire of User

Interface Satisfaction (QUIS, developed by Norman, Chin, and Shneiderman,(1987).

TESTS OF THE HYPOTHESES

The four performance hypotheses from the Second Evaluation were analyzed with independent one-tailed t tests. Data from the rating and questionnaire instruments were used to analyze the usefulness of the CCET in the Second Evaluation.

METHODS AND PROCEDURES

The Methods and Procedures section consists of three parts. The first part describes the development of the second version of CCET. The second part describes methods and procedures of the second evaluation.

DEVELOPMENT OF THE SECOND VERSION OF CCET

This section covers two areas. The first area describes the second version of CCET and compares it with the first one. The second area describes the courseware cost algorithm used in the second version.

Description of CCET

The second prototype version of CCET consisted of twelve different spreadsheets, the eight from the first prototype version and four additional ones. The additional ones were designed in response to the suggestions from the First Evaluation. The spreadsheets were still divided into the same five categories the first version did. Appendix G contains the screen prints of the second prototype. The courseware and personnel categories emerged unchanged from the first evaluation. All of the other categories had some changes.

Three changes improved the Input Spreadsheet. First, the user could enter the number of hours by media type by number of instructors required (up to two instructors). Second, the user could estimate some of the course costs across three different versions of a course. Third, the user could specify the number of instructors available for a course.

The Course Processing category of the CCET was changed by adding a new spreadsheet and by changing the method of estimating the number of instructors required in a course. The new spreadsheet explicitly showed the estimation of the number of classrooms required by type of classroom. The subjects in the First Evaluation suggested the CCET use the InterService Training Review Organization (ITRO) formula to estimate the number of instructors. The ITRO formula includes time factors for course delivery, course preparation, course maintenance, and other duties. The CCET already estimated the time and cost of course maintenance, so that factor was dropped from the formula.

Two new spreadsheets were added to the Classroom category of CCET. The two spreadsheets broke the classroom cost estimates into the facilities and equipment cost estimates for each type of classroom. This helped the user check the default equipment configurations against their specific equipment requirements and costs. If the default configurations were inaccurate, the user could change the standard configuration and costs to reflect their specific situation.

A new spreadsheet was added and changes to the existing one improved the Life Cycle Cost category of the CCET. The subjects from the First Evaluation suggested the changes to make the CCET more useful. The new spreadsheet permitted the user to compare up to three different course configurations on the cost estimate totals. The specific information about configurations was not retained except for the last configuration entered. The life cycle spreadsheet from the first version of CCET was changed to show the effects of inflation in years two through six.

Courseware Development Cost Algorithm

The courseware development information gathered during the First Evaluation was used to refine the courseware development cost estimation method used in the second prototype version of CCET. The method used in the first prototype version of CCET used specific labor categories (such as, junior instructional designer). The subjects in the First Evaluation thought specific labor categories were more detailed than necessary for the CCET. Also, some labor groups were unintentionally omitted from the first prototype version of CCET. Appendix G shows the modified method of estimating courseware development for the media categories. This spreadsheet aggregates labor categories into training development functions (for example, instructional design). Instructional design could involve more than one person, such as senior and junior instructional designers. Aggregating the labor categories into training development functions reflected the way the subjects viewed the process.

The courseware cost estimation method was changed to use three inter-related spreadsheets to develop the estimate. The first spreadsheet was used to calculate the ratios of the training development categories across media types. Data from the first spreadsheet was used to calculate values in the second. The second spreadsheet takes the ratio of contributions within a media type and derives the per cent contribution of each training development category. Using the first spreadsheet the training manager can check the changing ratio of contributions for each training development category across the media categories. With the second spreadsheet the training manager can check the per cent contribution by training development category within a media category. The training manager can go back to the first spreadsheet and adjust the ratios until the percentages and the ratios approach the desired values.

After adjusting the ratios of contributions across training development categories the training manager would look at the bottom line of the courseware development cost estimation tool: the estimated cost of developing one hour of courseware for each media. This cost estimate was computed as the sum of the products of the per cent contribution multiplied by the estimated number of development hours multiplied by the labor cost per hour for each labor category in a media category.

PROCEDURE FOR THE SECOND EVALUATION

This section consists of three parts. The first part describes the subjects used. The second part describes the procedure used. The third part discusses the data analysis procedures used. Appendix H contains the printed materials used in this evaluation.

Subjects

Eighteen training managers from three different USAF Air Training Command Technical Training Centers were used in the second evaluation. Their experience ranged from three to fifteen years as training managers. Some had experience in developing training programs using interactive videodisc or computer assisted instruction.

Procedure for the Second Evaluation

The procedure for the second evaluation followed the randomized control group post-test only experimental design. The control group had the users' manual to help them with the tasks. The experimental group had the users' manual and the performance aid (detailed flow charts) to help them.

The experimenter gave the subjects both an overview of the study and an overview of the goals of the second evaluation. Then they were given the Informed Consent form and the experimenter reviewed their rights as subjects in the experiment as they read the form. After they had signed the form, the session began.

The subjects were given the users' manual and the computer with CCET installed on the computer. CCET was set to display the Input Spreadsheet. The experimenter guided each subject through the users' manual. The subject was permitted to take as much time as needed to review the manual.

After the subjects in the experimental group had reviewed the manual, they were given the performance aid. The experimenter guided each subject through the performance aid, pointing out the structure of the charts and their contents. The subjects were then permitted to review the performance aid.

After each subject was finished reviewing the materials provided and had adjusted the computer display screen for comfort, the subject was given the scenarios and question sets. Then they started the three scenarios. Each scenario consisted of a set of input conditions followed by four questions. The subject had to change the Input Spreadsheet to reflect the input conditions. Then the subject had to get the answers to the questions from other spreadsheets in the CCET. The subject used combinations of two keys pressed simultaneously to change spreadsheets.

After the subject completed the third scenario, the subject was given the evaluation materials. The evaluation materials consisted of the spreadsheet rating questionnaire, the spreadsheet comment questionnaire, and QUIS. The experimenter told the subjects to use the CCET as necessary to refresh their memory or look at spreadsheets they had not seen in the scenarios. Also, the experimenter was available to answer any questions. On QUIS, the subjects were told to mark "not applicable" if, in their opinion, the question did not apply to their situation. The subjects were told to rate the spreadsheets and make their comments based on applying the tool to their job, not the experimental situation.

Data Analysis

The performance data were analyzed in separate *t*-Tests. This related to hypotheses one, two and three. The three performance variables (total time, number of spreadsheets, and number of errors) are assumed to be logically independent. For example, there are three different combinations of the three variables which result in the same number of errors but not the same number of spreadsheets used and/or total time to complete the task. First, you can make a mistake quickly (quick completion time). Second, you can make the mistake after spending a lot of time in the printed materials or after spending a long time on one spreadsheet (long completion time and few spreadsheets). Third, you could make the mistake after spending a considerable amount of time switching between spreadsheets (long completion time and many spreadsheets). Each of these possible combinations of the three variables resulted in changes in one of the other two variables and no change in the number of errors variable. Similar combinations of the three variables can be constructed to show changes in two of them does not necessarily mean the other one changed values. To test statistical dependence, a correlation analysis was done on the variables.

The fourth hypothesis was analyzed using a one-tailed t test. The number of charts used by the experimental group in the first scenario were compared to the number of charts used in the second. To do this the number of charts used to answer each question within a scenario was summed.

The data from the Rating the Spreadsheets Questionnaire were analyzed using the Friedman's Two-Way Analysis of Variance to test the fifth hypothesis. The overall difference in ratings between the groups was the primary point of interest. Another point of interest was looking for detached responses. That is, were there rating distributions with one or more zero frequency ratings between non-zero frequency ratings? Then the Friedman test was run to see if the groups rated the spreadsheets significantly differently.

The sixth hypothesis was tested using the information from the Rating the Spreadsheets Questionnaire and QUIS. Positive ratings on both the questionnaires support the usefulness of the CCET. Negative ratings indicate a lack of support.

RESULTS

This section consists of four parts. The first part presents the performance results (hypotheses one, two, three, and four). The second part presents the ratings of the spreadsheets. The third part presents the comments on the spreadsheets. These two parts relate to hypotheses five and six. The fourth part presents the results of the QUIS evaluation (hypothesis six).

PERFORMANCE RESULTS

The results from the performance evaluation were measured by: number of errors, number of spreadsheets used, and time to complete the scenarios. Within the first variable (errors) the type of error, number of errors, and where the error occurred were analyzed. The second variable consisted of the total number of spreadsheets used to complete all questions. The third variable consisted of the overall time to complete the scenarios.

Table 4 shows the results of the correlation analysis. The largest correlation between a pair of variables was

TABLE 4

Correlation Analysis of the Performance Variables

Variable	Errors	Time
Spreadsheets	-.064	.320
Errors		-.002

0.32. No significant correlation was found between the variables in the performance evaluation.

The one tail t test analysis revealed no statistically significant differences between the groups on the performance measures. These results fail to support hypotheses one, two, and three. Table 5 shows the results of the one tail t test analysis of the two groups on total session time, number of errors, and number of spreadsheets used to complete the questions.

The average number of errors was less than one error per subject for both groups. The overall error total was fourteen, and the error rate was 6.2%. The Manual and Charts (experimental) group made eight errors while the Manual Only (control) group made six. The type of error made was fairly evenly split between the two types (wrong spreadsheet and reading the wrong value from the correct spreadsheet) possible for both groups. Also, both groups made half or more of their errors on the same question: scenario two, question two. The types of errors on this one question differed between the groups. The Manual and Charts group made wrong spreadsheet errors and the Manual-Only group made mostly wrong value errors (two wrong values, one wrong spreadsheet). Comparing the observed distribution of

TABLE 5

Results of Performance Evaluation

Variable	Group	n	Mean	t	df	p>
Number of Errors	Manual and Charts	9	.89			
	Manual-Only	9	.67	.63	16	.53
Number of Spreadsheets	Manual and Charts	9	11.3			
	Manual-Only	9	11.6	-.18	16	.85
Total Time (minutes)	Manual and Charts	9	22.67			
	Manual-Only	9	23.5	-.44	16	.66

Notes:

1. n is the number of subjects.
2. t is the value of the Student t statistic.
3. df is the number of degrees of freedom associated with the statistic.
4. p> is the probability of obtaining that value of the Student t statistic or larger by chance.

errors in the first two scenarios to the distribution expected with the Chi Square test revealed a significant ($p<.001$) result. This indicates the errors were not randomly distributed.

The Manual-Only group made six errors. This represented an error rate of six per cent. Four subjects made one error and one subject made two. Half of the errors occurred because the subjects used the wrong spreadsheet. The other half occurred because the subjects recorded the wrong value from the right spreadsheet.

After the first scenario, the subjects did not use the charts very often. Three of the nine subjects in the Manual and Charts group never used a chart directly in answering a question. Another three of the nine subjects used the charts on four to six questions. Only one subject used the charts to help answer half the questions. The subjects consulted the charts only after giving up on answering the question using the information in the manual. A Chi Square analysis was done combining the number of subjects using charts on questions within a scenario. The result was significant ($p<.05$), showing spreadsheet usage was significantly different across the scenarios. The number of subjects using the performance aid decreased from six in the first scenario to three in the second to none in the third. This supports the fourth hypothesis.

RATING THE SPREADSHEETS RESULTS

The ratings of the spreadsheets had two general patterns: a concentration in values 1 and 2 and a more dispersed pattern across both the experimental and control groups. There was no significant difference between the groups in their ratings of the spreadsheets. This did not support hypothesis five. The subjects rated the two summary spreadsheets as being the most useful were (Life-Cycle Cost Estimate and Comparison of Course Alternatives). The spreadsheets the subjects rated the least useful differed between the groups. The Manual Only group rated the Classroom spreadsheets least useful. The Manual and Charts group rated the Personnel, Input, and Course Processing spreadsheets least useful.

The Manual Only group generally rated the spreadsheets the same as the Manual and Charts group. These subjects were more dispersed in their ratings, partially due to one subject. The subject rated all the spreadsheets a 6 (indicating dissatisfaction). The subject provided no additional written or verbal comments to explain the ratings. Only four 5's were given by all the rest of the subjects (no one else gave any 6's). In addition to this subject's ratings, there were two spreadsheet ratings with a gap in the ratings. These two spreadsheets were the two concerned with classroom maintenance. Some training managers said classroom maintenance was the responsibility of the resource managers, not the training manager (see Appendix J). Four of the spreadsheets had a median rating of one, five had a median rating of two, and two had a median rating of three. The modal rating was one for ten spreadsheets, the ratings of two and three tied for the mode on the Personnel spreadsheet.

The Manual and Charts group showed the same two patterns in rating the spreadsheets as did the Manual-only group. Only on the Input Spreadsheet was there a gap in the ratings. Five spreadsheets had a median of one and the other six had a median rating of two. The modal rating was one for ten spreadsheets (tied with two on the Classroom Standards, Maintenance Spreadsheet). The Course Processing Spreadsheet had a modal rating of two.

COMMENTS ON THE SPREADSHEETS RESULTS

Appendix I contains the results of the second evaluation form. Fifty two per cent of the comments were "None," another ten per cent of the comments were "OK." There was at least one suggested change for each spreadsheet. Also, there was at least one "Good" for each spreadsheet. Most of the suggestions were to extend the CCET to accommodate the full range of technical training courses. The positive nature of the comments provides anecdotal support for hypothesis six.

QUIS RESULTS

Appendix B contains a copy of QUIS. Questions rated by at least 66% of the subjects had a median rating of seven or higher, except two: overall reactions to the system (terrible...wonderful, 6.5) and characters on the computer screen (hard to read...easy to read, 6). Error and help messages were the ones most often rated n/a. The system failures question had the lowest median rating. That rating was a two. However, this question has a scale reversal (seldom...frequent). The modal rating was generally within one rating unit of the median. These results also provide evidence to support hypothesis six.

The ratings tended to be consistent within the question groupings provided. The percent n/a values varied within the groupings. For example, within the question grouping about highlighting, the values ranged from 11.2 to 44.4. In the relationship of terminology grouping, the range was from 0 to 33.3. In the reliability of the system grouping, the range was from 22.2 to 77.8.

DISCUSSION

This section consists of four parts. The first part discusses the performance evaluation. The second part discusses the ratings of the spreadsheets. The third part discusses the comments on the spreadsheets. The fourth part discusses the QUIS evaluation.

PERFORMANCE RESULTS

The groups did not perform significantly differently across the performance variables. Therefore, hypotheses one, two, and three of the second evaluation were not supported. Four reasons could account for this. First, the performance aid may not have aided performance on the tasks. Second, the small number of subjects made showing any differences difficult. Third, the small number of problems worked and time spent on the system also limited the amount of data collected. Fourth, the subjects in the Manual and Charts group did not regularly use the charts. So, the groups were not very different in how they worked the problems. All of these reasons provide explanations for the failure to find differences between the groups on any of the three variables tested.

Another explanation could have been the power of the test. Keppel (1982) described a method to test the power of a statistical test for a sample size and estimate of variance. This was used on the total session time data to check the power associated with the test results. The power test returned a value of 0.8 for the power of the test given the data and number of subjects. This value from the power test indicated the size of the sample was adequate to discern any statistically significant differences between the groups on this variable.

For the number of spreadsheets used, there was one subject in the Manual and Charts group that used 25 spreadsheets across the tasks. The next highest number used in either group was 16. Given the small number of subjects and the size of the difference between the subject that used 25 and the rest of the subjects, the results for one subject could overshadow any differences between the groups on the number of spreadsheets used.

The small number of errors may point out both the simplicity of the system and the fact that the tasks selected for the subjects to do did not provide a good test of the differences between the groups. The large number of errors in answering one question could indicate a problem with the question or the prototype. The number of errors when the Classroom Standards, Acquisition Spreadsheet was used instead of the Classroom Acquisition Spreadsheet implies the subjects were having difficulty differentiating between these spreadsheets.

The results supported hypothesis four. The subjects did refer to significantly fewer charts in the second scenario than in the first. This could mean the CCET was simple enough to understand that they did not need to refer to the charts after the first scenario. An alternate explanation is because the charts were physically separate from the CCET, the subjects viewed them as a distraction from the task.

SPREADSHEET RATING RESULTS

The results did not support the fifth hypothesis of the Second Evaluation. The groups did not rate the spreadsheets significantly differently. The main difference was in one subject from one group. This subject rated all the spreadsheets _6_, no other subject rated any spreadsheet less than a _5_.

Except for that subject, most of the subjects rated the spreadsheets favorably. The ratings were highest for the summary spreadsheets (Comparison and Life-Cycle Cost Estimate Spreadsheets) and lowest for the background spreadsheets (Classroom Standards, Acquisition and Classroom Standards, Maintenance Spreadsheets). The ratings for the other spreadsheets fell between those two categories. The lack of high ratings for the spreadsheets and the consistent low ratings from one subject could indicate the prototype is an interesting addition, but not essential to their job. Another explanation for the results comes from the length of time the users had with the system. There may have not been enough time for them to form strong opinions on most of the spreadsheets.

SPREADSHEET COMMENT RESULTS

The results from the comments on the spreadsheets also indicated support for the system. This supports hypothesis six. The comments paralleled the ratings in the number of positive comments. The suggested changes to the system mostly involved extending it beyond resident technical training and into other types of training. The small number of suggested changes could have come from four different reasons. First, the subjects were nearing the end of the session and needed to get on with other tasks. Second, the subjects only had a very limited time to use the prototype and did not have time to think about what it needed to be better. Third, the prototype did not impact their job very much, so, they did not review it in detail. Fourth, the prototype only needed minor changes they did not find in the short period of usage they had. Based on their oral reaction to the system the reason is probably a combination of the four, with a different combination for each subject.

QUIS

Two different types of information were gathered from the QUIS results. First, the median and modal ranking for the questions was obtained. Second, the per cent of n/a (not applicable) ratings for each question was obtained. The first result measured the users' satisfaction with the prototype on each question. The second result showed those parts of QUIS the subjects felt did not apply to the system they were evaluating. The second result also was used with the median and modal rating data to see if the ratings were consistent within the questionnaires groupings.

Overall, the users rated the prototype highly. This supports hypothesis six. Except for the 2 on the system failure question all the ratings were in the middle or upper third of the rating scale regardless of the per cent n/a.

All the subjects rated at least two questions n/a. At least half of the subjects rated almost half the questions n/a. Certain questions really did not apply to the system. There were no error messages, system failures, expert features, on-line help utility, or tutorials. Some subjects rated these questions anyway. Half or more of the subjects did not rate two questions, supplemental materials and reference materials, that did apply.

CONCLUSIONS

The subjects with the detailed flow charts did no better than the other subjects in the number of errors made and the number of spreadsheets used. Therefore, hypotheses one and two were not supported. The time to complete the tasks was not significantly different between the two groups. This fails to support hypothesis three even though hypothesis four was supported. Within the Manual-Only group the subjects tended to use only the Navigation page and not the rest of manual. Based on these results, the charts provided no additional benefit to the user. Two cautions go with that statement. First, the number of subjects was small. Second, each subject used the prototype for a short time.

The ratings and comments on the spreadsheets were not significantly different between the two groups. This fails to support the fifth hypothesis. The results did support the usefulness of the prototype (hypothesis six) and for further development and evaluation. The users liked the prototype, but need more time with it.

The responses to QUIS also support the prototype. This supports hypothesis six. Comments about QUIS show some of the terms used were not known by all the subjects. Also, sections of QUIS not related to the evaluation effort should be deleted from the questionnaire. This lowers the time requirement for the subjects to complete the questionnaire and they do not feel they have wasted time on irrelevant questions.

GENERAL DISCUSSION

The results from this study suggest the configuration of the system the users want, but lack enough completeness to give it a final form. For example, the interviews done before developing the prototypes revealed a requirement for at least two different output formats. The first format broke the costs down according to the instructional systems step (development, operations, and maintenance) they occurred in. The second format broke the costs down according to the Air Force cost accounting code they would fall under. The prototype versions of CCET only used the first format. None of the subjects cited a need for another format in either of the two evaluations. At least two different reasons could account for this. One, the other format is not necessary. Two, the other format is necessary, but the structure of the evaluation did not encourage the subject to review this requirement. A review of the reports the training managers use and generate supported the need for the second report format. Also, the scenarios and questionnaires did not encourage the subject to make judgements about the completeness of CCET to meet all of their requirements. Instead, they focused the subject on the adequacy of the prototype as a partial solution to the problem of estimating training course costs.

Finally, given the short period of evaluation, the subjects did not have the time to reflect at their leisure on their total information requirements. The focused nature and short duration of the evaluation supports the need for a longer-term evaluation of a decision support tool than was done in this study.

The methods used in this study produced specific information about the training managers' requirements for a cost estimating tool. The methods also provided specific guidance for the design, development, and evaluation of CCET. However, extending these results to other situations, subject groups, and decision tools requires reviewing the weaknesses of the overall experimental design. Earlier the two weaknesses brought up by Isaac and Michael (1985) were discussed. This study used a specialized decision tool and a small subject group to look at the various design, development, and evaluation tools. The representativeness mentioned earlier is especially relevant when these results are applied to other settings, subjects, or decision tools. So, extending these results beyond this study into other situations requires an evaluation of the applicability of the methods in the new setting.

Another weakness, subjective bias, is not as much a weakness of generalization beyond this study as it is a weakness within the study. The small number of subjects used limits the power of statistical tests and the amount of information reported on the open-ended questionnaires. One way to deal with this problem is to conduct a more thorough and long-term evaluation of the next version of the prototype.

CONCLUSIONS

The main conclusion of this study relates to the research question: was an iterative design process and rapid prototyping effective in designing decision support tools. The answer was a qualified yes. The number of subjects and the length of time they were available restricted the amount of information gathered from them. The subjects preferred using the prototype to watching the demonstration after a couple of minutes of use. But, the overall conclusion was the methods can be applied to field settings. One of the crucial decisions is selecting the focus of the application, design guidance or scientific proof. The amount of information necessary to guide a design effort and the amount of information necessary to draw scientific conclusions about the results of the evaluations may not be the same.

SUGGESTIONS FOR FURTHER WORK

The Suggestions for Further Work chapter consists of four sections. The first section contains suggestions on methodology. The second section contains suggestions on constructing an operational system. The third section contains suggestions on evaluating decision support tools. The fourth section contains suggestions for using QUIS.

METHODOLOGY

The different methods of using prototypes as evaluation tools needs further research. The advantages and limitations of each method would provide the designer with important information to guide the selection of a method. Clearly, to get performance information, the user must have access to the system. Also, it may be appropriate at the earlier stages of the design process to have the designer demonstrate the system and act as a human interface. More work needs to be done to place the various methods of using prototypes in the appropriate stages of the design process.

CONVERTING THE PROTOTYPE TO AN OPERATIONAL SYSTEM

One of the most obvious areas for follow-on work is in converting the prototype to an operational system. One enhancement of CCET the training managers asked for was the ability to input all their courses and then go back and change them as necessary. This could be done in the spreadsheet environment by developing a spreadsheet for each course. However, the tool would only store the results on the Compare Spreadsheet from different course configurations, and then only for three different versions. An alternative is to convert the system to a data base and report writer environment. The conversion must not change the ease of use and speed of result features of the prototype. More than any thing else, these two features were the most frequently praised by the training managers.

Converting to a data base environment would allow the training managers to make the long-term comparisons they wanted. Also, converting to a data base environment would encourage longitudinal validation of the variables and values used in the tool by simplifying keeping track of usage. A third advantage is ease of maintaining the system within a user location. If everyone uses a data base system, they only need to manipulate the data (course input data) records, there is only one copy of the formulas and standards to maintain. In the spreadsheet environment, each course is on one or more spreadsheets, so updating and maintaining the tool is more difficult.

Appendices I and J contain the comments from the subjects in both evaluations. Their comments contain many suggested improvements to the system. Also, the suggestion was made during the preliminary interviews that the output format needs to support the federal budget code for the type of money involved. For example, military trainees are paid out of the operations and maintenance military pay code. And it seems that each iteration raises new questions, points out different deficiencies, and brings up new uses for the prototype. So, a configuration management plan should be devised assigning the responsibility for maintaining and upgrading the system if the decision is made to field it.

EVALUATION

A weakness of this evaluation was the short time each subject had to use the system. The true value and evaluation of the system can only come after a much longer period of use.

Also, this tool was designed to help answer operational questions. The evaluation used questions similar to those found in the operational environment and used a similar setting. The emphasis is on "similar," the questions and the settings were not the actual working scenarios of the training managers. In addition to evaluating the tool over a longer time, it needs to be evaluated in the real environment with real questions.

QUIS

QUIS's comprehensiveness is its greatest strength and also its greatest weakness. The questionnaire was used in its entirety to see how the subjects responded to questions that may or may not have applied to the experiment. Clearly, the questionnaire could provide a wealth of data on a total system (hardware and software). For a software system like CCET too many sections were unnecessary.

The subjects also had problems with the terminology used in the questionnaire. Terms like "reverse video" were familiar only to the most computer literate. This introduces a frustration factor for the subject.

The combination of these problems shows a need to consider carefully using QUIS to gather data on user satisfaction. It may be necessary to break it down to the relevant sections and modify the language (or supplement with definitions) to make it appropriate for the user population. Breaking it down would entail revalidating the questionnaire, and providing the definitions could add to the time it takes to complete the questionnaire. Also, the definitions need to be checked with the users to see if they understand the definitions.

REFERENCES

Advanced Technology. (1984). First technical report, volumes 1, 2 ,3 (F41689-84-C-0012). Reston, VA: Advanced Technology.

Andersson, B. and Nilsson, S. (1964). Studies in the reliability and validity of the critical incident technique. Journal of Applied Psychology, 48, 398-403.

Anthony, R.N. (1965). Planning and control systems: a framework for analysis. Boston: Division of Research, Harvard Business School.

Banks, J. and Fabrycky, W.J. (1987). Procurement and inventory system analysis. New York: Prentice-Hall.

Braby, R., Henry, J.H., Parrish, W.F. Jr., and Swope, W.M. (1975). A technique for choosing cost-effective instructional delivery systems. Orlando, Fl.: Naval Training Equipment Center.

Carroll, J.M. (1982). The adventure of getting to know a computer. IEEE Computer, 15(11), 49-58.

Chin, J.P., Norman, K.L., and Shneiderman, B.(1987). Subjective user evaluation of CF Pascal programming tools. Human/Computer Interaction Laboratory, Center for Automation Research, University of Maryland, College Park, Md.

Defense Science Board (1972). Summer study on training and training technology. Washington, DC: Office of the Under Secretary of Defense, Research and Engineering.

Dick, W. and Carey, L. (1978). The systematic design of instruction. Dallas, Tx: Scott, Foresman.

del Galdo, E.M., Williges, R,C., Williges, B.H., and Wixon, D.R. (1986). An Evaluation of critical incidents for software design. In Proceedings of the Human Factors Society 30th Annual Meeting. Santa Monica, Ca: Human Factors Society.

Eagle Technology (1987). Evaluation of the acquisition process. Orlando, Fl: Eagle Technology.

Edney, P.J. (1972). A systems analysis of training. London: Pittman.

Edwards, A.L. (1957). Techniques of attitude scale construction. New York: Appleton-Century-Crofts.

Elmaghraby, S.E. (1966). The design of production systems. New York: Reinhold.

Fitts, P.M. and Jones, R.E. (1947). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. In H.W. Sinaiko (Ed.), Selected papers on human factors in the design and use of control systems. New York:Dover.

Goldstein, I.L. (1986). Training in organizations: needs assessment, development, and evaluation. Monterey, Ca: Brooks/Cole.

Gorry, G.A., and Scott Morton, M.S. (1971). A framework for management information systems. Sloan Management Review, 13 (1), 55-70.

Greene, J.H. (1974). Production and inventory control systems and decisions. Homewood, Il.: Irwin.

Hawley, J.K. and Frederickson, E.W. (1983). The training development decision support system: an aid to the design of cost-effective training programs. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 397-401). Santa Monica, CA: Human Factors Society.

HAY Systems (1986). A spreadsheet implementation of the TAEG report no. 16 cost model. Orlando, Fl: HAY systems.

Hendricks, D.K., Kilduff, P., Brooks, P., Marshak, R., and Doyle, B. (1982). Human engineering guidelines for management information displays. Human Engineering Laboratory, Aberdeen Proving Ground.

Hollander, M. and Wolfe, D.A. (1973). Nonparametric statistical methods. (pp 138-158) New York: Wiley

Isaac, S. and Michael, W.B. (1985). Handbook in research and evaluation. San Diego, Ca:EdITS.

Keppel, G. (1982). Design and analysis: a researcher's handbook, second edition. Englewood Cliffs, N.J.:Prentice-Hall.

Knapp M.I. and Orlansky, J. (1983). A cost element structure for defense training. Alexandria, Va.: Institute for Defense Analysis.

Kurstedt, H.A. Jr. (1986). Instructors handouts for information systems analysis and design. Unpublished course materials for IEOR 5410, Fall 1986, Virginia Polytechnic and State University, Blacksburg, Va.

Murdick, R.G. and Munson, J.C. (1986). MIS concepts and design. Englewood Cliffs, NJ: Prentice-Hall.

Norman, K.L., Chin, J.P., and Shneiderman, B. (1987). User evaluation of interactive work stations. Human/Computer Interaction Laboratory, Center for Automation Research, University of Maryland, College Park, Md.

Nunnally, J.C. (1967). Psychometric theory. New York: McGraw-Hill.

Orlansky, J. (1985). The cost-effectiveness of military training. In Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training. Brussels, Belgium: NATO Headquarters, Defense Research Group on the Defense Applications of Operations Research.

Orlansky, J. and String, J. (1979). Cost-effectiveness of computer-based instruction in military training. Washington, DC: Institute for Defense Analysis.

Schmidt, J.W. (undated). Model validation. Unpublished course materials for IEOR 5190, Virginia Polytechnic Institute and State University, Blacksburg, Va.

Shneiderman, B. (1987). Designing the user interface. Reading, Mass: Addison-Wesley.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. (pp 166-172) New York: McGraw-Hill.

Simon, H.A. (1960). The new science of management decision. New York: Harper and Row.

Singhal, A. (1986). An evaluation methodology to ensure the success of decision support tools. Unpublished masters thesis, Virginia Polytechnic Institute and State University, Blacksburg, Va.

Smith, S.L. (1986). Standards vs. guidelines for designing user interface software. Behavioral Information Technology, 5, 47-61.

Smith, S.L. and Mosier, J.N. (1986). Guidelines for designing user interface software. Technical Report ESD-TR-86-278. Bedford, Ma.: USAF Electronic Systems Division.

Sol, H.G. (1983). Prototyping: a methodological assessment. In R. Budde, K. Kuhlencamp, L. Mathiassen, and H. Zullighoven (Eds.), Approaches to prototyping (pp 368-382). New York: Springer-Verlag.

Sticha, P.J., Knerr, C.M., and Goldberg, S.L. (1984). Application of simulation and modeling to army training management. In Proceedings of the Human Factors Society 28th Annual Meeting, (pp. 1023-1027). Santa Monica, CA: Human Factors Society.

Williges, R.C., Williges, B.H., and Elkerton, J. (1987). Software interface design. In G. Salvendy (Ed.) Handbook of human factors. New York: Wiley.

APPENDIX A

Example Bipolar Questionnaire

QUESTIONNAIRE FOR RATING THE MODEL REPRESENTATIONS

Please rate the model representation on how useful it is in:

1. Helping you understand the Cost Estimation System

very useful						hinders understanding
1	2	3	4	5	6	7

2. Effect your accuracy/ability to find information in the Cost Estimation System

very useful						hinders accuracy
1	2	3	4	5	6	7

3. Effect your time to solve cost estimation problems with the Cost Estimation System

very useful						hinders solution
1	2	3	4	5	6	7

APPENDIX B

Questionnaire of User Interface Satisfaction (QUIS)

Last four digits of your student identification number ____________
Age _______

Sex Male Female

PART 1: Type of System to be Rated

1. Name of software __________ Name of hardware __________

2. Length of time you have worked on this system

less than 1 hour	6 months to less than 1 year
1 hours to less than 1 day	1 year to less than 2 years
1 day to less than 1 week	2 years to less than 3 years
1 week to less than 1 month	3 years of more
1 month to less than 6 months	

3. Average usage per week

less than one hour 4 to less than 10
one to less than 4 hours over 10 hours

PART 2: Past Experience

1. How many different types of computer systems (e.g., main frames and personal computers have you worked with?

none	3-4
1	5-6
2	more than 6

2. Of the following devices software and systems check those that you have personally used and are familiar with

keyboard	text editor	color monitor
numeric key pad	word processor	time-share system
mouse	file manager	personal computer
light pen	electronic spreadsheet	lap computer
touch screen	electronic mail	computer magazines
track ball	computer games	computer user's group
joy stick	video games	floppy disks

PART 3: User Evaluation of an Interactive Computer System

Please circle the numbers which most appropriately reflect your impressions about using this computer system. Not applicable = NA. Please add your written comments below the corresponding item.

Overall reactions to the system

terrible wonderful
1 2 3 4 5 6 7 8 9 NA

frustrating satisfying
1 2 3 4 5 6 7 8 9 NA

dull stimulating
1 2 3 4 5 6 7 8 9 NA

difficult easy
1 2 3 4 5 6 7 8 9 NA

inadequate power adequate power
1 2 3 4 5 6 7 8 9 NA

rigid flexible
1 2 3 4 5 6 7 8 9 NA

PART A SCREEN

1. Characters on the computer screen — hard to read easy to read
1 2 3 4 5 6 7 8 9 NA

1.1 Image of characters — fuzzy sharp
1 2 3 4 5 6 7 8 9 NA

1.2 Character shapes (fonts) — barely legible very legible
1 2 3 4 5 6 7 8 9 NA

2. Highlighting on the screen makes task easier — not at all very much
1 2 3 4 5 6 7 8 9 NA

2.1 User of reverse video — unhelpful helpful
1 2 3 4 5 6 7 8 9 NA

2.2 Use of blinking — unhelpful helpful
1 2 3 4 5 6 7 8 9 NA

3. Screen layouts make tasks easier		never always 1 2 3 4 5 6 7 8 9 NA
3.1 Amount of information that can be displayed screen		inadequate adequate 1 2 3 4 5 6 7 8 9 NA
3.2 Arrangement of information on screen		illogical logical 1 2 3 4 5 6 7 8 9 NA
4. Sequence of screens		confusing clear 1 2 3 4 5 6 7 8 9 NA
4.1 Next screen in a sequence	unpredictable predictable	1 2 3 4 5 6 7 8 9 NA
4.2 Going back to the previous screen		impossible easy 1 2 3 4 5 6 7 8 9 NA
4.3 Beginning middle and end of	confusing clearly marked	1 2 3 4 5 6 7 8 9 NA

PART B TERMINOLOGY AND SYSTEM INFORMATION

5. Use of terms throughout system	inconsistent consistent	1 2 3 4 5 6 7 8 9 NA
5.1 Task terms	inconsistent consistent	1 2 3 4 5 6 7 8 9 NA
5.2 Computer terms	inconsistent consistent	1 2 3 4 5 6 7 8 9 NA
6. Terminology relates to the work you are doing		unrelated related 1 2 3 4 5 6 7 8 9 NA
6.1 Computer terminology is used	too frequently appropriately	1 2 3 4 5 6 7 8 9 NA
6.2 Terms on the screen		ambiguous precise 1 2 3 4 5 6 7 8 9 NA

7. Messages which appear on screen — inconsistent / consistent
1 2 3 4 5 6 7 8 9 NA

7.1 Position of instructions on the screen — inconsistent / consistent
1 2 3 4 5 6 7 8 9 NA

8. Messages to the user — confusing / clear
1 2 3 4 5 6 7 8 9 NA

8.1 Instructions for commands of — confusing / clear
1 2 3 4 5 6 7 8 9 NA

8.2 Instructions for correcting errors — confusing / clear
1 2 3 4 5 6 7 8 9 NA

9. Computer keeps you informed about — never / always
1 2 3 4 5 6 7 8 9 NA

9.1 Performing an operation leads to a predictable result — never / always
1 2 3 4 5 6 7 8 9 NA

9.2 User can control amount of feedback — never / always
1 2 3 4 5 6 7 8 9 NA

10. Error messages — unhelpful / helpful
1 2 3 4 5 6 7 8 9 NA

10.1 Error messages clarify the — never / always
1 2 3 4 5 6 7 8 9 NA

10.2 Phrasing of error messages — unpleasant / pleasant
1 2 3 4 5 6 7 8 9 NA

PART C LEARNING

11. Learning to operate the system	difficult — easy	1 2 3 4 5 6 7 8 9 NA
11.1 Getting started	difficult — easy	1 2 3 4 5 6 7 8 9 NA
11.2 Learning advanced features	difficult — easy	1 2 3 4 5 6 7 8 9 NA
11.3 Time to learn to use the	slow — fast	1 2 3 4 5 6 7 8 9 NA
12. Exploration of features by trial and error	discouraged — encouraged	1 2 3 4 5 6 7 8 9 NA
12.1 Exploration of features	risky — safe	1 2 3 4 5 6 7 8 9 NA
12.2 Discovering new features	difficult — easy	1 2 3 4 5 6 7 8 9 NA
13. Remembering names and use of commands	difficult — easy	1 2 3 4 5 6 7 8 9 NA
13.1 Remembering specific rules about entering commands	difficult — easy	1 2 3 4 5 6 7 8 9 NA
14. Tasks can be performed in a straight-forward manner	never — always	1 2 3 4 5 6 7 8 9 NA
14.1 Number of steps per task	too many — just right	1 2 3 4 5 6 7 8 9 NA
14.2 Steps to complete a task	rarely — always	1 2 3 4 5 6 7 8 9 NA
14.3 Completion of sequence of steps	unclear — clear	1 2 3 4 5 6 7 8 9 NA

15. Help messages on the screen	confusing clear 1 2 3 4 5 6 7 8 9 NA
15.1 Accessing help messages	difficult easy 1 2 3 4 5 6 7 8 9 NA
15.2 Content of help messages	confusing clear 1 2 3 4 5 6 7 8 9 NA
15.3 Amount of help	inadequate adequate 1 2 3 4 5 6 7 8 9 NA
16. Supplement reference materials	confusing clear 1 2 3 4 5 6 7 8 9 NA
16.1 Tutorials for beginners	confusing clear 1 2 3 4 5 6 7 8 9 NA
16.2 Reference manuals	confusing clear 1 2 3 4 5 6 7 8 9 NA

PART D SYSTEM CAPABILITIES

17. System speed	too slow fast enough 1 2 3 4 5 6 7 8 9 NA
17.1 Response time for most operations	too slow fast enough 1 2 3 4 5 6 7 8 9 NA
17.2 Rate information is displayed	too slow fast enough 1 2 3 4 5 6 7 8 9 NA

18. Reliability of the system		unreliable reliable 1 2 3 4 5 6 7 8 9 NA
18.1 Operations are	undependable	dependable 1 2 3 4 5 6 7 8 9 NA
18.2 System failures occur		seldom frequently 1 2 3 4 5 6 7 8 9 NA
18.3 System warns the user about potential problems		never always 1 2 3 4 5 6 7 8 9 NA
19. System tends to be		noisy quite 1 2 3 4 5 6 7 8 9 NA
19.1 Mechanical devices such as fans disks and printers		noisy quite 1 2 3 4 5 6 7 8 9 NA
19.2 Computer tones beeps clicks etc		annoying pleasant 1 2 3 4 5 6 7 8 9 NA
20. Correcting your mistakes		difficult easy 1 2 3 4 5 6 7 8 9 NA
20.1 Correcting typos or mistakes		complex simple 1 2 3 4 5 6 7 8 9 NA
20.2 Ability to undo operations		inadequate adequate 1 2 3 4 5 6 7 8 9 NA
21. The needs of both experienced and inexperienced users are taken		never always 1 2 3 4 5 6 7 8 9 NA
21.1 Novices can accomplish tasks knowing only a few commands	with difficulty	easily 1 2 3 4 5 6 7 8 9 NA
21.2 Experts can use features shortcuts	with difficulty	easily 1 2 3 4 5 6 7 8 9 NA

APPENDIX C

First Version of CCET Screens

Length of Course

Lngth (days)	in/out Proc
78	5

Total Days	Total Weeks
83	17

Total Trainees

TPR	# of Recycles
89	17

Total Trainees	Total Trng Days
106	12,614

of Classes/yr Based on Class Size

Max Class Size	# of Classes
12.00	9

#of Classes/yr Based on Facilities

Max # of Classes	# of Classes
3.00	9

# of Shifts	1
# of Classes/Shift	3
# of Instructors	3
# of Trainees/Class	12

Course Data Input Spreadsheet

TPR	89
Attrition Rate	10%
Max Class Size	12
Max # of Classes	3
Revision Rate	5%

Media Mix	Del Hrs	Classroom Configuration		
Lecture	24	CAI	2.00	
Film/Tape	0	IVD	2.00	
CAI I	0	0= no new facilities		2= networked
CAI II	234	1= single classroom		3= telecomm
CAI III	40			
IVD I	188	Cost of Actual Equipment Laboratories		
IVD II	40	Acquisition	Maintenance	
Actual Equip I	100	$250,000.00	$25,000.00	
Actual Equip II	0	200,000.00	2,000.00	
Actual Equip III	0	0.00	0.00	

Courseware Development Cost Estimate Spreadsheet

	Hours	Cost/hr Del	Cost
Lecture	24.0	980.00	23,520.00
Film/Tape	0.0	2,456.00	0.00
CAI I	0.0	5,900.00	0.00
CAI II	234.0	8,820.00	2,063,880.00
CAI III	40.0	10,290.00	411,600.00
IVD I	188.0	5,880.00	1,105,440.00
IVD II	40.0	7,056.00	282,240.00
Actual Equip I	100.0	490.00	49,000.00
Actual Equip II	0.0	490.00	0.00
Actual Equip III	0.0	490.00	0.00
Total	626.0	Total	$3,935,680.00

Courseware Maintenance Cost Estimate Spreadsheet

	Hours	Cost/hr Del	Cost
Lecture	1.2	490.00	588.00
Film/Tape	0.0	1,474.00	0.00
CAI I	0.0	1,770.00	0.00
CAI II	11.7	2,058.00	24,078.60
CAI III	2.0	2,352.00	4,704.00
IVD I	9.4	2,352.00	22,108.80
IVD II	2.0	2,822.00	5,644.00
Actual Equip I	5.0	490.00	2,450.00
Actual Equip II	0.0	490.00	0.00
Actual Equip III	0.0	490.00	0.00
Revision Rate	0.05	Total	$59,573.40

Classroom Acquisition Cost Estimate Spreadsheet

Classroom Type	Number	Cost/clsrm	Cost
Lecture	1	900.00	900.00
Film/Tape	0	1,200.00	0.00
CAI single	0	34,400.00	0.00
networked	2	37,400.00	74,800.00
telecomm	0	35,400.00	0.00
IVD single	0	55,400.00	0.00
networked	2	58,400.00	116,800.00
telecomm	0	56,400.00	0.00
Actual Equip I	1	250,000.00	250,000.00
Actual Equip II	0	200,000.00	0.00
Actual Equip III	0	0.00	0.00
		Total	442,500.00

Classroom Maintenance Cost Estimate Spreadsheet

Classroom Type	Number	Cost/Classroom	Cost
Lecture	1.00	900.00	900.00
Film/Tape	0.00	1,200.00	0.00
CAI single	0.00	4,250.00	0.00
networked	2.00	4,550.00	9,100.00
telecomm	0.00	5,250.00	0.00
IVD single	0.00	6,350.00	0.00
networked	2.00	6,650.00	0.00
telecomm	0.00	7,350.00	0.00
Actual Equip I	1.00	25,000.00	25,000.00
Actual Equip II	0.00	2,000.00	0.00
Actual Equip III	0.00	0.00	0.00
		Total	$35,000.00

Personnel Cost Estimate Spreadsheet

Personnel Type	Number	Unit Cost	Cost
Instructors	3	35,000.00	105,000.00
Trainee	12,614	100.00	1,261,400.00
		Total	$1,366,400.00

Course Life Cycle Cost Estimate Spreadsheet

	Fiscal Year 1	2	3	4	5	6
Development	3,935,680.00					
Maintenance		59,573.40	53,616.06	48,254.45	43,429.01	39,086.11
Personnel						
Classroom Acq	442,500.00	1,366,400.00	1,229,760.00	1,106,784.00	996,105.60	896,495.04
Classroom Maint		35,000.00	31,500.00	28,350.00	25,515.00	22,963.50
Total	4,378,180.00	1,460,973.40	1,460,973.40	1,183,388.45	1,065,049.61	958,544.65
Cumulative Total	4,378,180.00	5,839,153.40	7,300,126.80	8,483,515.25	9,548,564.86	10,507,109.51

APPENDIX D

Courseware Cost Estimation Spreadsheet

Courseware Development/Maintenance Cost Table

Labor Category	Labor Cost	Lect	SFT	Media Type CAI1	CAI2	CAI3	IVD1	IVD2
Prog Mgr	46.00	.10	.10	.10	.10	.10	.10	.10
Sr Analyst	46.00	0	0	.05	.05	.05	.05	.05
Jr Analyst	33.00	0	0	.10	.10	.10	.10	.10
Sr Instr Tech	36.00	0	0	.05	.05	.05	.05	.05
Jr Instr Tech	26.00	0	.05	.10	.10	.10	.10	.10
ISD Analyst	33.00	.10	.10	.10	.10	.10	.10	.10
Programmer	23.00	0	0	.20	.25	.25	.25	.25
Tech Editor	20.00	.10	.10	.10	.10	.10	.10	.10
Graphic Artist	19.00	.05	.05	0	0	0	0	0
Typist	12.00	.20	.20	0	0	0	0	0
Instructor	25.00	.25	.30	.10	.05	.05	.05	.05
SME	25.00	.20	.20	.10	.10	.10	.10	.10
Dev-hr/Del-hr		2	3	5	7	10	10	12
Cost/hr Dev		49	74	148	206	294	294	353
Courseware Maintenance Factor		1.0	1.0	0.6	0.5	0.4	0.4	0.4
Cost/hr Maintenance		49	74	89	103	118	118	141

APPENDIX E

First Evaluation Subject Materials

Participant's Informed Consent Form

This is an attempt to test and improve a tool for estimating couse life cycle costs. To do this you'll use the tool to answer various questions about the course in the profile. While you're answering the questions, please write down any problems or good features you see in the tool. After you've finished answering the questions you'll go back over your impressions and rate them on their importance in a useful cost estimating tool. I'm also going to making notes as you answer the questions. Between both of our notes we'll try to define what you need in a cost estimating tool. Then you'll look at 4 different representations of the tool and rate them on their usefulness in helping you use the tool.

As a participant in this experiment you have certain rights, explained below. The purpose of this document is to obtain your written consent to participate in this effort.

1. You have the right to discontinue your participation in the study at any time for any reason. If you decide to terminate the experiment, inform the researcher.

2. You have the right to inspect your data and withdraw it from the experiment if you feel that you should for any reason. In general, data are processed and analyzed after a subject has completed the experiment. At that time, all identification will be removed and the data treated with anonymity. Therefore, if you wish to withdraw your data, you must do so immediately after your participation is completed.

3. You have the right to be informed of the overall results of the experiment. If you wish to receive a synopsis of the results, include your address with your signature below. If after receiving the synopsis, you would like more indepth information please contact Virginia Tech's Human Factors Laboratory and a full report will be made available to you.

The principal investigator is Dr. Robert Williges. The researcher is Gary Macomber. Both of these people can be contacted at the following address and phone number:

Human Factors Laboratory
302 Whittemore Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA 24601
(703) 961-4602

If you have any questions about the experiment or your rights as a participant, please do not hesitate to ask. The researcher will do his best to answer them.

Your signature below indicates you have read and understand your rights as a participant (as stated above), and that you consent to participate.

____

Participants Signature

____

Witness' Signature

____

____

____

Print your name and address if you wish to receive a summary of the experimental results

Evaluation of Cost Estimation Tool

Using your notes and all you remember about the system, please write down all of the things you liked about the system. As you write each one down please rate its importance to you on a scale of 1 to 10 with 1 being absolutely critical and 10 being nice to have but not essential.

Now, please write down all the things you didn't like about the system or features you think it should have that it currently doesn't. As you write each one down please rate its importance to you on a scale of 1 to 10 with 1 being absolutely critical and 10 being nice to have but not essential.

QUESTIONNAIRE FOR RATING THE MODEL REPRESENTATIONS

Model 1 Model Outline

Please rate the model representation on how useful it is in:

1. Helping you understand the Cost Estimation System

very useful						hinders understanding
1	2	3	4	5	6	7

2. Effect your accuracy/ability to find information in the Cost Estimation System

very useful						hinders accuracy
1	2	3	4	5	6	7

3. Effect your time to solve cost estimation problems with the Cost Estimation System

very useful						hinders solution
1	2	3	4	5	6	7

If you've any suggestions on how I can improve the representation of the system to help you use it, please write them below.

Model Outline

1. **Input Spreadsheet** (goes to 2.)

2. **Course Processing Spreadsheet** (goes to 3.a.1, 3.a.2, 3.b.1, 3.b.2, 3.b.3)

3.a. Courseware Cost Estimates

- 3.a.1. **Courseware Development Cost Estimate Spreadsheet** (goes to 4.)
- 3.a.2. **Courseware Maintenance Cost Estimate Spreadsheet** (goes to 4.)

3.b. Operations Cost Estimates

- 3.b.1. **Personnel Cost Estimate Spreadsheet** (goes to 4.)
- 3.b.2. **Classroom Acquisition Cost Estimate Spreadsheet** (goes to 4.)
- 3.b.3. **Classroom Maintenance Cost Estimate Spreadsheet** (goes to 4.)

4. **Course Life Cycle Cost Estimate Spreadsheet**

QUESTIONNAIRE FOR RATING THE MODEL REPRESENTATIONS

Model 2 Detailed Model Outline

Please rate the model representation on how useful it is in:

1. Helping you understand the Cost Estimation System

very useful 1	2	3	4	5	6	hinders understanding 7

2. Effect your accuracy/ability to find information in the Cost Estimation System

very useful 1	2	3	4	5	6	hinders accuracy 7

3. Effect your time to solve cost estimation problems with the Cost Estimation System

very useful 1	2	3	4	5	6	hinders solution 7

If you've any suggestions on how I can improve the representation of the system to help you use it, please write them below.

Detailed Model Outline

1. Input Spreadsheet

Input

Information	Source
Course/Media hours mix	user
CAI Classroom Configuration	user
IVD Classroom Configuration	user
Equipment Laboratory Acquisition Costs	user
Equipment Laboratory Maintenance Costs	user
Trained Personnel Requirement	user
Maximum Class Size	user
Maximum Number of Classes/Shift	user

Output

Information	Destination Spreadsheet
Course/Media Hrs Mix	Course Processing
CAI Classroom Config	Course Processing
IVD Classroom Config	Course Processing
Equip Lab Acq Costs	Course Processing
Equip Lab Maint Costs	Course Processing
Trained Personnel Req	Course Processing
Max Class Size	Course Processing
Max Number of Classes/Shift	Course Processing

2. Course Processing Spreadsheet

Information	Source Spreadsheet
Course/Media hours mix	user
CAI Classroom Configuration	user
IVD Classroom Configuration	user
Equipment Laboratory Acquisition Costs	user
Equipment Laboratory Maintenance Costs	user
Trained Personnel Requirement	user
Maximum Class Size	user
Maximum Number of Classes/Shift	user

Output

Information	Destination Spreadsheet
Course Hours/Media Mix	Courseware Development
Courseware Maint Media/Hrs Mix	Courseware Maintenance
Number of Training Days	Personnel
Number of Instructors	Personnel
Number of Classes	Classroom Acquisition
Number of Classes	Classroom Maintenance

3. Courseware Cost Estimates

3.a.1. Courseware Development

Input

Information	Source Spreadsheet
Course Media/Hours Mix	Course Processing

Output

Information	Destination Spreadsheet
Tot Est Courseware Dev Cost	Course Life Cycle

3.a.2. Courseware Maintenance

Input

Information	Source Spreadsheet
Courseware Maint Media/Hrs Mix	Course Processing

Output

Information	Destination Spreadsheet
Tot Est Courseware Maint Cost	Course Life Cycle

3. Operations Cost Estimates

3.b.1. Personnel

Input

Information	Source Spreadsheet
Number of Training Days	Course Processing
Number of Instructors	Course Processing

Output

Information	Destination Spreadsheet
Tot Yearly Est Personnel Costs	Course Life Cycle

3.b.2. Classroom Acquisition

Input

Information	Source Spreadsheet
Number of Classes	Course Processing

Output

Information	Destination Spreadsheet
Tot Est Classroom Acq Costs	Course Life Cycle

3.b.3. Classroom Maintenance

Input

Information	Source Spreadsheet
Number of Classes	Course Processing

Output

Information	Destination Spreadsheet
Tot Est Classroom Maint Cost	Course Life Cycle

4. Life Cycle Summary

Input

Information	Source Spreadsheet
Tot Est Courseware Dev Cost	Courseware Dev
Tot Est Courseware Maint Cost	Courseware Maint
Tot Yearly Est Personnel Costs	Personnel
Tot Est Classroom Acq Cost	Classroom Acq
Tot Est Classroom Maint Cost	Classroom Maint

QUESTIONNAIRE FOR RATING THE MODEL REPRESENTATIONS

Model 3 Model Chart

Please rate the model representation on how useful it is in:

1. Helping you understand the Cost Estimation System

very useful						hinders understanding
1	2	3	4	5	6	7

2. Effect your accuracy/ability to find information in the Cost Estimation System

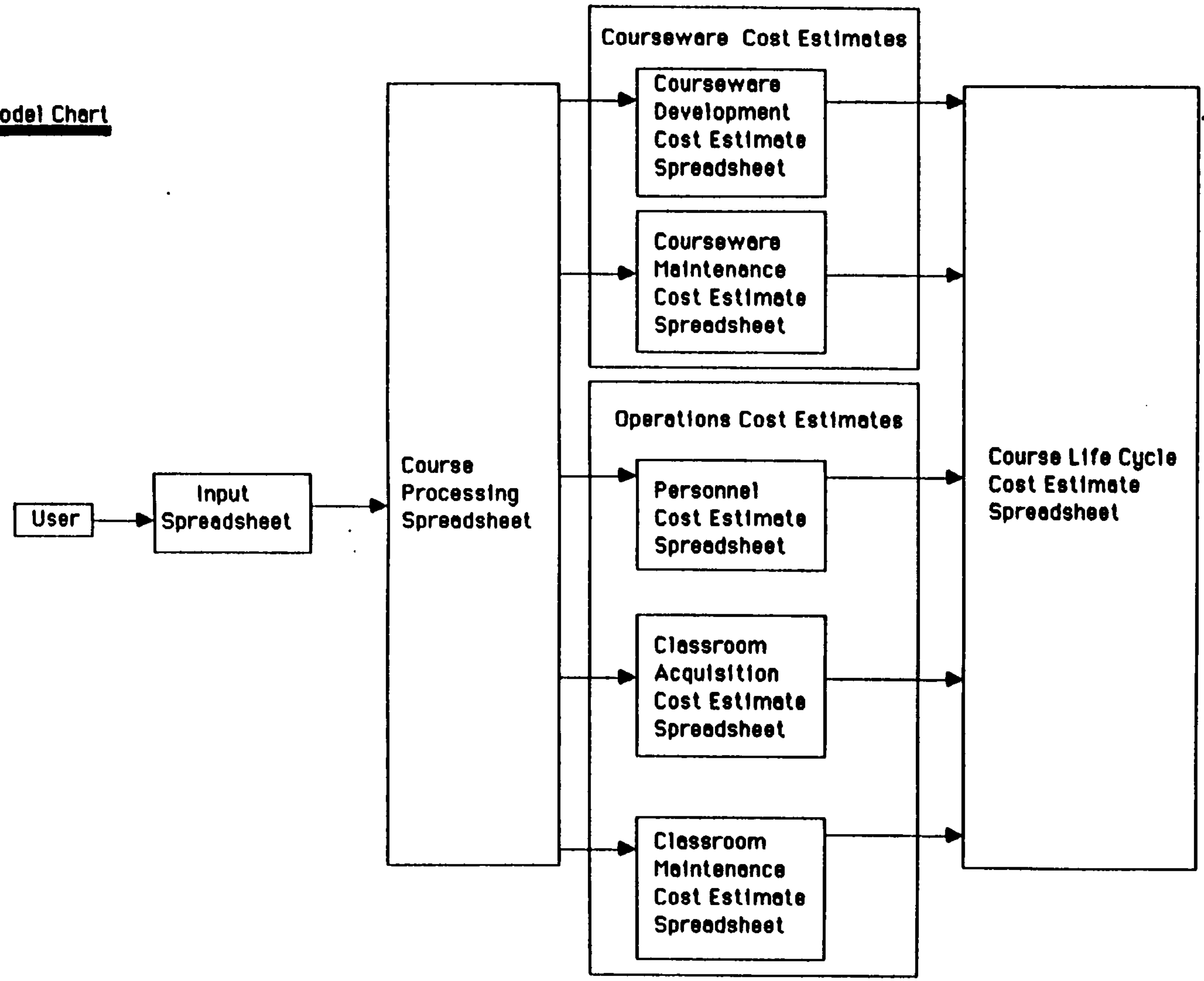
very useful						hinders accuracy
1	2	3	4	5	6	7

3. Effect your time to solve cost estimation problems with the Cost Estimation System

very useful						hinders solution
1	2	3	4	5	6	7

If you've any suggestions on how I can improve the representation of the system to help you use it, please write them below.

Model Chart


User
Input Spreadsheet
Course Processing Spreadsheet
Courseware Cost Estimates
Courseware Development Cost Estimate Spreadsheet
Courseware Maintenance Cost Estimate Spreadsheet
Operations Cost Estimates
Personnel Cost Estimate Spreadsheet
Classroom Acquisition Cost Estimate Spreadsheet
Classroom Maintenance Cost Estimate Spreadsheet
Course Life Cycle Cost Estimate Spreadsheet

QUESTIONNAIRE FOR RATING THE MODEL REPRESENTATIONS

Model 4 Detailed Model Chart

Please rate the model representation on how useful it is in:

1. Helping you understand the Cost Estimation System

very useful						hinders understanding
1	2	3	4	5	6	7

2. Effect your accuracy/ability to find information in the Cost Estimation System

very useful						hinders accuracy
1	2	3	4	5	6	7

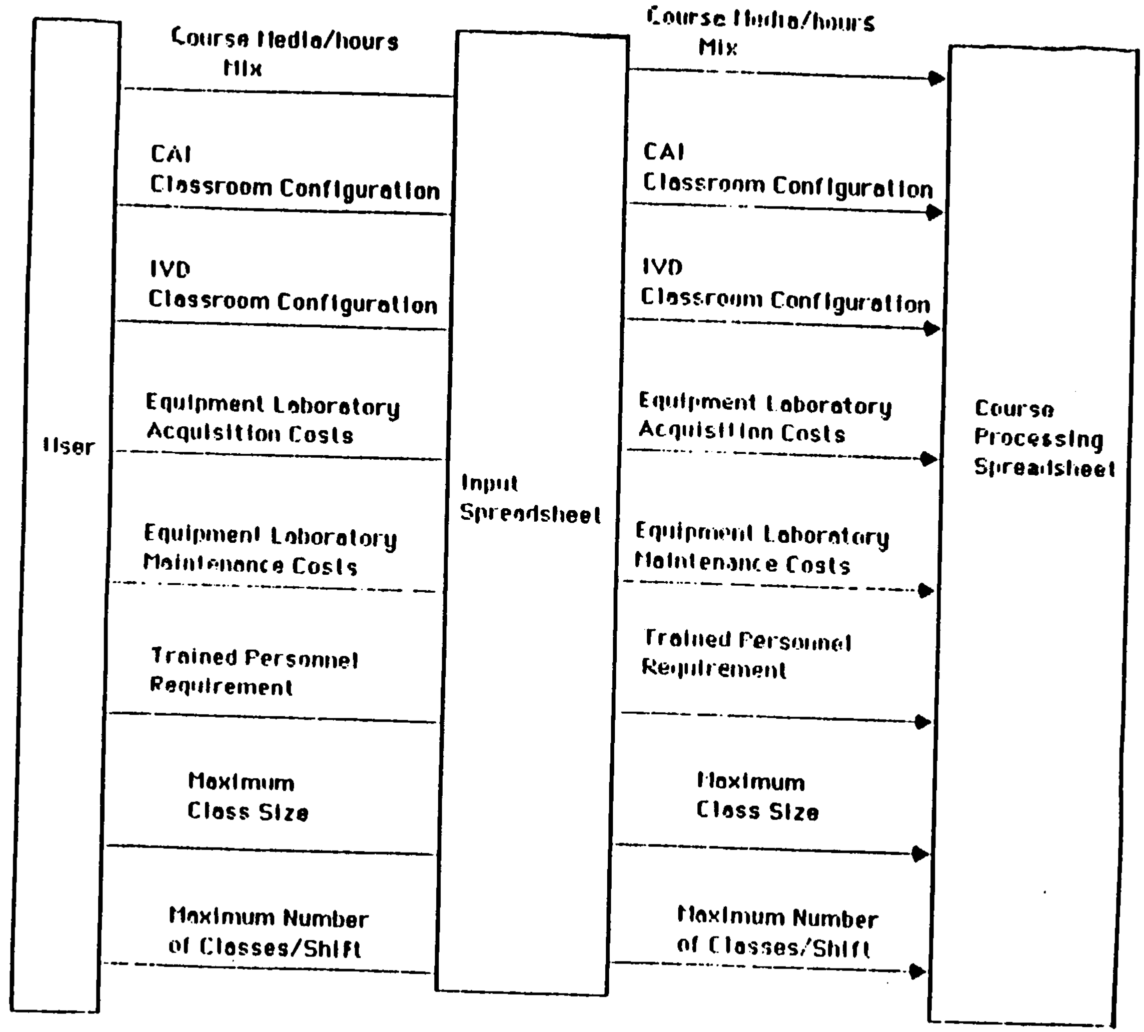
3. Effect your time to solve cost estimation problems with the Cost Estimation System

very useful						hinders solution
1	2	3	4	5	6	7

If you've any suggestions on how I can improve the representation of the system to help you use it, please write them below.

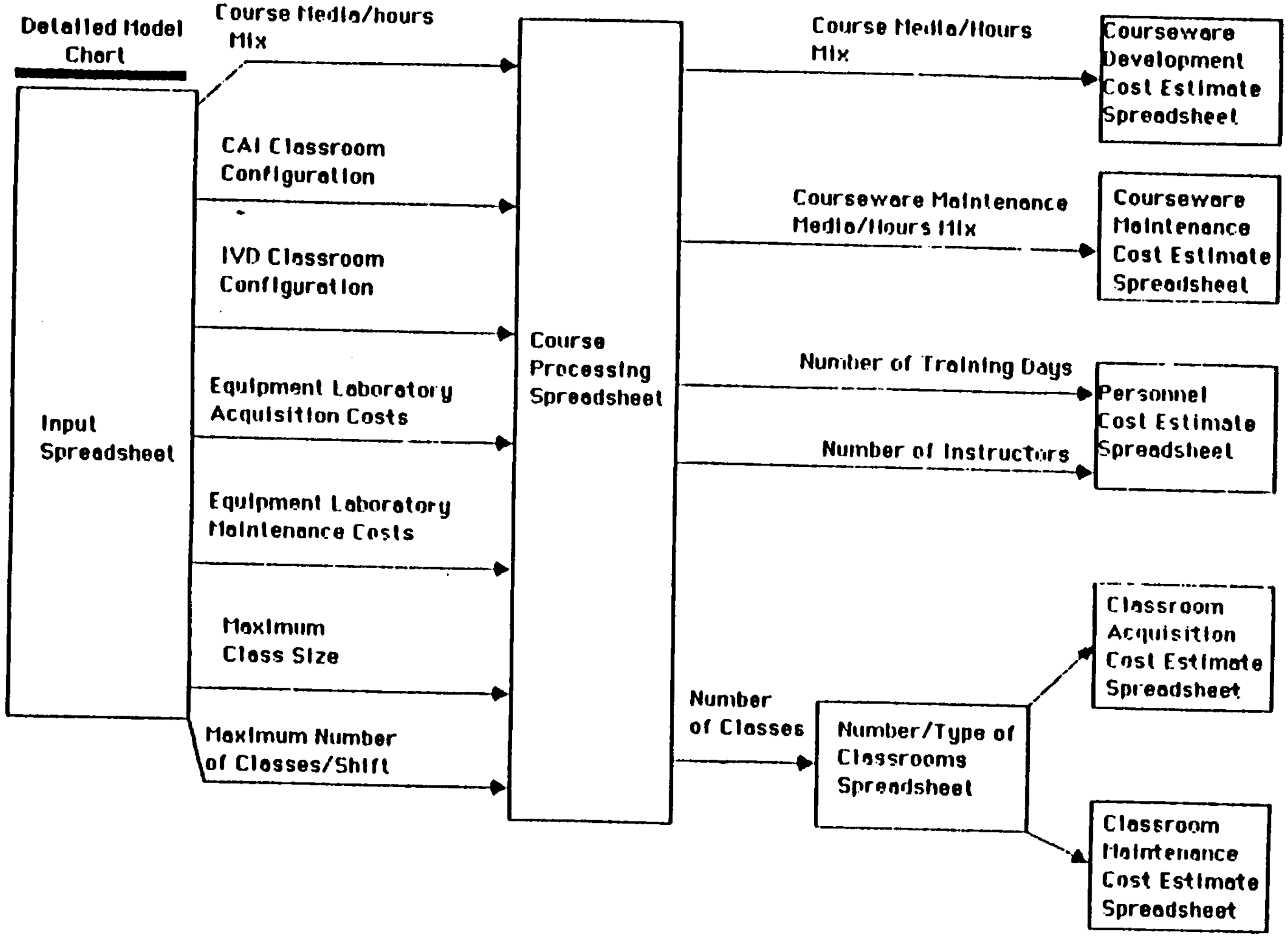
I

Detailed Model Chart



2

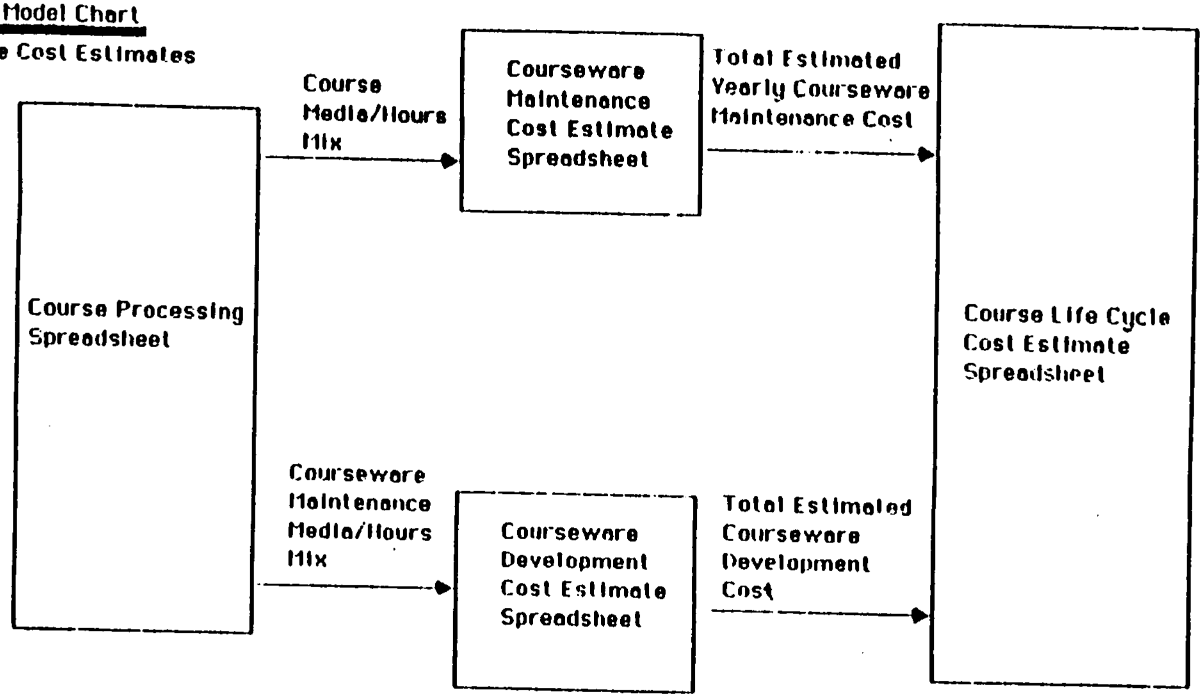
Detailed Model Chart



3

<u>Detailed Model Chart</u>

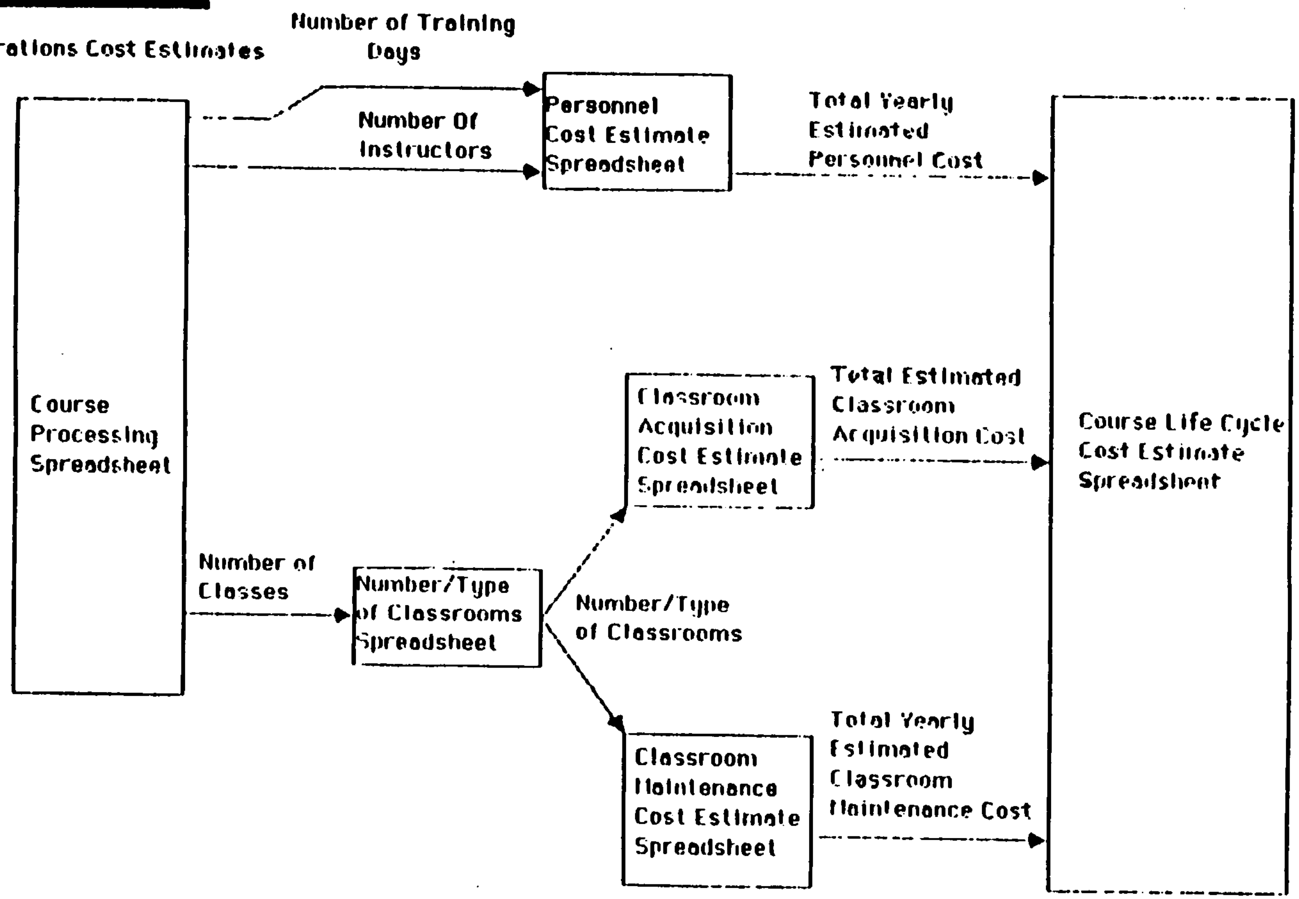
Courseware Cost Estimates



4

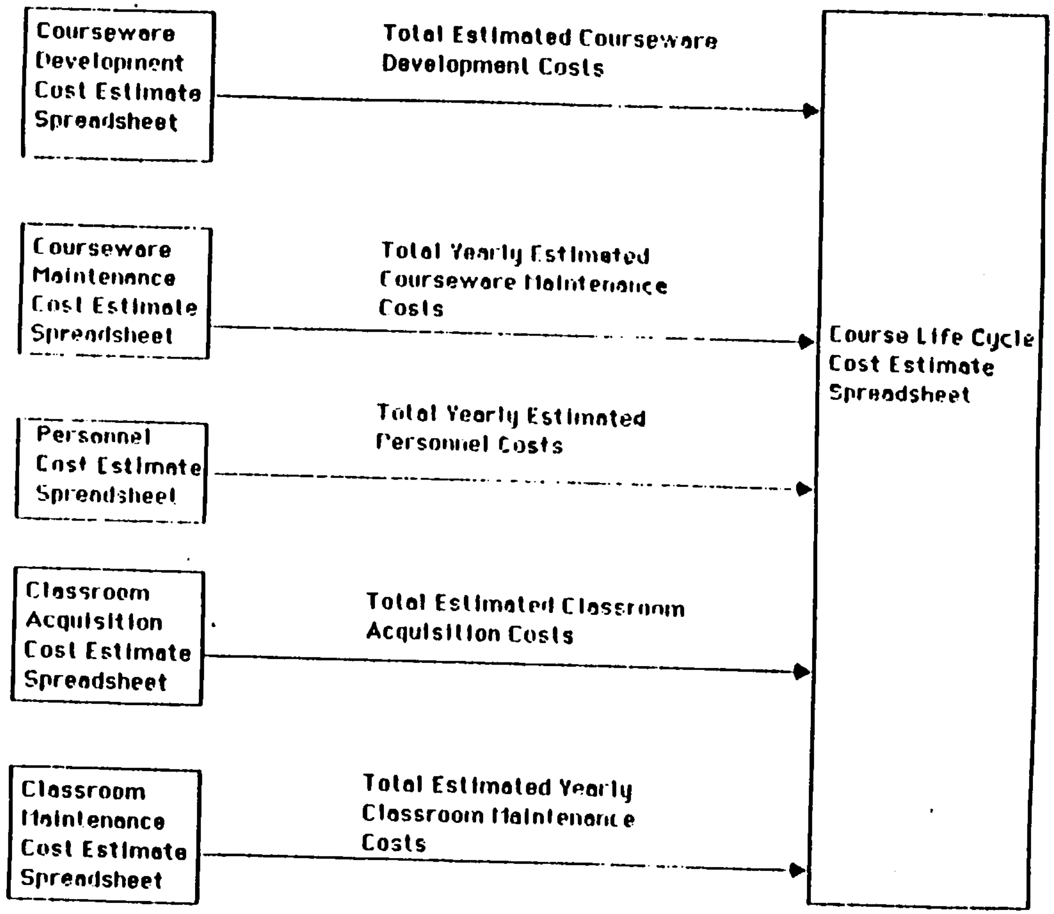
Detailed Model Chart

Operations Cost Estimates



5

Detailed Model Chart



APPENDIX F

First Evaluation Comments

Results of the First Evaluation of the Prototype

Using your notes and all you remember about the system, please write down all of the things you liked about the system. As you write each one down please rate its importance to you on a scale of 1 to 10 with 1 being absolutely critical and 10 being nice to have but not essential.

Rating	Comment
1	Quickly gives life cycle cost and changes in that final output as you change the input.
3	Certainly would serve as a quick reference tool for short term planning.
6	Production of course resource estimate mechanized.
8	Capability for snapshot course cost comparison.
	Ease of inputting data.
	Ability to quickly compare several possible options.
8	The straight forward method in which data is arranged and the ease in which that data can be changed.
8	Ease of operation.
9	Comprehensiveness.

Now, please write down all the things you didn't like about the system or features you think it should have that it currently doesn't. As you write each one down please rate its importance to you on a scale of 1 to 10 with 1 being absolutely critical and 10 being nice to have but not essential.

Rating	Comment
2	Should have means to account for multiple instructor requirement (MIR).
2	Doesn't figure in my multiple instructor requirements (MIR) or family groupings to the best of my recollection.
2	Should have means to show course development over a longer period than one year.
2	Accommodate unusual recycle rates in relation to attrition.

8 Ability to input actual costs for equipment in CAI classrooms - necessary for definitive costs.

8 Ability to input actual student man day costs, vice estimates.

8 Need more realistic dev/del time ratios. Industry standards reflect up to 400+ hours per 1 hour of IVD (some studies reflect as much as 800/1).

Needs to be able to provide cost data for all types of ATC courses including type 6.
Some factors that contribute to life cycle cost not in software: amount of washback time required for course, instructor qualification time/money.

In general should have more options for unique input: within categories already recognized and for unforeseen relevant data.

None

APPENDIX G

Second Version of CCET Screens

Course Data Input Spreadsheet

Course Model #	1	
Inflation Rate	5%	
# of Instructors	0	

TPR	50	
Attrition Rate	10%	
Max Class Size	12	
Max # of Classes	3	
Revision Rate	5%	

Course Configuration

	Del Hrs # of Instrs 1	2
Lecture	200	0
Film/Tape	140	0
CAI I	80	0
CAI II	100	0
CAI III	24	0
IVD I	40	0
IVD II	80	0
Equipment Lab 1	150	0
Equipment Lab 2	75	0

Classroom Configuration

CAI	2.00
IVD	2.00

0= no new facilities 2= networked
1= single classroom 3= telecomm

Cost of Equipment Laboratories

Acquisition	Maintenance
$250,000.00	$25,000.00
200,000.00	20,000.00

Comparison of Different Course Models

Cost Category	Model 1	Model 2	Model 3
Courseware			
Development	3,386,620.00	3,386,620.00	3,386,620.00
Maintenance	56,914.90	56,914.90	56,914.90
Operations			
Personnel	1,148,000.00	1,029,000.00	1,148,000.00
Classroom Acq	621,050.00	621,050.00	621,050.00
Classroom Mnt	64,505.00	46,505.00	46,505.00
Total Est L-C Cost	10,956,545.27	10,208,383.79	10,859,983.91
% of Model 1		93%	99%

Length of Course

Lngth(days)	in/out Proc
111	5

Total Days	Total Weeks
116	24

Total Trainees

TPR	# of Recycles
50	10

Total Trainees	Total Trng Days
60	10,080

# of Classes/yr Based on Class Size

Max Class Size	# of Classes
12.00	6

#of Classes/yr Based on Facilities

Max # of Classes	# of Classes
3.00	6

# of Shifts	1
# of Classes/Shift	3
# of Instructors	4
# of Trainees/Class	7

Number Of Classrooms Spreadsheet

Classroom Type	Del Hrs	% of Course	# of Clsrms
Lect/Film/Tape	340.00	0.38	1.00
CAI	204.00	0.23	2.00
IVD	120.00	0.13	1.00
Equipment Lab 1	150.00	0.17	1.00
Equipment Lab 2	75.00	0.08	1.00
Equipment Lab 3	0.00	0.00	0.00
Total	889.00	1.00	6.00

Courseware Development Cost Estimate Spreadsheet

	Hours	Cost/hr Del	Cost
Lecture	200.0	1.100.00	220,000.00
Film/Tape	140.0	2.800.00	392.000.00
CAI I	30.0	5.100.00	408.000.00
CAI II	100.0	7,400.00	740,000.00
CAI III	24.0	10,400.00	249,600.00
IVD I	40.0	7,410.00	296.400.00
IVD II	80.0	10,414.00	833.120.00
Actual Equip I	150.0	1.100.00	165.000.00
Actual Equip II	75.0	1.100.00	82.500.00
Actual Equip III	0.0	1,100.00	0.00
Total	889.0	Total	$3,386.620.00

Courseware Maintenance Cost Estimate Spreadsheet

	Hours	Cost/hr Del	Cost
Lecture	10.0	490.00	4,900.00
Film/Tape	7.0	1,474.00	10.318.00
CAI I	4.0	1,770.00	7,080.00
CAI II	5.0	2.058.00	10.290.00
CAI III	1.2	2.352.00	2.822.40
IVD I	2.0	2.352.00	4.704.00
IVD II	4.0	2.822.00	11.288.00
Actual Equip I	7.5	490.00	3.675.00
Actual Equip II	3.8	490.00	1,837.50
Actual Equip III	0.0	490.00	0.00
Revision Rate	0.05	Total	$56,914.90

Classroom Size and Equipment Spreadsheet

CLASSROOM Type	Size	EQUIPMENT Type	Number	Cost/Unit	Classroom Cost
Lect/Film/Tape	1,200	overhead	1	900.00	8,300.00
		film	1	600.00	
		tape	1	800.00	
CAI Single	1,200	terminals	21	1,500.00	42,500.00
		file server	1	5,000.00	
CAI Network (serves 4 clsrms)	1,200	terminals	21	1,500.00	43,750.00
		CPU	1	25,000.00	
CAI Telecomm (serves 4 clsrms)	1,200	terminals	21	1,500.00	43,125.00
		file server	1	5,000.00	
		telecomm	1	2,500.00	
IVD Single	1,200	terminals	21	1,500.00	74,000.00
		vdisk player	21	1,500.00	
		file server	1	5,000.00	
IVD Network (serves 4 clsrms)	1,200	terminals	21	1,500.00	75,250.00
		vdisk player	21	1,500.00	
		CPU	1	25,000.00	
IVD Telecomm (serves 4 clsrms)	1,200	terminals	21	1,500.00	74,625.00
		vdisk player	21	1,500.00	
		file server	1	5,000.00	
		telecomm	1	2,500.00	

Classroom Acquisition Cost Estimate Spreadsheet

Classroom Type	Number	Cost/clsrm	Cost
Lect/Film/Tape	1	8.300.00	8.300.00
CAI single	0	42.500.00	0.00
networked	2	43.750.00	87.500.00
telecomm	0	43.125.00	0.00
IVD single	0	74.000.00	0.00
networked	1	75.250.00	75.250.00
telecomm	0	74.625.00	0.00
Actual Equip I	1	250,000.00	250.000.00
Actual Equip II	1	200,000.00	200.000.00
Actual Equip III	0	0.00	0.00
		Total	621,050.00

Classroom Maintenance Cost Estimate Spreadsheet

Classroom Type	Number	Cost/Classroom	Cost
Lect/Film/Tape	1.00	1,430.00	1,430.00
CAI single	0.00	4.850.00	0.00
networked	2.00	4,975.00	9,950.00
telecomm	0.00	4,912.50	0.00
IVD single	0.00	8,000.00	0.00
networked	1.00	8,125.00	8,125.00
telecomm	0.00	8.062.50	0.00
Actual Equip I	1.00	25,000.00	25,000.00
Actual Equip II	1.00	2,000.00	2,000.00
Actual Equip III	0.00	0.00	0.00
		Total	$46,505.00

Course Life Cycle Cost Estimate Spreadsheet

	Fiscal Year 1	2	3	4	5	6
Courseware						
Development	3,386,620.00					
Maintenance		56,914.90	59,760.65	62,748.68	65,886.11	69,180.42
Operations						
Personnel		1,148,000.00	1,205,400.00	1,265,670.00	1,328,953.50	1,395,401.18
Classroom Acq	621,050.00					
Classroom Maint		64,505.00	67,730.25	71,116.76	74,672.60	78,406.23
Total	4,007,670.00	1,269,419.90	1,269,419.90	1,399,535.44	1,469,512.21	1,542,987.82
Cumulative Total	4,007,670.00	5,277,089.90	6,546,509.80	7,946,045.24	9,415,557.45	10,958,545.27

APPENDIX H

Revised Courseware Cost Estimation Spreadsheet

COURSEWARE DEVELOPMENT COST ESTIMATES PER HOUR DELIVERED

Adjustment by Comparative Contribution of Labor Types

Number of "Parts" by Labor Category

LABOR	LECT	SFT	CAI1	CAI2	CAI3	IVD1	IVD2	TOT
MGT	0.50	0.50	0.50	1.00	1.50	1.25	2.00	7.25
ID	1.00	1.25	1.00	2.00	3.00	3.00	4.00	15.25
PROG	0.00	0.00	1.00	2.00	4.00	1.50	3.00	11.50
VIDEO	0.00	1.00	0.25	0.75	1.00	3.00	4.00	10.00
MEDIA	0.00	0.00	0.75	1.25	2.00	2.00	4.00	10.00
WP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	7.00
SME	2.00	2.00	1.00	1.50	2.00	1.50	2.00	12.00
TOTAL	4.50	5.75	5.50	9.50	14.50	13.25	20.00	73.00

This section shows the the relative per cent by the Instructional Type and Labor Category

LABOR	LECT	SFT	CAI1	CAI2	CAI3	IVD1	IVD2	AVE
MGT	0.12	0.09	0.09	0.10	0.10	0.09	0.10	0.10
ID	0.22	0.22	0.18	0.21	0.21	0.23	0.20	0.21
PROG	0.00	0.00	0.18	0.21	0.28	0.11	0.15	0.13
VIDEO	0.00	0.17	0.05	0.08	0.07	0.23	0.20	0.11
MEDIA	0.00	0.00	0.14	0.13	0.14	0.15	0.20	0.11
WP	0.22	0.17	0.18	0.11	0.07	0.08	0.05	0.13
SME	0.44	0.35	0.18	0.16	0.13	0.11	0.10	0.21
TOTAL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

COURSEWARE DEVELOPMENT COST ESTIMATES PER HOUR DELIVERED

Number of development hours per hour delivered

	LECT	SFT	CAI1	CAI2	CAI3	IVD1	IVD2
DEV/DEL	40	100	175	250	350	250	350

Cost of development per hour delivered for each labor category and total development cost.

LABOR	LECT	SFT	CAI1	CAI2	CAI3	IVD1	IVD2	$/HR
MGT	204	400	731	1210	1665	1084	1610	46
ID	302	739	1081	1789	2462	1924	2380	34
PROG	0	0	890	1473	2703	792	1470	28
VIDEO	0	521	238	592	724	1698	2100	30
MEDIA	0	0	596	822	1206	943	1750	25
WP	133	260	477	394	362	283	262	15
SME	433	869	795	986	1206	707	875	25
TOTAL	1083	2789	4808	7266	10328	7431	10447	

The total cost for development is based on (the per cent by labor category for an instructional category) multiplied by (the number of development hours per hour delivered) multiplied by (the cost per hour for a labor category

APPENDIX I

Second Evaluation Materials

Session Checklist

Introduction
- Overview of Dissertation/what we're doing ____
- Informed Consent ____

Using the Tool
- Introduce the manual: Macros and Glossary ____
- Introduce the tasks: Charts and scenarios ____

	Running Time	Elapsed Time
Start data collection		
Start Scenario 1 ____	End data entry ____	
Answer Question 1	____	____
Spreadsheets used: ____		
Answer Question 2	____	____
Spreadsheets used: ____		
Answer Question 3	____	____
Spreadsheets used: ____		
Answer Question 4	____	____
Spreadsheets used: ____		
Start Scenario 2 ____	End data entry ____	
Answer Question 1	____	____
Spreadsheets used: ____		
Answer Question 2	____	____
Spreadsheets used: ____		
Answer Question 3	____	____
Spreadsheets used: ____		
Answer Question 4	____	____
Spreadsheets used ____		
Start Scenario 3 ____	End data entry ____	
Answer Question 1	____	____
Spreadsheets used: ____		
Answer Question 2	____	____
Spreadsheets used: ____		
Answer Question 3	____	____
Spreadsheets used: ____		
Answer Question 4	____	____
Spreadsheets used: ____		

Evaluating the tool
- Overview of the evaluations ____
- Evaluation 1: Rating the Spreadsheets ____
- Evaluation 2: Improving the Spreadsheets ____
- Evaluation 3: QUIS ____

Participant's Informed Consent Form

This is an attempt to test and improve a tool for estimating course life cycle costs. To do this you'll use the tool to answer various questions about the course models in the scenarios. While you're answering the questions, please write down any problems or good features you see in the tool. I'm also going to be making notes as you answer the questions. This information will be used to find weaknesses or gaps in thesystem. After you've completed the scenarios you'll have an opportunity to rate the tool and make suggestions on improvements.

As a participant in this experiment you have certain rights as explained below. The purpose of this document is to obtain your written consent to participate in this effort.

1. You have the right to discontinue your participation in the study at any time for any reason. If you decide to terminate the experiment, inform the researcher.

2. You have the right to inspect your data and withdraw it from the experiment if you feel that you should for any reason. In general, data are processed and analyzed after a subject has completed the experiment. At that time, all identification will be removed and the data treated with anonymity. Therefore, if you wish to withdraw your data, you must do so immediately after your participation is completed.

3. You have the right to be informed of the overall results of the experiment. If you wish to receive a synopsis of the results, include your address with your signature below. If after receiving the synopsis, you would like more indepth information please contact Virginia Tech's Human Factors Laboratory and a full report will be made available to you.

The researcher is Gary Macomber. You can contact him at the following address and phone number:

Human Factors Laboratory
302 Whittemore Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA 24601
(703) 961-4602

If you have any questions about the experiment or your rights as a participant, please do not hesitate to ask. The researcher will do his best to answer them.

Your signature below indicates you have read and understand your rights as a participant (as stated above), and that you consent to participate.

____

Participants Signature

____

Witness' Signature

____

____

____

Print your name and address if you wish to receive a summary of the experimental results

Background

Problem

This tool was designed to provide a training manager with an estimate of the development, acquisition, personnel, and maintenance costs for a training course. These cost estimates are displayed across a six year life cycle. The tool can be used in two ways. The first way is to use the values for the cost factors provided with the tool. The second way is to tailor the cost factors to your specific situation.

This tool was developed to help training personnel answer course life cycle cost questions. These questions take various different forms. One is to compare the cost of using different media mixes in a course. Another is to look at the cost of changing the throughput in a course. A third type is the comparison of the costs of different training types, such as: field training, resident training, on-the job training, and contractor training.

This tool was designed to answer the first two types of questions. The third type involves some additional information beyond that the tool was designed to handle. Extending the tool to answer that type of question would first involve analyzing the differences between the various training environments and determining the appropriate cost factors for each one.

Environment

This tool was designed to model the cost factors in USAF Technical Training. In particular, resident basic courses (ABR or OBR).

Tool Development

Assumptions

This tool was designed based on a set of assumptions about the training environment and the cost factors that are related to it. Also, some assumptions were made about the types of questions and the types of answers the training manager would have and need.

The basic assumption underlying the tool's structure is that a training course's cost relationships could be modeled with a series of related spreadsheets using average and approximate values. Further, it assumes that courseware development and maintenance efforts can be costed based on the estimated number of delivery hours. For traditional instruction this may not be too large an assumption. However, for computer assisted instruction and interactive videodisc instruction it may be a large assumption. Only through carefully documenting the actual costs of many development efforts and deriving a means to categorize them can we assess the effect of this assumption on the cost estimate's relationship to the actual cost.

Another assumption underlying the tool's structure is the way it defines and relates costs. The costs for instructors and trainees include the costs for base support, branch and wing personnel, and an estimate of the other supervisory levels above the base level. Also, courseware maintenance is treated separately from classroom operations. Courseware maintenance costs are related to the amount of material and the complexity of the maintenance effort.

Constraints

This tool is constrained by its structure and the variables included explicitly and implicitly. This means the "fit" of the tool to the course being estimated is directly related to how well the course structure matches the tool's structure and how well the variables and values of the course match those of the tool. This is the basic constraint of all models and simulations. As a user, you should first compare the training course structure of the tool with your target course. Then you should compare the cost variables and values to the data you have from the target or similar courses.

The largest constraint comes from one of the design guidelines used: keep it simple. In building a tool that requires only 23 or fewer pieces of information about a course to prepare a 6 year life cycle cost estimate a lot of flexibility is designed out. For example, the tool's structure assumes all the training is done at a single site. No allowance is explicitly made for travel. Neither are other training modes (OJT, field training, contractor training) explicitly considered. Extension to these other training modes is seen as a further enhancement, first the tool needs to be tested against the environment it was designed for.

Navigation

Macro Keys

<alt> <I>	Course Input Spreadsheet
<alt> <L>	Life Cycle Cost Estimate Spreadsheet
<alt> <D>	Courseware Development Cost Estimate Spreadsheet
<alt> <M>	Courseware Maintenance Cost Estimate Spreadsheet
<alt> <P>	Personnel Cost Estimate Spreadsheet
<alt> <R>	Course Data Processing Spreadsheet
<alt> <Q>	Classroom Acquisition Cost Estimate Spreadsheet
<alt> <C>	Classroom Maintenance Cost Estimate Spreadsheet
<alt> <E>	Classroom Standards, Acquisition
<alt> <F>	Classroom Standards, Maintenance
<alt> <N>	Number of Classrooms Spreadsheet
<alt> <T>	Compare Different Course Models Spreadsheet

Instructions: To Start the Program

DOS

1. At the Dos Prompt (**A>**) type **LUCID.**
2. After LUCID has loaded (the DOS prompt has returned), press the **<control>** and the Left **<shift>** keys simultaneously.

LUCID

1. Press the </> key (the menu will appear across the top of the screen).
2. Press the **<S>** key (the highlight will move to the **STORAGE** menu item and a submenu will appear.
3. Press the **<L>** key (the highlight will move to the **LOAD** submenu item).
4. Using the arrow keys, move the highlight to **LIFE2.**
5. Press **<ENTER>.**
6. Use the MACRO keys (navigation section) to navigate within the tool.

Instructions: To End the Program

1. Press the </> key (the menu will appear across the top of the screen).
2. Press the **<E>** key (the **EXIT** submenu will appear).
3. If you want to save the changed spreadsheets as they are, press **<S>**.
4. If you want to return the spreadsheets to the values from before this session, press **<L>**.

DOS

1. Once the DOS prompt (**A>**) has returned and the red light indicating disk usage has gone out, remove the disk(s) from the drive(s).
2. Turn off the power.

GLOSSARY

Classroom Acquisition Cost Estimate Spreadsheet: <alt> <>
Uses the number and type of classroom data from the Course Processing Spreadsheet to estimate the classroom acquisition costs.

Classroom Maintenance Cost Estimate Spreadsheet:
Uses the number and type of classroom data from the Course Processing Spreadsheet to estimate the classroom maintenance costs.

Classroom Standards, Acquisition:
Sets the acquisition costs for each type of classroom based on classroom size and equipment.

Classroom Standards, Maintenance:
Sets the maintenance costs for each type of classroom based on classroom size and equipment.

Compare Different Course Model Spreadsheet:
Uses the courseware development and maintenance, classroom acquisition and maintenance, personnel, and estimated life cycle costs from the Life Cycle Cost Estimate Spreadsheet from up to three different models defined by the model # input on the Input Spreadsheet.

Course Processing Spreadsheet:
Uses the parameters from the Input Spreadsheet to calculate the data for the five cost estimate spreadsheets.

Courseware Development Cost Estimate Spreadsheet:
Uses the number of hours of instruction to be developed by media type data from the Course Processing Spreadsheet to calculate the total estimated cost of courseware development.

Courseware Maintenance Cost Estimate Spreadsheet:

Uses the number of hours of instruction to be maintained by media type data from the Course Processing Spreadsheet to calculate the total estimated cost of courseware maintenance.

Input Spreadsheet:

Used to set the course parameters for the model. Sends the parameters to the Course Processing Spreadsheet.

Life Cycle Cost Estimate Spreadsheet:

Uses the cost estimates from the courseware and operations categories and projects them across a six year life cycle. The first year involves courseware development and classroom acquisition. The other five years consist of personnel costs, classroom maintenance, and courseware maintenance.

Number of Classrooms Spreadsheet:

Uses the course media mix parameters from the Input Spreadsheet and the number of classes from the Course Processing Spreadsheet to estimate the number of classrooms needed by type of classroom.

Personnel Cost Estimate Spreadsheet:

Uses the number of training days per year and the number of instructors from the Course Processing Spreadsheet to estimate the cost of personnel.

CAI 1:

This refers to the simplest form of computer assisted instruction. The courseware is not graphics or programming intensive.

CAI 2:

This refers to the middle range of difficulty in amount of graphics and programming. This level includes the simpler forms of gaming and operation. It's considered to be similar to a part task trainer.

CAI 3:

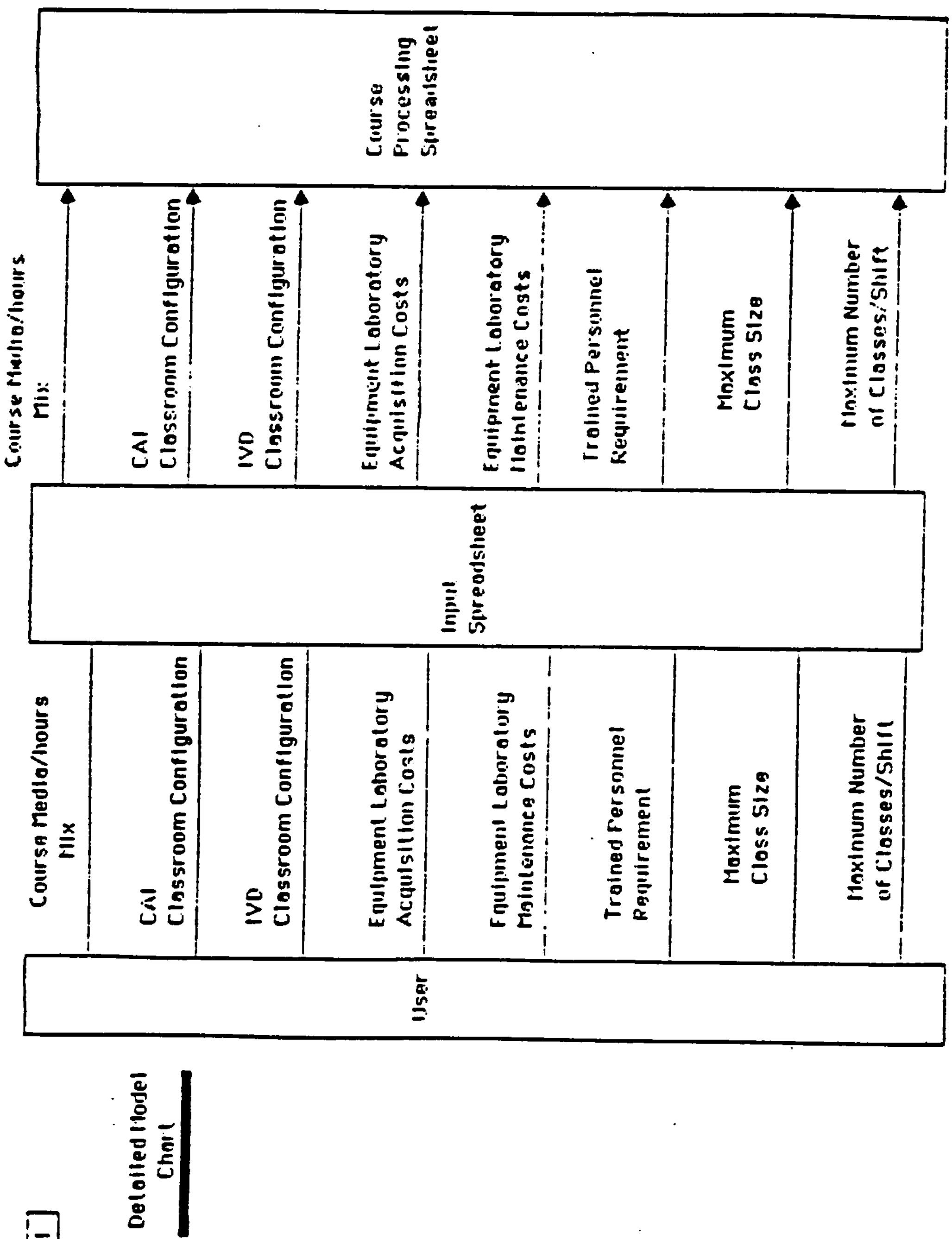
This refers to the most complex level of CAI. It indicates a high level of interactivity between the system and the student. This level compares to an operational simulation.

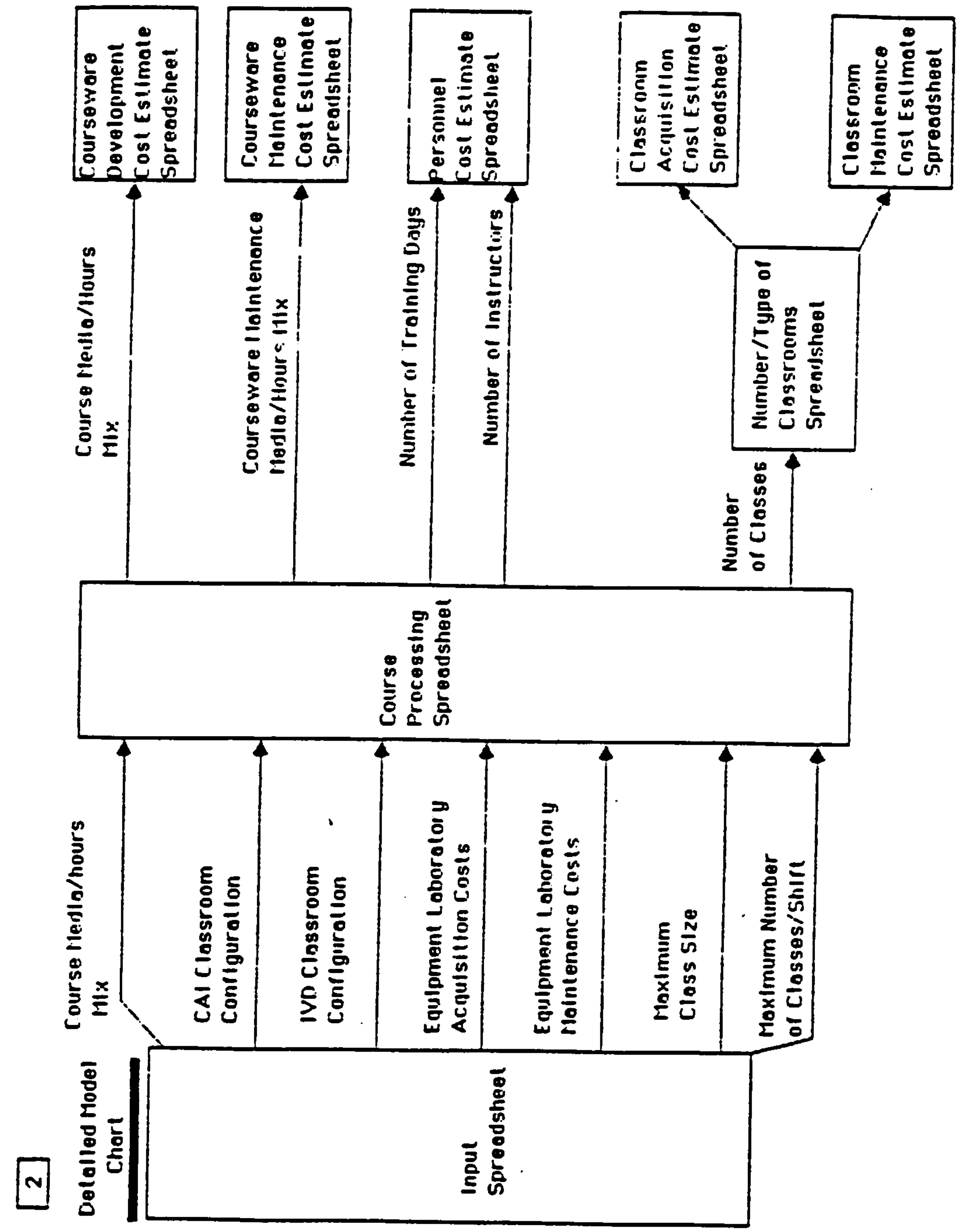
IVD 1:

This refers to the simplest level of IVD. It involves a low level of interactivity, little or no use of sound on the disc, and not much need for overlaying the images with graphics.

IVD 2:

This refers to a more complex level of IVD. Complexity can come from several different areas. The course could involve simulation/gaming (like CAI 3). Or the programming effort could be more complex due to the use of digitized sound or the use of graphic overlays.

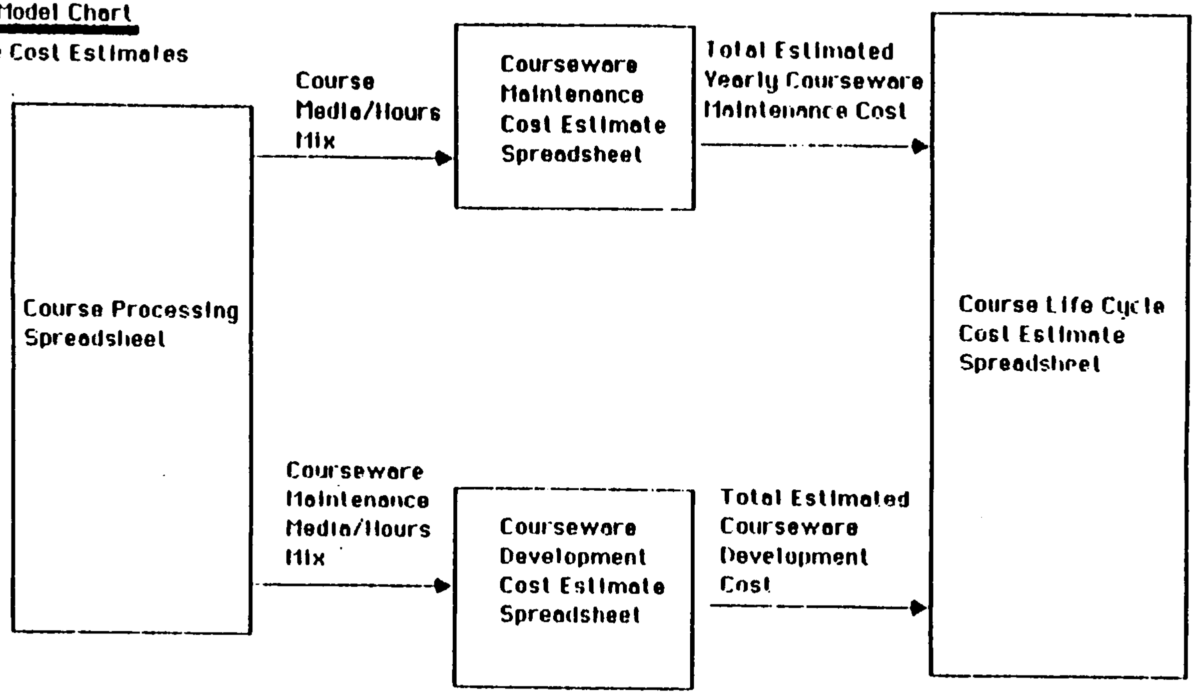

1
Detailed Model Chart
User
Course Media/hours Mix
CAI Classroom Configuration
IVD Classroom Configuration
Equipment Laboratory Acquisition Costs
Equipment Laboratory Maintenance Costs
Trained Personnel Requirement
Maximum Class Size
Maximum Number of Classes/Shift
Input Spreadsheet
Course Media/hours Mix
CAI Classroom Configuration
IVD Classroom Configuration
Equipment Laboratory Acquisition Costs
Equipment Laboratory Maintenance Costs
Trained Personnel Requirement
Maximum Class Size
Maximum Number of Classes/Shift
Course Processing Spreadsheet


2
Detailed Model Chart
Input Spreadsheet
Course Media/hours Mix
CAI Classroom Configuration
IVD Classroom Configuration
Equipment Laboratory Acquisition Costs
Equipment Laboratory Maintenance Costs
Maximum Class Size
Maximum Number of Classes/Shift
Course Processing Spreadsheet
Course Media/Hours Mix
Courseware Maintenance Media/Hours Mix
Number of Training Days
Number of Instructors
Number of Classes
Courseware Development Cost Estimate Spreadsheet
Courseware Maintenance Cost Estimate Spreadsheet
Personnel Cost Estimate Spreadsheet
Number/Type of Classrooms Spreadsheet
Classroom Acquisition Cost Estimate Spreadsheet
Classroom Maintenance Cost Estimate Spreadsheet

3

<u>Detailed Model Chart</u>

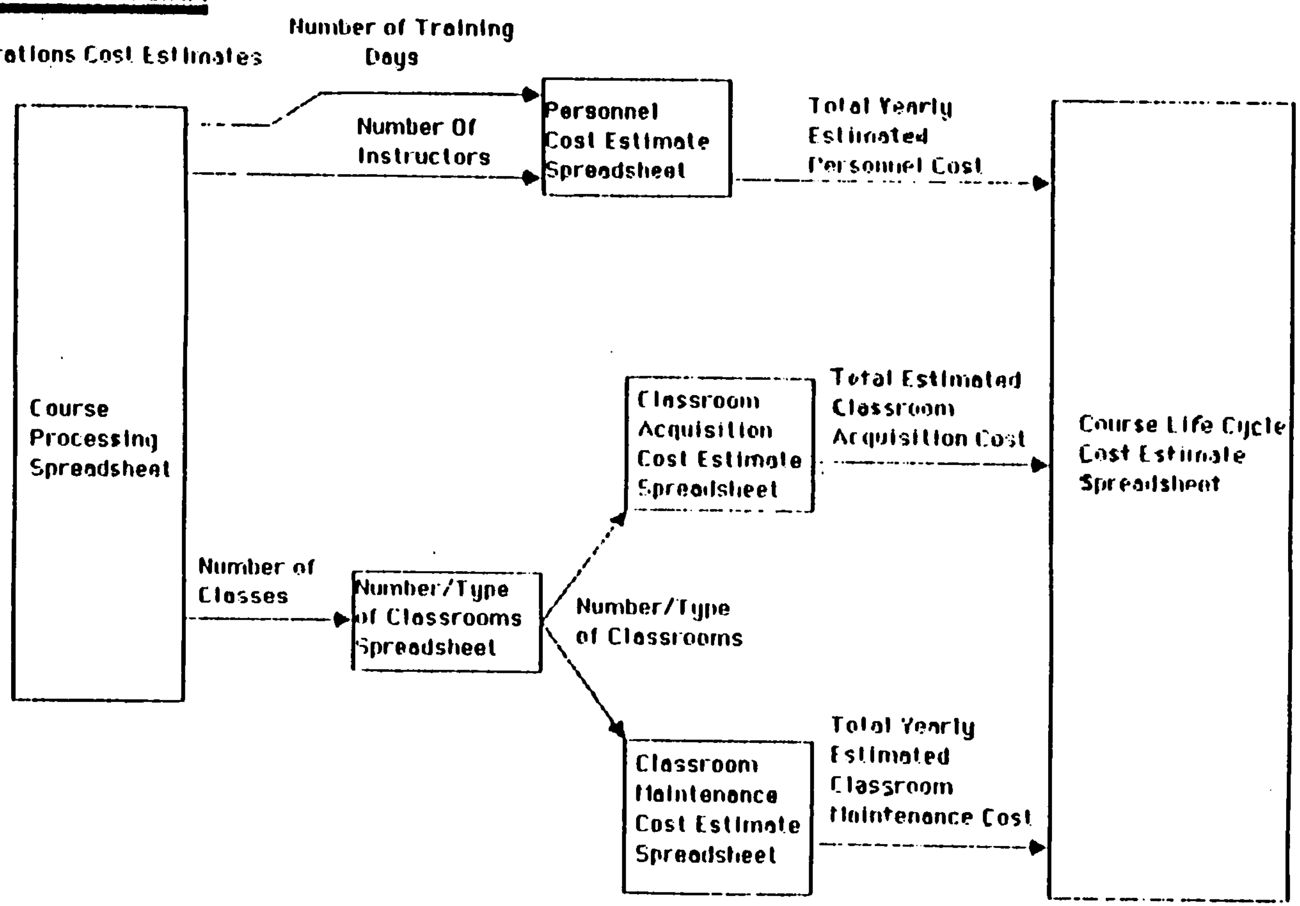
Courseware Cost Estimates



4

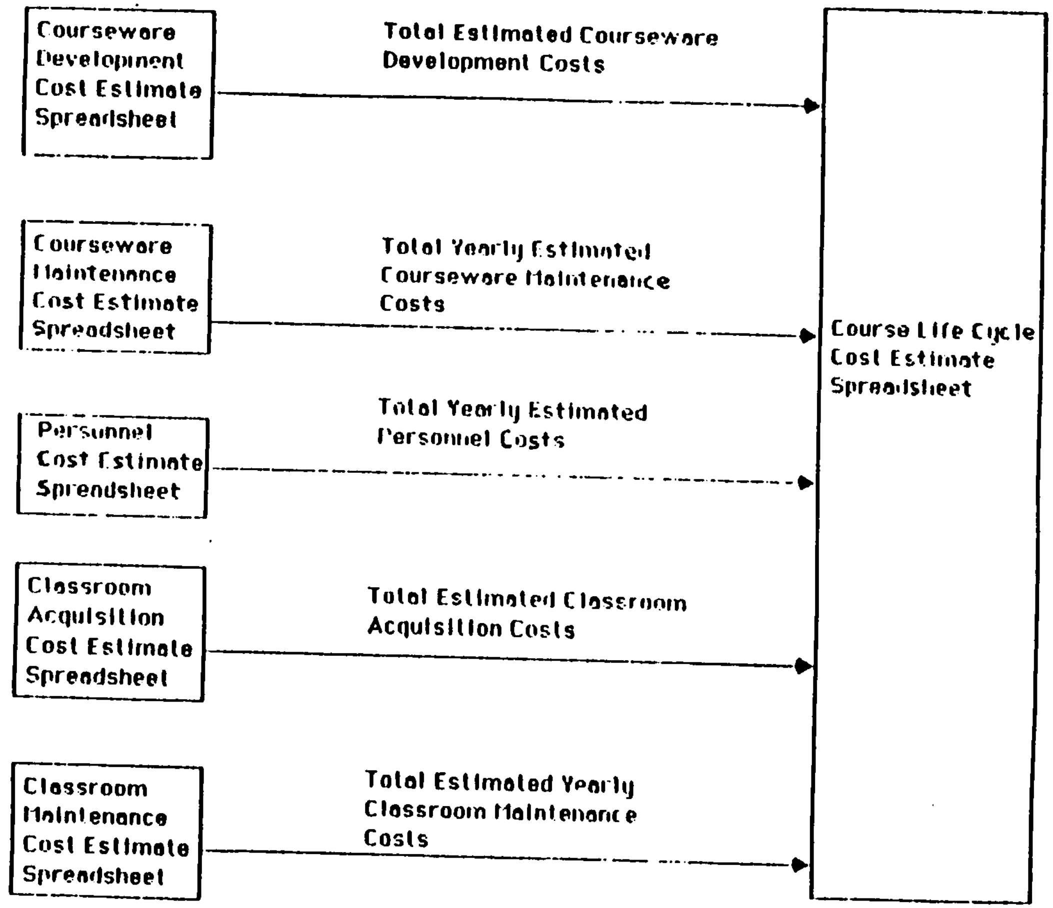
Detailed Model Chart

Operations Cost Estimates



5

Detailed Model Chart



Scenarios

This packet contains 3 scenarios for the Training/Cost Model and a question set for each scenario. These scenarios contain all the information necessary to answer the questions. The scenarios are part of the evaluation of the tool. They provide a means to obtain some performance measures and to give training planners some hands on experience with the tool. There is no time limit on the tasks, but you will be timed. This is to provide us with an estimate of the time required to use the tool. Also, your performance will be monitored, this provides us with information on how we should organize the information in the tool to make it easier to use. Use any of the materials you've been given to help you find the answers to the questions. When you are ready please turn the page and begin.

Scenario 1

Task: To develop the life cycle cost profile from the course data given below and the life cycle cost tool provided.

Course Data:

Course model information:

Course Model #:	1
Inflation Rate	.05
# of Instructors	0

(note: # of instructors is used only to fix the number of instructors, if 0 is input then the model determines the number of instructors required)

Performance information:

Trained Personnel Requirement:	200
Historical Attrition Rate:	.10
Maximum Class Size	20
Maximum Number of Classes at one time	5
% of course revised each year	.10

Course Media Mix:	instructional hours
Lecture:	300
Film/Slide/Tape:	100
CAI 1:	0
CAI 2:	0
CAI 3:	0
IVD 1:	0
IVD 2:	0
Actual Equipment Laboratory 1	75
Actual Equipment Laboratory 2	50

Classroom Configuration:

Type	Configuration
CAI	2
IVD	2

Cost of Actual Equipment Laboratories

	Acquisition	Maintenance
Laboratory 1	$150,000	$3,000/yr
Laboratory 2	250,000	3,000

Question Set 1

Based on Scenario 1, find the answers to the following questions.

1. How many training days are estimated for this course configuration?

2. What is the total estimated cost of courseware development?

3. What is the total estimated number of classrooms?

4. What is the total estimated personnel cost for the first year of operations?

Scenario 2

Task: Develop the life cycle cost profile from the course data given below and the life cycle cost tool provided, then answer the questions on the next page.

Course Data:

Course model information:

Course Model #:	2
Inflation Rate	.05
# of Instructors	0

(note: # of instructors is used only to fix the number of instructors, if 0 is input then the model determines the number of instructors required)

Performance information:

Trained Personnel Requirement	200
Historical Attrition Rate	.08
Maximum Class Size	20
Maximum Number of Classes at one time	5
% of course revised each year	.10

Course Media Mix:	instructional hours
Lecture:	235
Film/Slide/Tape:	80
CAI, 1	24
CAI, 2	40
CAI, 3	0
IVD, 1	20
IVD, 2	0
Actual Equipment Laboratory 1	65
Actual Equipment Laboratory 2	40

Classroom Configuration:

Type	Configuration
CAI	2
IVD	2

Cost of Actual Equipment Laboratories

	Acquisition	Maintenance
Laboratory 1	$150,000	$3,000/yr
Laboratory 2	250,000	3,000

Question Set 2

Based on Scenario 2 find the answers to the following questions.

1. What is the total estimated courseware maintenance cost for the first year of operations?

2. What is the estimated cost of acquiring Lect/Film/Tape classrooms for this course model

3. What is the total estimated life cycle cost for this course configuration?

4. What is the estimated cost of instructors for this course configuration?

Scenario 3

Task: Develop the life cycle cost profile from the course data given below and the life cycle cost tool provided.

Course Data:

Course model information:

Course Model #:	3
Inflation Rate	.05
# of Instructors	0

(note: # of instructors is used only to fix the number of instructors, if 0 is input then the model determines the number of instructors required)

Performance information:

Trained Personnel Requirement	200
Historical Attrition Rate	.05
Maximum Class Size	20
Maximum Number of Classes at one time	5
% of course revised each year	.10

Course Media Mix:	instructional hours
Lecture:	150
Film/Slide/Tape:	50
CAI 1	30
CAI 2	100
CAI 3	30
IVD 1	30
IVD 2	22
Actual Equipment Laboratory 1	35
Actual Equipment Laboratory 2	25

Classroom Configuration:

Type	Configuration
CAI	2
IVD	2

Cost of Actual Equipment Laboratories

	Acquisition	Maintenance
Laboratory 1	$150,000	$3,000/yr
Laboratory 2	250,000	3,000

Question Set 3

Task: Based on the 3 course models described above, answer the following questions.

1. Which of the three has the lowest total estimated life cycle cost?

2. What are the three total estimated life cycle costs?

Model 1:

Model 2:

Model 3:

3. Which cost estimates changed from Model 1 to Model 2?

4. Which cost estimates changed from Model 2 to Model 3?

Now that you've worked with the tool a little and have had a chance to see most of the parts of it, I'd like your suggestions and ratings of the tool and its usefulness to you. To do this there are 3 different question sets. The first set asks you to rate the importance of each spreadsheet to you in answering the questions you might be asked. The second set asks you for any changes, either additions or deletions, or even other ways of calculating or representing the information that would make the tool more useful for you. The third set is the Questionnaire of User Interface Satisfaction (QUIS) this questionnaire was designed to assess how you feel about the tool in terms of some standard comparisons. Some of the questions in QUIS might not apply to this situation, so please circle n/a for those questions you feel don't apply to you.

Questionnaire on the Usefulness of the Spreadsheets

1. Input Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

2. Course Processing Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

3. Courseware Development Cost Estimate Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

4. Courseware Maintenance Cost Estimate Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

5. Classroom Acquisition Cost Estimate Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

6. Classroom Maintenance Cost Estimate Spreadsheet

very useful						unneeded
1	2	3	4	5	6	7

7. Personnel Cost Estimate Spreadsheet

very useful 1	2	3	4	5	6	unneeded 7

8. Life Cycle Cost Estimate Spreadsheet

very useful 1	2	3	4	5	6	unneeded 7

9. Compare Different Course Models Spreadsheet

very useful 1	2	3	4	5	6	unneeded 7

10. Classroom Standards, Acquisition

very useful 1	2	3	4	5	6	unneeded 7

11. Classroom Standards, Maintenance

very useful 1	2	3	4	5	6	unneeded 7

Suggestions for Improving the Spreadsheets

1. Input Spreadsheet

2. Course Processing Spreadsheet

3. Courseware Development Cost Estimate Spreadsheet

4. Courseware Maintenance Cost Estimate Spreadsheet

5. Classroom Acquisition Cost Estimate Spreadsheet

6. Classroom Maintenance Cost Estimate Spreadsheet

7. Personnel Cost Estimate Spreadsheet

8. Life Cycle Cost Estimate Spreadsheet

9. Compare Different Course Models Spreadsheet

10. Classroom Standards, Acquisition

11. Classroom Standards, Maintenance

QUESTIONNAIRE OF USER INTERFACE SATISFACTION

Last four digits of your student identification number ____________

Age _______

Sex Male Female

PART 1: Type of System to be Rated

1. Name of software __________ Name of hardware __________

2. Length of time you have worked on this system

less than 1 hour	6 months to less than 1 year
1 hours to less than 1 day	1 year to less than 2 years
1 day to less than 1 week	2 years to less than 3 years
1 week to less than 1 month	3 years of more
1 month to less than 6 months	

3. Average usage per week

less than one hour	4 to less than 10
one to less than 4 hours	over 10 hours

PART 2: Past Experience

1. How many different types of computer systems (e.g., main frames and personal computers have you worked with?

none	3-4
1	5-6
2	more than 6

2. Of the following devices software and systems check those that you have personally used and are familiar with

keyboard	text editor	color monitor
numeric key pad	word processor	time-share system
mouse	file manager	personal computer
light pen	electronic spreadsheet	lap computer
touch screen	electronic mail	computer magazines
track ball	computer games	computer user's group
joy stick	video games	floppy disks

PART 3: User Evaluation of an Interactive Computer System

Please circle the numbers which most appropriately reflect your impressions about using this computer system. Not applicable = NA. Please add your written comments below the corresponding item.

Overall reactions to the system

terrible wonderful
1 2 3 4 5 6 7 8 9 NA

frustrating satisfying
1 2 3 4 5 6 7 8 9 NA

dull stimulating
1 2 3 4 5 6 7 8 9 NA

difficult easy
1 2 3 4 5 6 7 8 9 NA

inadequate power adequate power
1 2 3 4 5 6 7 8 9 NA

rigid flexible
1 2 3 4 5 6 7 8 9 NA

PART A SCREEN

1. Characters on the computer screen — hard to read easy to read
1 2 3 4 5 6 7 8 9 NA

1.1 Image of characters — fuzzy sharp
1 2 3 4 5 6 7 8 9 NA

1.2 Character shapes (fonts) — barely legible very legible
1 2 3 4 5 6 7 8 9 NA

2. Highlighting on the screen makes task easier — not at all very much
1 2 3 4 5 6 7 8 9 NA

2.1 User of reverse video — unhelpful helpful
1 2 3 4 5 6 7 8 9 NA

2.2 Use of blinking — unhelpful helpful
1 2 3 4 5 6 7 8 9 NA

3. Screen layouts make tasks easier — never / always
1 2 3 4 5 6 7 8 9 NA

3.1 Amount of information that can be displayed screen — inadequate / adequate
1 2 3 4 5 6 7 8 9 NA

3.2 Arrangement of information on screen — illogical / logical
1 2 3 4 5 6 7 8 9 NA

4. Sequence of screens — confusing / clear
1 2 3 4 5 6 7 8 9 NA

4.1 Next screen in a sequence — unpredictable / predictable
1 2 3 4 5 6 7 8 9 NA

4.2 Going back to the previous screen — impossible / easy
1 2 3 4 5 6 7 8 9 NA

4.3 Beginning middle and end of — confusing / clearly marked
1 2 3 4 5 6 7 8 9 NA

PART B TERMINOLOGY AND SYSTEM INFORMATION

5. Use of terms throughout system — inconsistent / consistent
1 2 3 4 5 6 7 8 9 NA

5.1 Task terms — inconsistent / consistent
1 2 3 4 5 6 7 8 9 NA

5.2 Computer terms — inconsistent / consistent
1 2 3 4 5 6 7 8 9 NA

6. Terminology relates to the work you are doing — unrelated / related
1 2 3 4 5 6 7 8 9 NA

6.1 Computer terminology is used — too frequently / appropriately
1 2 3 4 5 6 7 8 9 NA

6.2 Terms on the screen — ambiguous / precise
1 2 3 4 5 6 7 8 9 NA

7. Messages which appear on screen	inconsistent consistent 1 2 3 4 5 6 7 8 9	NA
7.1 Position of instructions on the screen	inconsistent consistent 1 2 3 4 5 6 7 8 9	NA
8. Messages to the user	confusing clear 1 2 3 4 5 6 7 8 9	NA
8.1 Instructions for commands of	confusing clear 1 2 3 4 5 6 7 8 9	NA
8.2 Instructions for correcting errors	confusing clear 1 2 3 4 5 6 7 8 9	NA
9. Computer keeps you informed about	never always 1 2 3 4 5 6 7 8 9	NA
9.1 Performing an operation leads to a predictable result	never always 1 2 3 4 5 6 7 8 9	NA
9.2 User can control amount of feedback	never always 1 2 3 4 5 6 7 8 9	NA
10. Error messages	unhelpful helpful 1 2 3 4 5 6 7 8 9	NA
10.1 Error messages clarify the	never always 1 2 3 4 5 6 7 8 9	NA
10.2 Phrasing of error messages	unpleasant pleasant 1 2 3 4 5 6 7 8 9	NA

PART C LEARNING

11. Learning to operate the system	difficult easy	1 2 3 4 5 6 7 8 9 NA
11.1 Getting started	difficult easy	1 2 3 4 5 6 7 8 9 NA
11.2 Learning advanced features	difficult easy	1 2 3 4 5 6 7 8 9 NA
11.3 Time to learn to use the	slow fast	1 2 3 4 5 6 7 8 9 NA
12. Exploration of features by trial and error	discouraged encouraged	1 2 3 4 5 6 7 8 9 NA
12.1 Exploration of features	risky safe	1 2 3 4 5 6 7 8 9 NA
12.2 Discovering new features	difficult easy	1 2 3 4 5 6 7 8 9 NA
13. Remembering names and use of commands	difficult easy	1 2 3 4 5 6 7 8 9 NA
13.1 Remembering specific rules about entering commands	difficult easy	1 2 3 4 5 6 7 8 9 NA
14. Tasks can be performed in a straight-forward manner	never always	1 2 3 4 5 6 7 8 9 NA
14.1 Number of steps per task	too many just right	1 2 3 4 5 6 7 8 9 NA
14.2 Steps to complete a task	rarely always	1 2 3 4 5 6 7 8 9 NA
14.3 Completion of sequence of steps	unclear clear	1 2 3 4 5 6 7 8 9 NA

15. Help messages on the screen	confusing clear 1 2 3 4 5 6 7 8 9 NA
15.1 Accessing help messages	difficult easy 1 2 3 4 5 6 7 8 9 NA
15.2 Content of help messages	confusing clear 1 2 3 4 5 6 7 8 9 NA
15.3 Amount of help	inadequate adequate 1 2 3 4 5 6 7 8 9 NA
16. Supplement reference materials	confusing clear 1 2 3 4 5 6 7 8 9 NA
16.1 Tutorials for beginners	confusing clear 1 2 3 4 5 6 7 8 9 NA
16.2 Reference manuals	confusing clear 1 2 3 4 5 6 7 8 9 NA

PART D SYSTEM CAPABILITIES

17. System speed	too slow fast enough 1 2 3 4 5 6 7 8 9 NA
17.1 Response time for most operations	too slow fast enough 1 2 3 4 5 6 7 8 9 NA
17.2 Rate information is displayed	too slow fast enough 1 2 3 4 5 6 7 8 9 NA

18. Reliability of the system	unreliable reliable	1 2 3 4 5 6 7 8 9 NA
18.1 Operations are	undependable dependable	1 2 3 4 5 6 7 8 9 NA
18.2 System failures occur	seldom frequently	1 2 3 4 5 6 7 8 9 NA
18.3 System warns the user about potential problems	never always	1 2 3 4 5 6 7 8 9 NA
19. System tends to be	noisy quite	1 2 3 4 5 6 7 8 9 NA
19.1 Mechanical devices such as fans disks and printers	noisy quite	1 2 3 4 5 6 7 8 9 NA
19.2 Computer tones beeps clicks etc	annoying pleasant	1 2 3 4 5 6 7 8 9 NA
20. Correcting your mistakes	difficult easy	1 2 3 4 5 6 7 8 9 NA
20.1 Correcting typos or mistakes	complex simple	1 2 3 4 5 6 7 8 9 NA
20.2 Ability to undo operations	inadequate adequate	1 2 3 4 5 6 7 8 9 NA
21. The needs of both experienced and inexperienced users are taken	never always	1 2 3 4 5 6 7 8 9 NA
21.1 Novices can accomplish tasks knowing only a few commands	with difficulty easily	1 2 3 4 5 6 7 8 9 NA
21.2 Experts can use features shortcuts	with difficulty easily	1 2 3 4 5 6 7 8 9 NA

APPENDIX I

Second Evaluation Comments

Input Speadsheet

1.no problem with any of them; they all seem fairly straight forward.
2. Maybe format, but it is but it is easy to follow as is.
3. How about when you press enter the cursor moves to the next field?
4,5,6,11,13. None
7,8. Add item for complexity.
9. Protect labels; add washback rate and program class size.
10. The spreadsheet is very well laid out and easy to use. Familiarization could make this a useful tool for the T.M.
12. Good, usable, corresponds nicely with PCD Part 1 and PMS course info.
14. Add block for course # and STS/CTS info.
15. Too busy. Could possibly use help frame to define terms.
16. All spreadsheets look good. The number seem adequate.
17. Good
18. OK

Course Processing Spreadsheet

1,4,5,8,10,11,13,16. None
2. Should indicate max capability of TPR given fixed resources.
3. Total training days - should be something like "student man-days" or "student tng days".
6. Info on same line as label (not below)
7. Add travel status cost - TDY <u>vs</u> PCS.
9. "in/out proc"
12. In/Out processing entry really not a function of training mgt. personnel. Established by regulations and varies with TTC. Bottom screen info particularly helpful for TPR conference and Wartime expansion info.
14. No additions except add columns or codes for pipeline, cross train, ANG, AFR personnel.
15. Not familiar with the term recycle. I think you are referring to those students who washback and repeat the unit of instruction. However, you don't break a course down into units.
17. Good
18. OK

Courseware Development Cost Estimate Spreadsheet

1,4,5,6,7,8,10,11,13,14,16. None
2. Are cost estimates for revisions the same as new courseware?
3,9,15. OK
12. Good info for budgeting and Justification. Depending on who is using it, usefulness could vary from must know to nice to know.
17. Good
18. Seems this should change between course complexities.

Courseware Maintenance Cost Estimate Spreadsheet

1,4,5,6,7,8,10,11,16. None
2. Could be combined with Courseware Development.
3,9,15. OK
12. Primarily a resources function. Nice to know from training manager perspective. Budgeting info best use.
14,17. Good
18. think this should have a few variable inputs: Hi Tech vs non Hi Tech.

Classroom Acquisition Cost Estimate Spreadsheet

1,2,4,5,6,8,10,11,13,16. None
3,14,17. Good
7. Add climate control.
9,15,18. OK
12. Very necessary in Course Resource Estimate Phase!

Classroom Maintenance Cost Estimate Spreadsheet

1,2,4,5,6,8,10,11,13,16. None
3,9,18. OK
7. Add climate control.
12. Primarily a resources function. Nice to know from training manager perspective. Budgeting info best use.
14,17. Good
15. What factors make up classroom maintenance?

Personnel Cost Estimate Spreadsheet

1,2,5,6,7,8,10,11,16. None
3. Is the trainee unit cost adjustable? It should be, based on different locales and environments
4. Need to add overhead cost (TDB personnel)
9,15,18. OK
12. Good info to determine "bang for the buck".
13. Like it.
14,17. Good

Life Cycle Cost Estimate Spreadsheet

1,2,4,5,6,7,8,10,16. None
3,9,15. OK
11. Need on screen prompt to remind user that more info is available.
12. A must!
13. Like it.
14,17,18. Good

Compare Different Course Models Spreadsheet

1,2,4,5,8,10,11,13,16. None
3,18. Good
6. Add cost per student.
7,15. OK
9. A++
12. Excellent wrap-up device.
14. Good comparison information.
17. Very Good

Classroom Standards, Acquisition

1,2,4,5,6,7,8,10,11,13,16. None
3,14. Good
9. A+
12. After initial planning stages, another resources tool
15. Would like to see more flexibility in manipulation of the variables.
17. Very Good
18. OK+

Classroom Standards, Maintenance

1,2,4,5,6,7,8,10,11,13,16. None
3,18. Good
9. A+
12. Primarily a resources function. Nice to know from training manager perspective. Budgeting info best use.
14. Good information
15. Would like to see more flexibility in manipulation of the variables.
17. Very Good

Other Comments

3. How about two other outputs?
    a. Set parameters with number of students and number of classrooms - can we handle that load?
    b. How many students can we train with available classrooms?
6. Need a lab cost spreadsheet.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2