FAA-AM-82-2

A GENERIC MODEL FOR EVALUATION OF THE FEDERAL AVIATION ADMINISTRATION AIR TRAFFIC CONTROL SPECIALIST TRAINING PROGRAMS

James O. Boone
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma

March 1982

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, D.C. 20591

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-82-2		

4. Title and Subtitle	5. Report Date
A GENERIC MODEL FOR EVALUATION OF THE FEDERAL AVIATION ADMINISTRATION AIR TRAFFIC CONTROL SPECIALIST TRAINING PROGRAMS	March 1982
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
James O. Boone	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, Oklahoma 73125	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	
Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591	14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under Tasks AM-C-80/81-PSY-87.

16. Abstract

The Systems Analysis Research Unit at the Civil Aeromedical Institute (CAMI) has developed a generic model for Federal Aviation Administration (FAA) Academy training program evaluation. The model will serve as a basis for integrating the total data base into a common format across all training programs. The model consists of four components: (1) design, (2) implementation, (3) formative, and (4) summative evaluation. Design evaluation is an assessment of the comprehensive implementation plan; implementation evaluation is a determination that the plan is completely and accurately implemented according to prescription; formative evaluation is a continual monitoring of the program to keep the process reliable, stable, and on track; and summative evaluation monitors the product of the training program. The design evaluation relies on the task, knowledge, and skills analysis and the documents in the implementation plan. The implementation evaluation makes use of the data from frequent status studies. Formative and summative evaluations make use of statistics and mathematical modeling, primarily linear regression models, to monitor the process and products of the programs and to estimate and determine the impact of changes made to the programs.

17. Key Words	18. Distribution Statement
Program Evaluation Math Modeling Training Air Traffic Control	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	29	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGEMENTS

Acknowledgements are given to Leland Page for providing Figures 2, 3. 4, 5, and 6, to Allan VanDeventer and Linda Ritchie for providing the report formats in Appendices A and B, and to Jo Ann Steen for the preparation of this manuscript.

A GENERIC MODEL FOR EVALUATION OF THE FEDERAL AVIATION ADMINISTRATION AIR TRAFFIC CONTROL SPECIALIST TRAINING PROGRAMS

I. Introduction.

In a large training institution such as the Federal Aviation Administration (FAA) Academy, several independent training programs operate simultaneously. As new technology becomes available for training, especially in the computer field, new training methods are frequently implemented. The new simulation facility for radar Air Traffic Control Specialist (ATCS) training and the new PLATO computer-based instruction system are examples of these advances. It is redundant and incoherent to develop a new program evaluation for each new development in ATCS training methods. Consequently the Systems Analysis Research Unit at the Civil Aeromedical Institute (CAMI) has developed a generic program evaluation model for Academy training programs. While the ATCS training programs were the primary aim of the model, it is appropriate for Airway Facility or Flight Standards training programs. The generic model allows research at CAMI on Academy programs to be integrated into our total systems approach by making specific application of the generic model to any new Academy development. By consistent application of the generic model, the data collected on programs will be compatible with our continuing data base and offer a means of expanding our total picture of Academy training programs in an integrated fashion.

II. Description of the Program Evaluation Model Components.

Program evaluation is designed to accomplish several tasks. These tasks are to (i) define exactly what the program is, its purposes and goals, (ii) document the exact structure of the program, (iii) define the process in the program (a logical step-by-step explanation) that achieves the goals, (iv) monitor the process to insure that any breakdown in the program during implementation or operation can be identified, (v) measure the outcomes of the program to determine if it is accomplishing its goals, and (vi) define and document any program revisions made to change the process, including the basis for the change and how this alters the structure and paths to produce the desired results. This paper describes a generic model for ATCS training program evaluation. The four components of the model are (i) design evaluation, (ii) implementation evaluation, (iii) formative evaluation, and (iv) summative evaluation (3).

Program design and implementation evaluations, as the terms imply, occur at the beginning of the program. Formative and summative evaluations occur simultaneously and serve to evaluate the process and course of the program as well as its products. Each of these evaluation components uses the techniques of statistics, math modeling, and various reporting systems.

Design Evaluation. Program design evaluation involves insuring the proper development of several tasks that make up the program implementation plan. First, the overall objectives of the program must be clearly defined. Every expected outcome of the program should be listed. The outcomes should be organized by broad categories and related to the objectives of the program. All curricula objectives, student assessment techniques and instruments, and teaching/training lesson plans must be based firmly on thorough task, knowledge, and skills analyses. A task analysis is a careful documentation of all the tasks performed in controlling air traffic and their relative importance and interaction. A knowledge and skills analysis is a determination of the knowledge and skills and knowledge and skill levels required to perform each task. Consequently, the task, knowledge, and skills analyses serve as the precise and clear job sample on which the student curricula, assessment, and teaching/training lesson plans are based. This is a very crucial and important step.

Next the teaching/training and assessment process or methods must be operationally defined. This involves a logically connected step-by-step explanation of the methods to be employed in accomplishing each of the outcomes and measuring the accomplishment of each of the objectives. This should include the use of any teaching equipment or aids. Flowcharts, PERT, tables, GANTT charts, and graphs should be used as appropriate in defining the process. Careful documentation of every step should be made during this evaluation phase by the evaluation staff with regular reports to the responsible supervisor on the progress of the design. The completed implementation plan should be clear enough that any competent educational expert could carry out the design. Figure 1 illustrates the process of specifying the program design requirements.

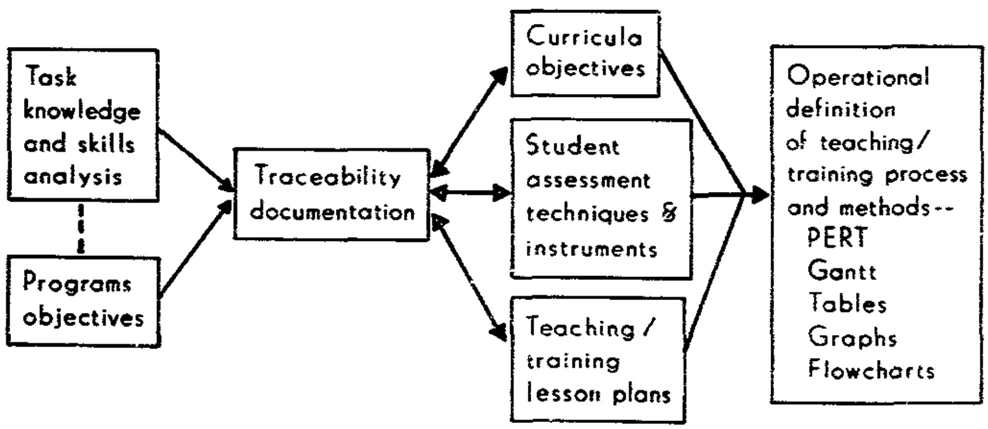

FIGURE 1. SPECIFICATION OF PROGRAM DESIGN REQUIREMENTS.

In the case of automated ATCS training systems, the design phase has several additional components. First, the operational requirements (Figure 1) from the task, knowledge, and skills analysis are stated in terms of the functional products that a training system must produce. This is a clear description of the visible workings/outcomes of the needed training system. The functional requirements should contain only the essentials necessary to simulate the required operational activity for the determined level of training. Figure 2 describes this step.

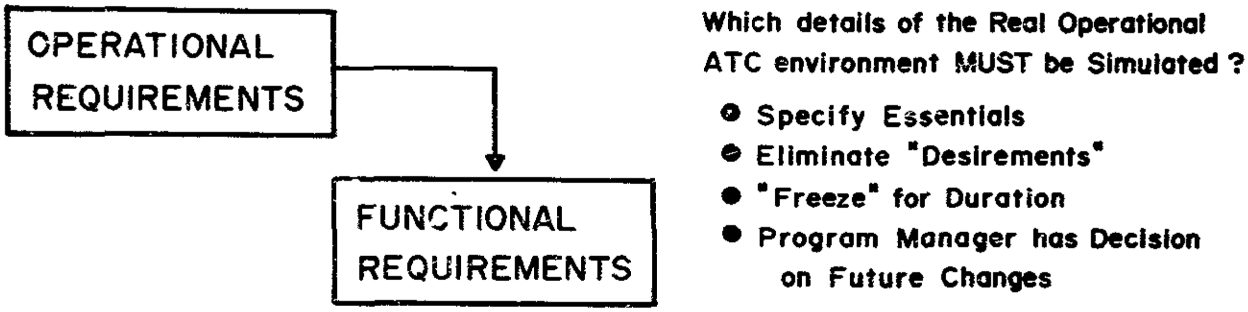

FIGURE 2. FUNCTIONAL REQUIREMENTS.

It is at this stage that computer-derived measures to assess student performance are stated as a functional requirement. Particular care must be taken in this phase to eliminate any unnecessary requirements. As pointed out by Page (2), the optimal, cost-efficient point on the complexity function is the minimal system required to satisfy the needed functional activity (see Figure 3).

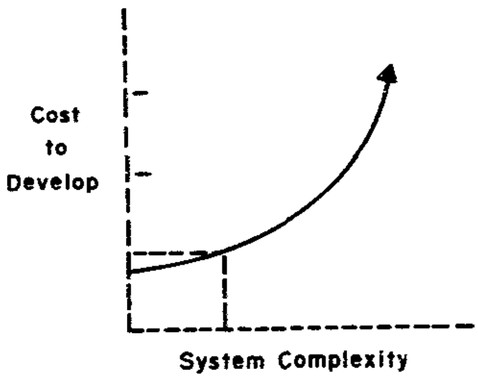

FIGURE 3. THE COMPLEXITY TRADEOFF.

The next design phase concerns the engineer more than the educational technologist; however, the educational technologist is involved in this stage and should be aware of the process. This phase is the design

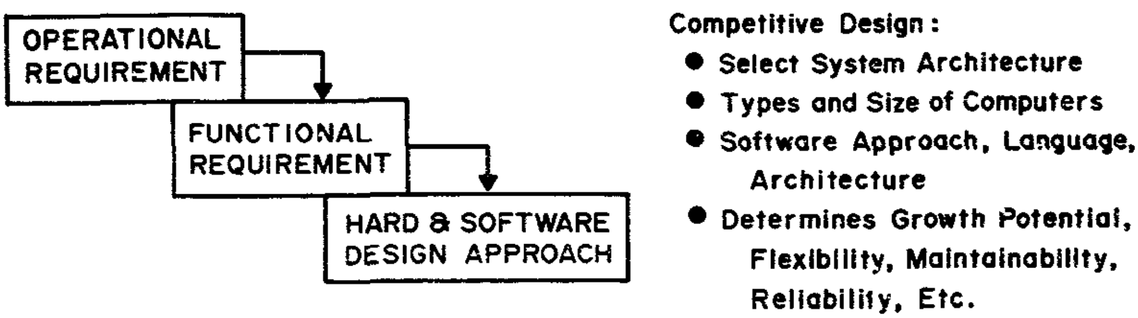

FIGURE 4. THE DESIGN APPROACH.

approach. As Figure 4 points out, this step includes the selection of the most reliable and cost-efficient minimal system architecture. The educational technologist acts as a consultant to the system engineer to insure that the selected system performance will satisfy the functional requirements.

Page (2) points out several reasons why it is very important to make correct judgments about the system during the design approach phase:

(i) The developer has to make the corrections;
(ii) The impact on program cost is less;
(iii) The cost to the user after system delivery (maintenance) is much less; and
(iv) The system will be less troublesome to use early in the operational phase.

Figure 5 further illustrates the impact on cost of making errors that must be corrected later in the development process. The two lines on the graph show the relative cost for making a large number of errors versus many fewer errors.

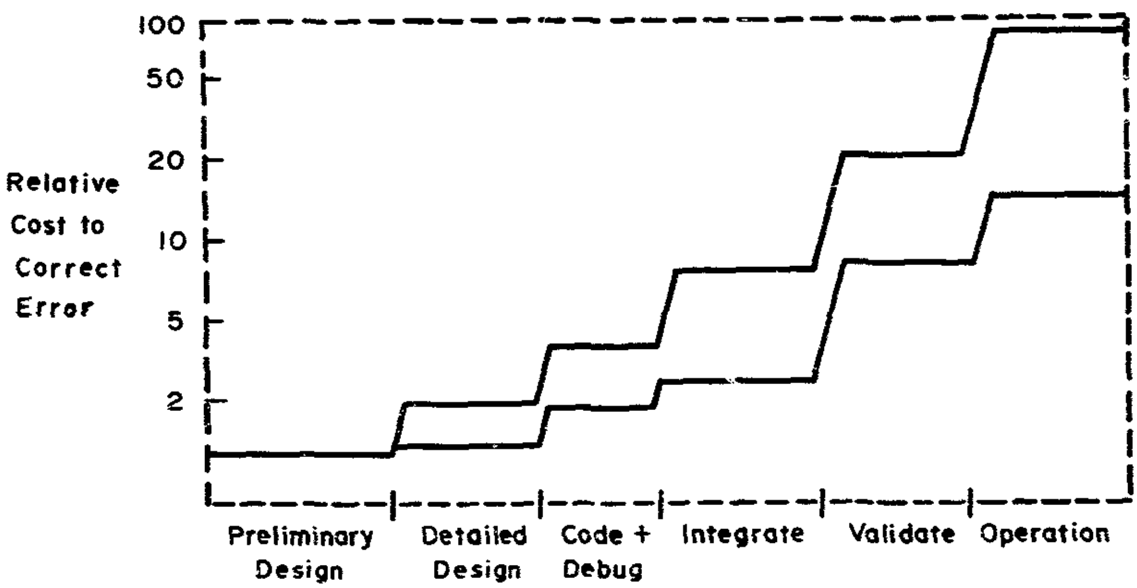

FIGURE 5. ERROR VERSUS COST.

The remainder of the development process during the design evaluation consists of the detailed design, hardware and software development, and system testing. The detailed design and hardware and software development are engineering tasks; however, the educational technologist again acts as a consultant to insure that the product satisfies the needs of the training requirements. Figure 6 depicts the entire process. The system testing phase is particularly important to the educational technologist, since this is a demonstration of the system's ability to perform the functional requirements as specified. Care should be taken to insure that the test is a valid demonstration, covering all aspects of the functional requirements

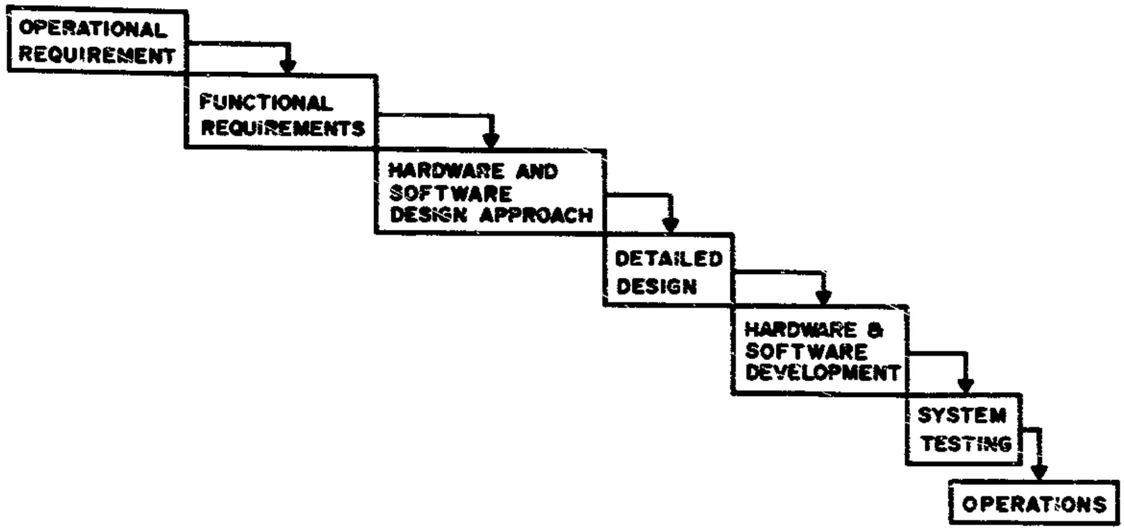

FIGURE 6. THE DEVELOPMENT PROCESS.

with the stated system reliability. Corrections after the system is set intact can be very costly. The system test should serve as the final checkpoint to catch all remaining bugs in the system and incongruencies with the functional specifications.

Implementation Evaluation. The implementation evaluation phase monitors program implementation and insures and documents that the program was implemented strictly according to the design. Any changes made to the design during implementation should be carefully documented and the design revised. The implementation evaluation stage insures that the stated process is operational, intact, and stable. This evaluation is generally accomplished by means of frequent status studies during the implementation stage. Data is collected (usually by surveying the responsible personnel) on each aspect of the process and a determination made about the state of implementation. Direct observations should also be made on a periodic schedule. The status studies are generally made into a report for decision-makers with suggestions to improve or expedite implementation. Shortcomings in implementation are noted in each report. Figure 7 is a flowchart depicting the process of implementation evaluation.

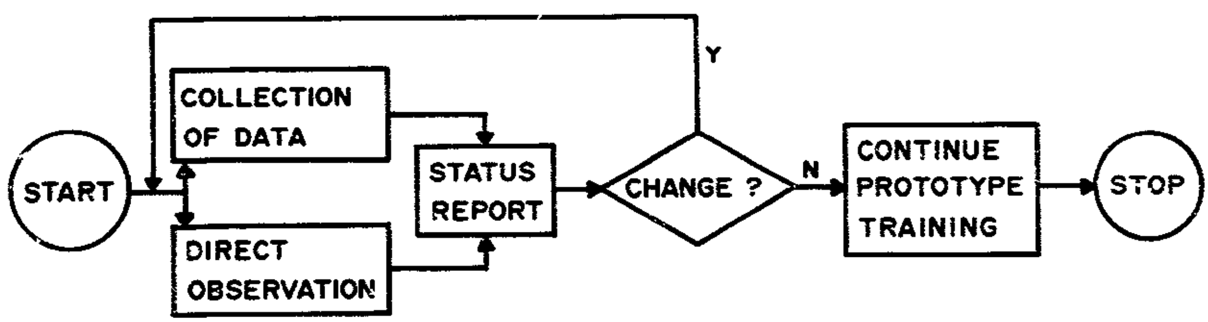

FIGURE 7. FLOWCHART OF IMPLEMENTATION EVALUATION PHASE.

Formative Evaluation. When the program is determined to be operational, intact, and sufficiently stable, formative and summative evaluations begin. Formative evaluation is an ongoing process that insures that the program remains on target. It is the process of continually collecting data and statistics related to training criteria, i.e., how well students are doing in training. This is a monitoring process to gauge the operational stability of the program and the quality of students coming into the program. It is also a method for monitoring compliance with Equal Employment Opportunity Commission (EEOC) guidelines.

The data base for formative evaluation should be extensive. It should contain information for each individual on the current EEOC and Office of Personnel Management (OPM) minority status code, all pertinent attitude information such as expectation and the set/information given to them prior to coming to the Academy, individual and composite scores for selection tests, other information used for points in selection such as education, experience, and veteran's preference, pass/fail information, and all training scores for academic and lab phases. Item responses for all tests during the training phase should also be maintained.

On a periodic basis, statistics and reports should be summarized for research purposes and for transmittal to decision-makers. Statistics should include sample size, means, standard deviations, intercorrelations, pass/fail rates, reliabilities on tests and labs, tests for parallelism on different forms of the same measure, and item parameters, i.e., item difficulty, item discrimination, and the validation of parallel laboratory problems and new items for parallel tests. These statistics should be maintained on record in both computer backup files and hard copy. Further, the statistics should be calculated by input and be cumulative up to and including the most recent input. Administration formative evaluation reports should include sample size, means, and intercorrelations on all relevant measures, and pass/fail rates stratified by minority status, sex, prior experience, predevelopmental/noncompetitive entry, veteran's preference, educational level, option, and region. Appendix A contains sample reports for formative evaluation.

When, based on the formative summary data, there appears to be a problem in how the training program is running, the evaluator has the resonsibility to alert the appropriate administrative personnel and prepare a concise report identifying the problem areas. Isolating the exact area of concern may require some mathematical modeling. The attitude information, where appropriate, should be employed as a covariate in the modeling. Modeling will be discussed in detail later.

Summative evaluation. Summative evaluation is a continual assessment of the quality of the products of the program. While formative evaluation is summarized on an input-by-input basis and serves as an immediate feedback loop for ongoing program revisions if needed, summative evaluation occurs on a larger scale across a longer time span (e.g., on a yearly basis). Formative evaluation is concerned with internal program accuracy and stability, program reliability, and content and/or concurrent demonstrations of validity.(For example, are the measures reliable? Are the objectives well matched with curricula content? Do the pass/fail rates remain stable?) Summative evaluation, however, is a check on the quality of the output from the stabilized program. The summative evaluation is a test of predictive or criterion validity. It is a measure of the on-the-job success of those who pass the Academy training, and the relationship of how well the candidates performed in the Academy compared to how well they performed on-the-job. The so-called validity coefficient is the measure of this relationship.

The summative data base should consist of several components. It is a comprehensive tracking of the career progression of every successful Academy candidate. It should contain data for every individual on types of facilities where the person has been employed, measures of job performance at each of these sites (criterion measures), type of attrition and why, whether a person changed options and why, whether a person was maintained by the agency in a non-2152 (ATCS) position, and as much attitude and demographic information as possible (e.g., divorce, aspects of the job the person likes or dislikes, etc.).

Statistics and reports should be summarized from the summative data base on a regular schedule for research and as information for decision-making. Statistics should include sample sizes, means, standard deviations, intercorrelations, validity coefficients, attrition rates, and

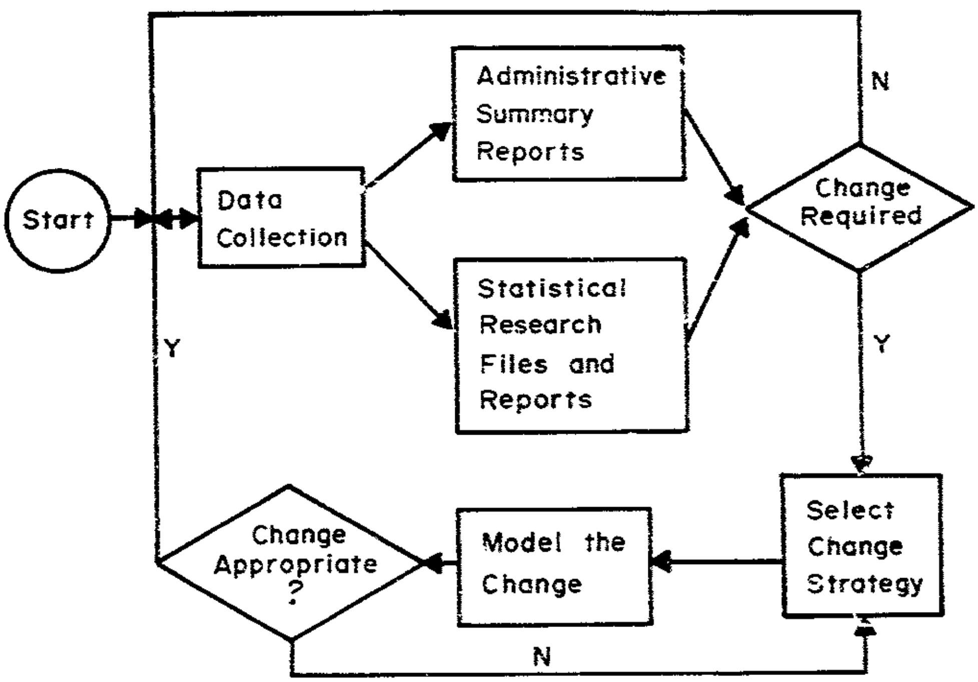

FIGURE 8. FLOWCHART OF THE GENERAL PROCESS FOR BOTH FORMATIVE AND SUMMATIVE EVALUATION.

mathematical modeling, using the attitude information as a covariate. Administrative summative reports should include sample size, means, intercorrelations, and validity coefficients on all relevant data, and attrition data should be stratified by minority status, sex, prior experience, predevelopmental/non-competitive entry, veteran's preference, educational level, reasons for attrition, 2152/non-2152 attrition, option, and region.

If the summative data base demonstrates a problem in the program, a need for a major program revision may be indicated. The data should be reviewed very carefully, employing mathematical modeling to isolate the source of the problem. As in the formative evaluation, the decision-makers should be alerted to the problem but, in addition, in the case of the summative data, policymakers and Academy officials should be alerted. Major program revisions require careful planning and more detailed attention than revisions based on formative data. Appendix B contains an example of the summative reports. Figure 8 flowcharts the general process for both formative and summative evaluation.

<u>The interaction and dynamic nature of the program evaluation components</u>. Figure 9 contains a summary of the four components of the ATCS program evaluation model. The descriptions in Figure 9 imply an interaction between the formative and summative evaluations. The formative evaluation is designed, through constant analyses and feedback mechanisms, to serve as a guidance system in keeping the program on track toward meeting the stated curricular objectives. It serves to stabilize the methods employed in teaching and training the curricular objectives. The summative evaluation is designed to inform policymakers as to whether the methods employed in meeting the curricular objectives and/or whether teaching to these stated objectives actually produces a successful ATCS. If the training methods are not stable or the curricular training objectives are not met, and/or these shortcomings are not detected and corrected within a very short time period by the formative evaluation, it is impossible for the summative evaluation to determine whether the present training methods being employed and/or the present curricular objectives are producing the product being viewed. The interaction between formative and summative evaluation is depicted in Figure 10.

This interaction between summative and formative evaluation has several implications: (i) It implies that a program should be very carefully designed and implemented initially. As previously mentioned, this means performing thorough task, knowledge and skills analyses and a <u>careful matching</u> between the job samples taken from the analyses and the curricular objectives, assessment techniques, and training methods. (ii) The interaction also implies that the summative evaluation assesses how successfully the formative evaluation is working. An unstable program produces confusing and inconsistent summative data. (iii) The last implication relates to program changes: When can program revisions be made; how large a change can be made based on formative and summative data; and what type of evaluation is required given that a change is made.

	DESIGN EVALUATION	IMPLEMENTATION EVALUATION	FORMATIVE EVALUATION	SUMMATIVE EVALUATION
OBJECTIVE	To state goals; develop, define, and document curricula, objectives, modules and assessment tools. Develop and define logically connected step-by-step process to achieve goals. Automated training systems require monitoring the system development.	To insure that the design was fully and correctly implemented and is intact and stable.	To insure that the program stays on target and to add and evaluate refinements changes to the program in a systematic manner. More concerned with program reliability.	To measure the quality of the final program product. What is the program payoff? More concerned with program validity.
METHOD	Careful documentation and description of major systems and subsystems of the program; use of flowcharts, PERT, tables, graphs, and general systems analysis technology.	Frequent status studies with data indicating the extent of implementation for each area of the process. Regular status reports are issued.	Maintenance of an ongoing data base that measures program stability and collection of data that would be sensitive to any change introduced. Regular reports to management.	Collection and analysis of data on the success of graduates of the program. Periodic reports on the data analysis.
RELATION TO DECISION-MAKING	Determining the best plan of action to accomplish the program objectives.	Status reports offer suggestions for improvement and indicate where implementation is falling short.	Offers information on program stability through regular reports and information on the effects of any program change. Also offers information on EEOC compliance.	Offers information on the quality of the program product that is especially useful for long-term planning and needed change.

FIGURE 9. SUMMARY OF THE FOUR COMPONENTS OF THE ATCS PROGRAM EVALUATION MODEL.

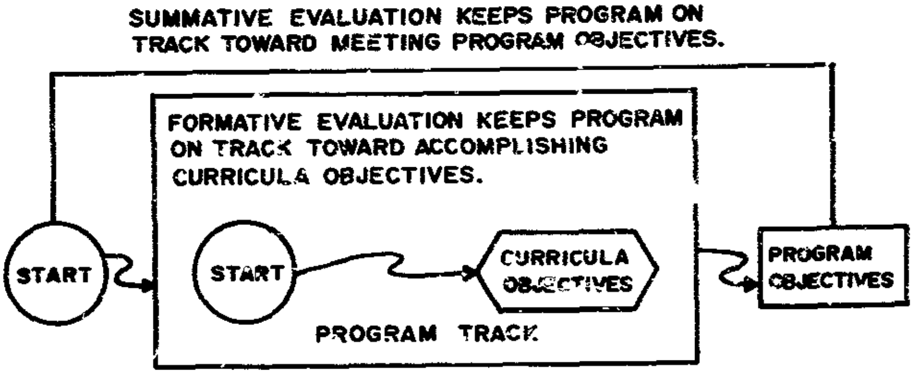

FIGURE 10. INTERACTION BETWEEN FORMATIVE AND SUMMATIVE EVALUATION.

Program changes can be classified as (i) program adjustments, (ii) changing a program component, (iii) adding or subtracting a program component, and (iv) a major program restructuring.

Program adjustments are changes that affect a common element across several program components. They are small or medium changes. Large program adjustments would fall under the category of major restructuring. Program adjustments usually take the form of changes in presentation of lesson material, small curricula adjustments, modifications in the types of assessment devices, or changing the item format in tests. Generally, the formative evaluation process can offer sufficient information to evaluate such a change. In a small number of cases a medium program adjustment may require summative data to evaluate the change.

Changing a program component can vary from a small to a large change. Small changes would include changing items in a component test, reordering the sequence of test items, or minor modifications in a program component curriculum. Small program component changes can be sufficiently evaluated by the formative evaluation data. Medium component changes would include changing the sequence of the component in the program or a medium curriculum change in the component. Medium changes require formative evaluation data and usually also require summative evaluation data. A large change is a major revamping of the component and requires design, implementation, formative, and summative evaluation.

Adding or subtracting a program component, even in a conservative sense, represents a medium or (usually) a large change. In either case, data required to evaluate the effect of adding or subtracting a component include design, implementation, formative, and summative evaluation.

A major restructuring of the program essentially requires the same process as a beginning program and involves design, implementation, formative and summative evaluation. Evaluation of a major restructuring should place more emphasis on the design and implementation evaluation than any of the other types of change. Program changes and the required types of evaluation are summarized in Figure 11.

ACTIVITY	SMALL	MEDIUM	LARGE
Program Adjustments	Formative	Formative Summative	N/A
Change a Program Component	Formative	Formative Summative	Design Implementation Formative Summative
Add/Subtract a Program Component	N/A	Design Implementation Formative Summative	Design Implementation Formative Summative
Major Program Restructuring	N/A	N/A	Design Implementation Formative Summative

FIGURE 11. TYPE OF EVALUATION REQUIRED FOR PROGRAM CHANGES.

The operation of the total model is dynamic and interactive. Each component is dependent on the correct accomplishment of the other components. The dependency, while overlapping, is somewhat linear. Accurate summative evaluation depends on accurate formative evaluation, accurate program implementation, and accurate design. Accurate formative evaluation depends on accurate implementation and design, and so forth back to the task, knowledge and skills analysis used in the design evaluation. Figure 12 is a schematic path diagram of the interaction and dependency among the four components.

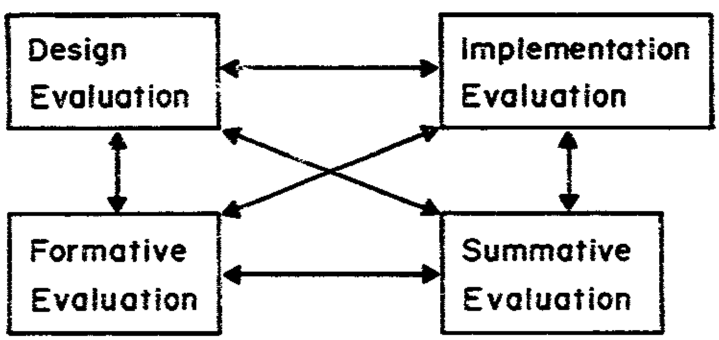

FIGURE 12. THE INTERACTIVE CLOSED LOOP PROGRAM EVALUATION STRUCTURE.

III. Mathematical Models in Formative and Summative Evaluation.

The linear model and intervening variables. The mathematical models used in formative and summative evaluation center on principles of linear regression. The most common phenomenon of interest is how each measure is related to another, i.e., the regression of one measure on another. The simple equation for linear regression is:

$$Y' = a + bX, \tag{1}$$

where X = independent measures, a = the value where the regression line intercepts the Y axis, b = the regression coefficient and Y' = the predicted dependent scores. Two sets of measures (X and Y) can be plotted as in

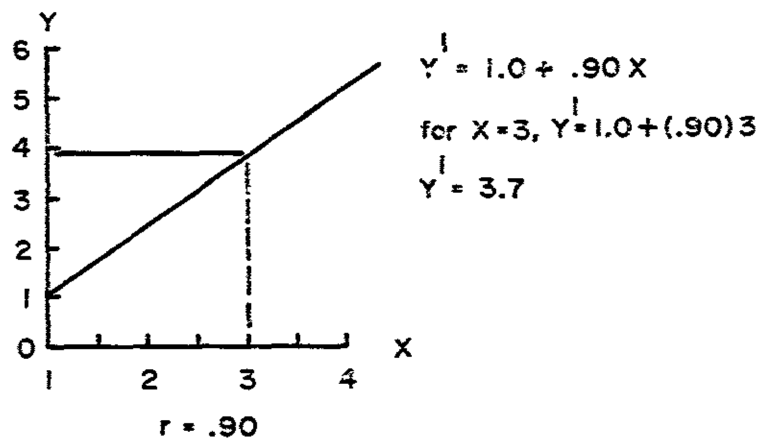

FIGURE 13. PLOT OF TWO SETS OF MEASURES.

Figure 13. As viewed in the plot, the regression line intercepts the Y axis at 1 and the slope of the regression line = .90. The slope indicates the predicted change in Y for each unit change in X. Consequently, it is easily seen how linear regression offers a means to predict or estimate values of one measure from values of another measure. The closer the data points on the graph are to a straight line, the better the prediction. A good example is plotting Academy scores (X) by a success measure on the job.

To explicate, suppose a researcher were to take measures of the same phenomena two times and then plot the two occasions. Identical measurement processes and conditions would yield a graph like Figure 14. The slope would be 45 degrees, placing the intercept through the origin at $a = 0$ and the slope, the increment in Y for a unit change in X, at $b = 1.0$. This result, as noted above, is contingent on two factors in the process being identical: (i) perfectly accurate measures on both occasions, and (ii) identical relevant conditions. If either of these factors were altered, the regression slope and/or intercept most probably would change.

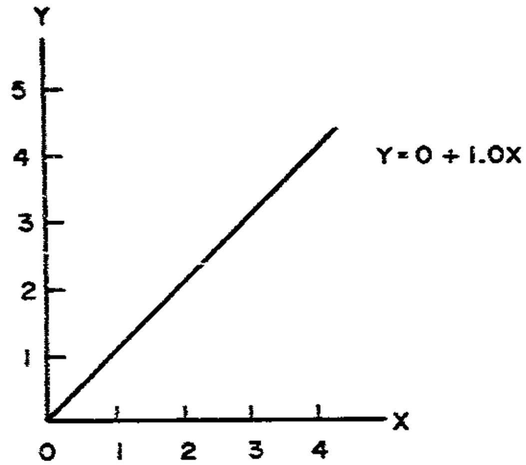

FIGURE 14. PLOT OF TWO IDENTICAL MEASURES.

The factors that would alter the regression line in the above example are referred to as: (i) measurement error and (ii) intervening variables. Figure 15 illustrates the concept.

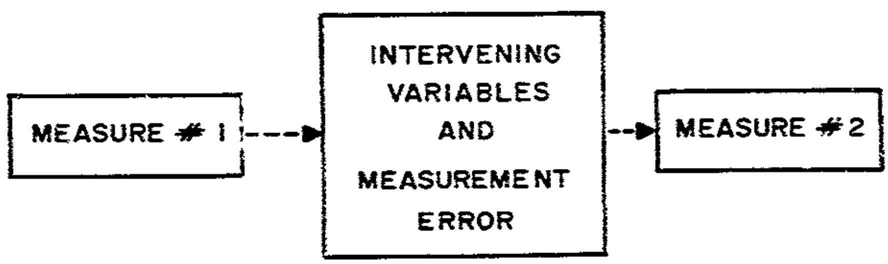

FIGURE 15. THE CONCEPT OF INTERVENING VARIABLES.

Measurement error is usually assumed to be symmetrically distributed about the true value and consequently, when summed, equals zero and has no effect on the analyses.

One of the major uses of the linear model in program evaluation is the identification of intervening variables and determining their impact on dependent variables. For example, what intervening variables affect the relationship between Academy scores and field success (the validity coefficient)? In experimental design the intervening variables are generally viewed as independent variables and the measures affected as dependent variables. Suppose one wanted to determine the impact of motivation (independent variable) on Academy success (dependent variable). Several methods are employed to identify intervening variables and their impact on dependent measures.

Methods for identifying and determining the impact of intervening variables. If one can identify intervening variables and their impact on dependent measures, then a transformation equation can be generated to map X on to Y and the relationship between X and Y can be mathematically explained and quantified. The methods employed to accomplish this depend on the nature of the variables involved and the assumptions one employs in the model.

Assuming the linear model, the process can be explained as follows. Suppose, for example, we have two measures of the same variable, and a plot of the measures appears as in Figure 16.

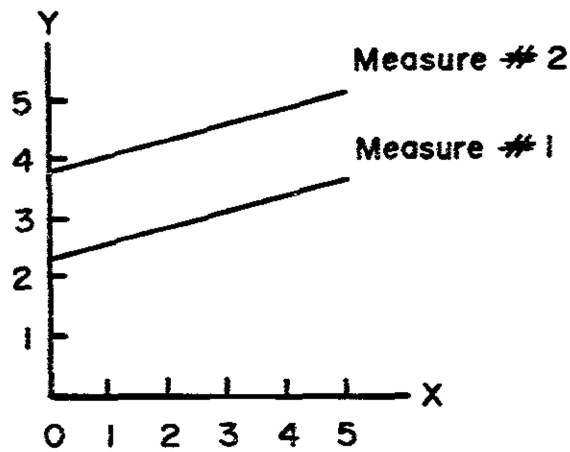

FIGURE 16. PLOT ON MEASURE #1, MEASURE #2, AND REGRESSION LINES.

Further, suppose we suspect that an intervening variable Z was the reason for that difference. It would then stand to logic that if the measures (Y_i) were adjusted to account for the influence of Z then the regression lines should be equal. Returning to the linear equation (1)

$$Y^1 = a + bX, \qquad (1)$$

we can see that the estimated Y, (Y^1), contains some error of estimation unless all the X and Y points lie on a straight line. So,

$$Y^1_i - Y_i = Y(E)_i . \qquad (2)$$

This is the error made in estimating Y for the element i. We can now state an equation for Y_i using the linear model.

$$Y_i = a + bX_i + Y(E)_i \qquad (3)$$

From equations (2) and (3) it is also easily seen that Y(E) is uncorrelated with X since Y(E) is that component in Y which is unexplained by X.

The aim of the adjustment of the measures (Y_i) is to remove the influence (or the relationship) of Z on the measures (Y). Thus applying the linear equation to estimate Y from Z,

$$Y^{1} = A + bZ, \tag{4}$$

we can produce

$$Y^{1}_i - Y_i = Y(E)_i . \tag{5}$$

Y(E) is now a Y value minus the influence of Z. If we plot the Y(E) values separately for M#1 and M#2 as in Figure 15, we have a picture of the plot without the influence of Z. If the regression lines for M#1 and M#2 are identical after removing the influence of Z, then Z can be used to explain the difference in the regression lines for M#1 and M#2 and Z can be used to adjust Y to equate M#1 and M#2 regression lines. This procedure obviously assumes parallel regression lines since it is the intercept that is being adjusted. Consequently, the adjustment to Y can be stated in the linear equation as,

$$Y(A)_i = Y_i - b(Z_i - a) + Y(E)_i , \tag{6}$$

where $Y(a)$ = the adjusted Y for element i. Or, the estimated adjusted Y is,

$$Y(A)^{1}_i = Y_i - b(Z_i - a). \tag{7}$$

The method just explained for removing the influence of Z is a univariate process, checking one variable at a time. Often, intervening variables that account for the differences in regression lines may be correlated. Consequently, the effects they account for are not additive, as seen in Figure 17.

Univariate analyses can help select variables in which their total effects (unique effect + shared effect) are shown; however, if two or more intervening variables share, to a large degree, in their effect on the Y regression lines under two separate conditions, M#1 and M#2, then all the variables are not needed to explain the difference nor to adjust Y. One must employ an analysis that utilizes the unique contribution of variables in producing the differences in M#1 and M#2 without the spurious addition of overlapping effects shared by one or more variables.

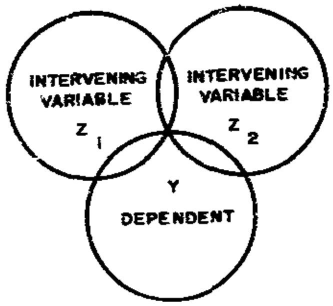

FIGURE 17. OVERLAP OF INTERVENING VARIABLES Z_1 AND Z_2 WITH Y AND THEMSELVES.

<u>Multiple regression</u>. Multiple linear regression is a method of analyzing the shared and unique contributions of more than one independent variable X_i (i=1...k) to the variation of one independent measure, Y (Kerlinger, 1973). By variation we mean how the measures in Y are different from each other. For example, if all Academy candidates were equally successful on-the-job regardless of their Academy scores or selection test scores, then the variance in the success measure is zero and we have no need to analyze the measure.

The model for multiple regression is an extension of our previous single variable regression of X on Y. For an estimated Y, Y^1,

$$Y^1 = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k, \qquad (8)$$

and for Y,

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e. \qquad (9)$$

The procedures in multiple regression are an extension of the previous dicussion on adjusting Y for the influence of a third variable, Z, prior to correlating X and Y. In multiple regression the variance in Y that can be explained by the first variable is partialed out. The remaining variance in Y that can be explained by the second variable without duplicating or overlapping that expressed by the first variable is then partialed out. This process continues until all the independent variables, X_i, have been considered. The relationship of X_i and Y without duplication or overlap among X_i is termed the multiple R. The multiple R squared, R^2, expresses the proportion of variation in Y explained by X_i. If all X_i were uncorrelated, not duplicative, then the multiple R^2 would be the simple sum of all the squared correlations, r (see Figure 13 to review "r"), of X_i and Y.

$$R^2_{y.12...k} = r^2_{y1} + r^2_{y2} + ... + r^2_{yk} \quad (10)$$

However, if the X_i are correlated, then R^2 is the sum of all the r^2 of X_i and Y with the duplication and overlap partialed out.

$$R^2_{y.12...k} = r^2_{y1} + r^2_{y(2.1)} + ... + r^2_{y(k.12...k-1)}, \quad (11)$$

where $r^2_{y(2.1)}$ is read, the correlation of variable 2 and Y with the effects of variable 1 partialed out.

Discriminant analysis. Since the Academy programs are pass/fail, it is a common question to ask which measures best discriminate between passing students and failing students. As an example, suppose we had measures on motivation, level of education, prior experience, and Academy scores and we wanted to know which of these best discriminated between students who pass

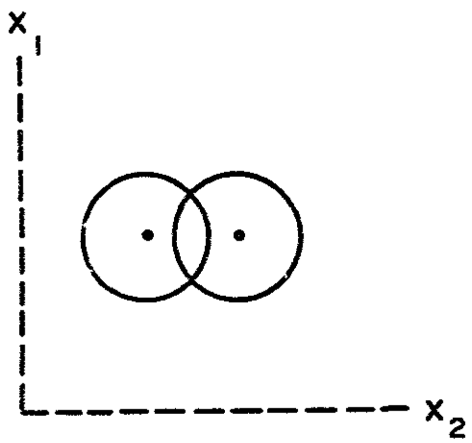

FIGURE 18. PLOTTED DATA FOR THE TWO VARIABLES, TWO GROUP CASE.

field training to full performance level (FPL) and those who fail. Then we would perform a discriminant analysis. To explain this procedure, we go to the most simple case. Suppose we have the simple case of two suspected intervening variables and two groups. If the data for the two groups were expressed on a graph where the axes were the two predictor variables x_1 and x_2, the data could be shown as coordinates of the two variables.

Forming the weighted sum of the intervening variables would create a new variable, Y.

$$Y^{1} = v_1 x_1 + v_2 x_2 , \qquad (12)$$

where v = the weights employed. This may be recognized as another linear equation similar to those previously discussed under "multiple regression." The question is how to express the measure Y on our graph, or, more accurately, how can a Y axis be indicated in the way x_1 and x_2 are? The answer is, the desired axis can be demonstrated by locating the coordinates represented by the two weights, v_1 and v_2, and drawing a line from this point to the origin of the x_1 and x_2 axis (see Figure 19). The data coordinates can now be projected onto the new Y axis as separ te distributions for each group.

The following is a representation of the scheme described above for four different weighted sums of x_1 and x_2.

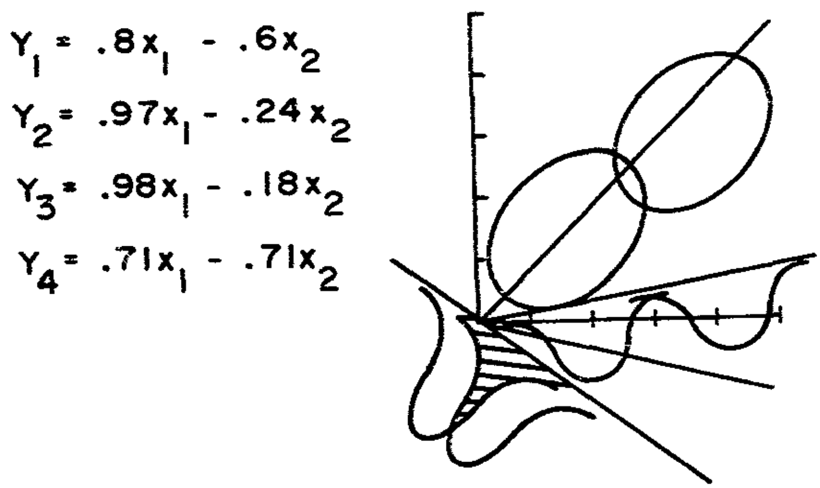

FIGURE 19. PROJECTIONS OF TWO GROUPS ON FOUR AXES REPRESENTING LINEAR COMBINATIONS OF THE ORIGINAL VARIABLES.

It can be noted from the representation that the projected distributions on the Y axes are separated differently. Some of the projected distributions overlap more than others. The problem, then, is to define the Y axis in such a manner that the projected distributions overlap the least. Obviously, in order to do that, a means to measure the overlap must be determined.

One means to define the overlap might be to subtract the means of Y_1 and Y_2 and divide that difference by the standard deviation of one of the groups.

$$\frac{\bar{Y}_1 - \bar{Y}_2}{S_y} \tag{13}$$

However, this would express the difference only in terms of one of the standard deviations.

A more equitable way to do this is to pool the within groups standard deviation. This is accomplished in the following manner.

$$S_{y(w)} = \sqrt{\frac{(n_1 - 1)S_{y1}^2 - (n_2 - 1)S_{y2}^2}{n_1 + n_2 - 2}}, \tag{14}$$

where $S_{y(w)}$ is the pooled within groups standard deviation, n is the group sample size, and S_y^2 is the variance for each group.

Now, a more stable and equitable measure of overlap can be expressed

$$f^2 = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{S_{y(w)}^2}, \tag{15}$$

where f is the measure of overlap in the two distributions.

To extend this measure to more than two groups,

$$f_k^2 = \frac{VAR(\bar{Y})}{S^2_{y(w)}}, \quad (16)$$

where

$$VAR(\bar{Y}) = \frac{\sum_{g=1}^{k} (\bar{Y}_g - \bar{Y}.)^2}{k-1}, \quad (17)$$

with

$$Y. = \frac{\bar{Y}_1 + \bar{Y}_2 + \ldots + \bar{Y}_k}{k}, \quad (18)$$

and

$$S^2_{y(w)} = \frac{\sum_{g=1}^{k} (n_g - 1) S^2_{y(g)}}{N-k}, \quad (19)$$

this being the within groups mean square, $MS_{(w)}$.

In order to take unequal n into account in the numerator above,

$$MS_{(b)} = \frac{\sum_{g=1}^{k} n_g (\bar{Y}_{(g)} - \bar{Y}.)^2}{k-1}, \quad (20)$$

where

$$\bar{Y}. = \frac{\Sigma n_g \bar{Y}_{(g)}}{N}, \quad (21)$$

and $\bar{Y}.$ is the grand mean of Y in the total sample in all k groups.

Collecting things together, we have,

$$f_k^2 = \frac{MS_{(b)}}{MS_{(w)}}, \qquad (22)$$

with $MS_{(b)}$ = mean squares between and $MS_{(w)}$ = mean squares within.

This formula is generally expressed as,

$$f_k^2 = \frac{SS_{(b)}/(k-1)}{SS_{(w)}/(N-k)} = \frac{SS_{(b)}}{SS_{(w)}} X \frac{N-k}{k-1}, \qquad (23)$$

and, since the multipliers (N-k)/(k-1), are constant for any given problem, they can be omitted, yielding

$$h = \frac{SS_{(b)}}{SS_{(w)}}, \qquad (24)$$

where $SS_{(b)}$ = sum of squares between and $SS_{(w)}$ = sum of squares within.

$$SS_{(b)} = \sum_{g=1}^{k} n_g (\bar{Y}_{(g)} - \bar{Y}.)^2, \qquad (25)$$

and

$$SS_{(w)} = \sum_{g=1}^{k} (\bar{Y}_{(g)} - \bar{Y}.)^2$$

$$= \sum_{g=1}^{k} \sum_{i=1}^{n_g} (\bar{Y}_{(g)i} - \bar{Y}_{(g)})^2, \qquad (26)$$

$Y_{(g)i}$ being the Y score of the ith individual in the gth group. The quantity h is termed the criterion.

The problem, then, in performing this analysis is to express the criterion, h, as a function of the weights v_1, v_2, ..., v_p, and to determine by differential calculus the set of weights which maximize h. The weights then express the relative contribution of each intervening variable in explaining the differences in the two groups.

Summary of mathematical models. Basically three examples of linear models were described, as well as the general notion of linear regression. The three examples are by no means inclusive of all linear models; however, the ones presented are the most frequently used in the ATCS program evaluation model. Linear regression models are particularly useful in program evaluation since the major function of any screening program is to best predict on-the-job success and to determine the most efficient subset of measures that can be used to do that. Linear regression is also very useful for estimating the impact of various proposed changes to the program. Without mathematical models, program evaluation would be extremely difficult at best.

IV. Summary.

The Systems Analysis Research Unit at CAMI has developed a generic model for Academy training program evaluation. The model will serve as a basis for integrating the total data base into a common format across all training programs. The model consists of four components; (i) design, (ii) implementation, (iii) formative, and (iv) summative evaluation. Design evaluation is an assessment of the comprehensive implementation plan; implementation evaluation is a determination that the plan is completely and accurately implemented according to prescription; formative evaluation is a continual monitoring of the program to keep the process reliable, stable, and on track; and summative evaluation monitors the product of the training program. The design evaluation relies on the task, knowledge, and skills analysis and on the documents in the implementation plan. The implementation evaluation makes use of data from frequent status studies. Formative and summative evaluations make use of statistics and mathematical modeling, primarily linear regression models, to monitor the process and products of the programs and to estimate and determine the impact of changes made to the programs.

REFERENCES

1. Kerlinger, F. N.: Multiple Regression in Behavioral Research. New York: Holt, Rhinehart, and Winston, 1973.

2. Page, L. F.: Technology in Air Traffic Control Training and Simulation. PROCEEDINGS OF THE THIRD INTERNATIONAL LEARNING TECHNOLOGY CONGRESS AND EXPOSITION, 2:212-216, 1980.

3. Stufflebeam, D. L., et al.: Educational Evaluation and Decision Making. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1971.

APPENDIX A

Examples of Reports for Formative Evaluation

CAMI, SYSTEMS ANALYSIS RESEARCH UNIT
9-JUN-81 REPORT NUMBER BURRE12

PERFORMANCE BY REGION FOR 1980 ENROUTE INPUT 12 STARTING 10-FEB-81

REGION	N	PASS		FAIL		WITHDRAW		NO SHOW		RADAR COMPOSITE		
		#	%	#	%	#	%	#	%	AVERAGE	SD	N
ALASKA	0	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.00	0
EASTERN	5	4	80.00	1	20.00	0	0.00	0	0.00	75.20	10.43	5
CENTRAL	4	3	75.00	1	25.00	0	0.00	0	0.00	71.50	3.32	4
GREAT LAKES	11	6	54.55	5	45.45	0	0.00	0	0.00	71.82	5.19	11
NEW ENGLAND	1	1	100.00	0	0.00	0	0.00	0	0.00	71.00	0.00	1
NORTHWEST	0	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.00	0
PACIFIC-ASIA	1	0	0.00	1	100.00	0	0.00	0	0.00	52.00	0.00	1
ROCKY MOUTAIN	4	4	100.00	0	0.00	0	0.00	0	0.00	77.00	5.94	4
SOUTHERN	0	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.00	0
SOUTHWEST	0	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.00	0
WESTERN	3	3	100.00	0	0.00	0	0.00	0	0.00	77.00	6.00	3
TOTAL	29	21	72.41	8	27.59	0	0.00	0	0.00	72.90	7.37	29

CAMI, SYSTEMS ANALYSIS RESEARCH UNIT
9-JUN-81 REPORT NUMBER 80RXE12

MEANS BY CLASS FOR 1980 ENROUTE INPUT 12 STARTING 10-FEB-81

	8105			8106			8107			8108			TOTAL INPUT		
	N	MEAN	SD	N	MEAN	SD	N	MEAN	SD	N	MEAN	SD	N	MEAN	SD
BLOCK TEST 1	9	94.00	4.58	7	88.86	5.52	5	92.40	3.58	8	85.50	6.21	29	90.14	6.02
BLOCK TEST 2	9	89.78	4.63	7	84.86	6.09	5	88.40	2.61	8	86.00	5.45	29	87.31	5.16
RADAR CPT	9	85.56	4.77	7	83.71	5.71	5	85.60	6.23	8	82.25	9.41	29	84.21	6.58
PROB 1 TECHNICAL ASSESSMT	9	72.11	13.38	7	60.86	13.64	5	62.60	24.03	8	72.00	16.39	29	67.72	16.39
PROB 1 INSTRUCTOR ASSESMT	9	86.78	8.48	7	66.29	15.93	5	73.60	24.85	8	87.00	9.12	29	79.62	16.34
PROB 1 LAB SCORE	9	79.33	10.07	7	63.71	13.92	5	68.00	24.29	8	79.50	12.57	29	73.66	15.65
PROB 2 TA	9	56.11	18.85	7	58.71	16.13	5	48.60	21.85	8	69.13	16.83	29	59.03	18.64
PROB 2 IA	9	79.00	12.36	7	69.71	17.90	5	70.20	25.40	8	83.75	11.29	29	76.55	16.49
PROB 2 LS	9	67.67	13.75	7	64.43	16.18	5	59.60	19.42	8	76.50	12.44	29	67.93	15.47
PROB 3 TA	9	54.67	21.26	7	52.86	12.76	5	59.80	29.58	8	66.25	11.34	29	58.31	18.74
PROB 3 IA	9	73.89	21.18	7	78.29	10.67	5	72.80	26.18	8	85.50	6.99	29	77.97	17.00
PROB 3 LS	9	64.44	20.96	7	65.43	11.37	5	66.20	27.55	8	76.00	8.83	29	68.17	17.48
PROB 4 TA	9	55.11	11.04	7	58.14	15.87	5	63.40	15.85	8	62.13	21.17	29	59.21	15.75
PROB 4 IA	9	80.00	12.87	7	78.57	11.57	5	71.40	13.87	8	81.88	11.64	29	78.69	12.26
PROB 4 LS	9	67.56	10.55	7	68.29	12.94	5	67.60	14.57	8	72.13	16.13	29	69.00	12.92
PROB 5 TA	9	54.33	22.14	7	56.14	23.19	5	46.80	27.49	8	61.50	21.11	29	55.45	22.34
PROB 5 IA	9	71.78	16.45	7	72.14	18.71	5	69.00	19.54	8	83.13	8.90	29	74.52	16.03
PROB 5 LS	9	63.22	18.83	7	64.00	20.77	5	58.00	23.02	8	72.25	14.69	29	65.00	18.67
AVERAGE TA	9	58.33	6.75	7	57.14	4.71	5	56.20	19.66	8	66.00	12.24	29	59.79	11.23
AVERAGE IA	9	78.33	4.39	7	73.00	6.22	5	71.40	20.07	8	84.25	6.80	29	77.48	10.38
LAB AVERAGE (ALL 5)	9	66.78	4.58	7	65.43	6.40	5	63.60	19.37	8	74.38	8.98	29	68.00	10.28
RADAR CST	9	85.11	5.58	7	80.57	6.19	5	84.40	7.13	8	81.25	5.65	29	82.83	6.04
LAB AVERAGE (BEST 4)	9	73.33	5.70	7	69.00	4.86	5	66.80	18.75	8	77.75	8.99	29	72.38	10.09
LAB COMPOSITE SCORE	9	73.11	4.08	7	70.57	5.19	5	70.40	14.64	8	76.25	5.47	29	72.90	7.37

CAMI, SYSTEMS ANALYSIS RESEARCH UNIT
9-JUN-81 REPORT NUMBER 80REE12

RADAR LAB STATUS BY ENTRY TYPE AND BY MINORITY CATEGORY FOR 1980 ENROUTE INPUT 12 STARTING 10-FEB-81

	PASS % = # PASSED / # ENTERED								FAIL % = # FAILED / # ENTERED							
	ALL		COMPETITIVE		PREDEV		CEP		ALL		COMPETITIVE		PREDEV		CEP	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
MEN	21.	77.78	20.	80.00	1.	100.00	0.	0.00	6.	22.22	5.	20.00	0.	0.00	1.	100.00
WOMEN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	2.	100.00	2.	100.00	0.	0.00	0.	0.00
MINORITIES	2.	66.67	1.	100.00	1.	100.00	0.	0.00	1.	33.33	0.	0.00	0.	0.00	1.	100.00
NONMINORITIES	19.	73.08	19.	73.08	0.	0.00	0.	0.00	7.	26.92	7.	26.92	0.	0.00	0.	0.00
MEN MIN	2.	66.67	1.	100.00	1.	100.00	0.	0.00	1.	33.33	0.	0.00	0.	0.00	1.	100.00
MEN NONMIN	19.	79.17	19.	79.17	0.	0.00	0.	0.00	5.	20.83	5.	20.83	0.	0.00	0.	0.00
WOMEN MIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00
WOMEN NONMIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	2.	100.00	2.	100.00	0.	0.00	0.	0.00
TOTAL	21.	72.41	20.	74.07	1.	100.00	0.	0.00	8.	27.59	7.	25.93	0.	0.00	1.	100.00

	WITHDRAW/NO SHOW % = # WD OR NS / # ENTERED								ALL ENTRIES % = # ENTERED / TOTAL # ENTERED							
	ALL		COMPETITIVE		PREDEV		CEP		ALL		COMPETITIVE		PREDEV		CEP	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
MEN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	27.	93.10	25.	92.59	1.	100.00	1.	100.00
WOMEN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	2.	6.90	2.	7.41	0.	0.00	0.	0.00
MINORITIES	0.	0.00	0.	0.00	0.	0.00	0.	0.00	3.	10.34	1.	3.70	1.	100.00	1.	100.00
NONMINORITIES	0.	0.00	0.	0.00	0.	0.00	0.	0.00	26.	89.66	26.	96.30	0.	0.00	0.	0.00
MEN MIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	3.	10.34	1.	3.70	1.	100.00	1.	100.00
MEN NONMIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	24.	82.76	24.	88.89	0.	0.00	0.	0.00
WOMEN MIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00	0.	0.00
WOMEN NONMIN	0.	0.00	0.	0.00	0.	0.00	0.	0.00	2.	6.90	2.	7.41	0.	0.00	0.	0.00
TOTAL	0.	0.00	0.	0.00	0.	0.00	0.	0.00	29.	100.00	27.	100.00	1.	100.00	1.	100.00

CAMI, SYSTEMS ANALYSIS RESEARCH UNIT
9-JUN-81 REPORT NUMBER 80RFE12

4 OF 5 SCORING IMPACT ON RADAR LAB FOR 1980 ENROUTE INPUT 12 STARTING 10-FEB-81

	PASS				FAIL				WD/NS		TOTAL
	5 OF 5		4 OF 5		5 OF 5		4 OF 5				
	#	%	#	%	#	%	#	%	#	%	#
MEN	21	77.78	24	88.89	6	22.22	3	11.11	0	0.00	27
WOMEN	0	0.00	2	100.00	2	100.00	0	0.00	0	0.00	2
MINORITIES	2	66.67	3	100.00	1	33.33	0	0.00	0	0.00	3
NONMINORITIES	19	73.08	23	88.46	7	26.92	3	11.54	0	0.00	26
COMPETITIVE	20	74.07	24	88.89	7	25.93	3	11.11	0	0.00	27
PREDEVELOPMENTAL	1	100.00	1	100.00	0	0.00	0	0.00	0	0.00	1
COOPERATIVE ED	0	0.00	1	100.00	1	100.00	0	0.00	0	0.00	1
TOTAL	21	72.41	26	89.66	8	27.59	3	10.34	0	0.00	29

CAMI, SYSTEMS ANALYSIS RESEARCH UNIT
9-JUN-81 REPORT NUMBER 80RGE12

RADAR LAB GRADE RANGES BY SUBGROUPS FOR 1980 ENROUTE INPUT 12 STARTING 10-FEB-81

GRADE RANGE	TOTAL GRP #	TOTAL GRP %	MEN #	MEN %	WOMEN #	WOMEN %	MINORITY #	MINORITY %	NONMIN #	NONMIN %	MEN MIN #	MEN MIN %	MEN NONMIN #	MEN NONMIN %	WOMEN MIN #	WOMEN MIN %	WOMEN NONM #	WOMEN NONM %
90-100	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
80-89	3	10.34	3	11.11	0	0.00	1	33.33	2	7.69	1	33.33	2	8.33	0	0.00	0	0.00
75-79	9	31.03	9	33.33	0	0.00	1	33.33	8	30.77	1	33.33	8	33.33	0	0.00	0	0.00
70-74	9	31.03	9	33.33	0	0.00	0	0.00	9	34.62	0	0.00	9	37.50	0	0.00	0	0.00
65-69	6	20.69	4	14.81	2	100.00	1	33.33	5	19.23	1	33.33	3	12.50	0	0.00	2	100.00
60-64	1	3.45	1	3.70	0	0.00	0	0.00	1	3.85	0	0.00	1	4.17	0	0.00	0	0.00
50-59	1	3.45	1	3.70	0	0.00	0	0.00	1	3.85	0	0.00	1	4.17	0	0.00	0	0.00
LT 50	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TOTALS	29	100.00	27	100.00	2	100.00	3	100.00	26	100.00	3	100.00	24	100.00	0	0.00	2	100.00

GRADE RANGE	COMPETITIV #	COMPETITIV %	PREDEV #	PREDEV %	CEP #	CEP %	ATC EXP #	ATC EXP %	NO ATC EXP #	NO ATC EXP %	VET PREF #	VET PREF %	NO VET PRF #	NO VET PRF %	COL DEGREE #	COL DEGREE %	NO DEGREE #	NO DEGREE %
90-100	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
80-89	3	11.11	0	0.00	0	0.00	3	20.00	0	0.00	3	18.75	0	0.00	0	0.00	3	16.67
75-79	8	29.63	1	100.00	0	0.00	5	33.33	3	23.08	6	37.50	3	25.00	3	30.00	5	27.78
70-74	9	33.33	0	0.00	0	0.00	5	33.33	4	30.77	6	37.50	2	16.67	2	20.00	7	38.89
65-69	5	18.52	0	0.00	1	100.00	2	13.33	4	30.77	1	6.25	5	41.67	3	30.00	3	16.67
60-64	1	3.70	0	0.00	0	0.00	0	0.00	1	7.69	0	0.00	1	8.33	1	10.00	0	0.00
50-59	1	3.70	0	0.00	0	0.00	0	0.00	1	7.69	0	0.00	1	8.33	1	10.00	0	0.00
LT 50	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TOTALS	27	100.00	1	100.00	1	100.00	15	100.00	13	100.00	16	100.00	12	100.00	10	100.00	18	100.00

APPENDIX B

EXAMPLE OF A SUMMATIVE REPORT

EN ROUTE TRACKING STUDY

GROUP	SIZE	PERCENT
PASS ACADEMY	824	88.0%
FAIL ACADEMY	80	8.5%
NO SHOW OR WITHDRAW	32	3.4%
TOTAL	936	100.0%
STILL ACTIVE IN 2152 (OF PASSES)	698	84.7%
NOT ACTIVE IN 2152 (OF PASSES)	126	15.3%
TOTAL PASSES	824	100.0%
TOTAL STILL ACTIVE	698	74.6%
TOTAL NOT ACTIVE	238	25.4%
TOTAL DEVELOPMENTALS	936	100.0%

FOR THOSE STILL ACTIVE IN 2152 OPTION (N=698):

LAB PHASE COMPOSITE SCORE = 81.7
SUPERVISOR RATING SCORE = 5.1

CORRELATION OF LAB COMPOSITE WITH
SUPERVISOR RATING = .212
CORRECTED FOR RESTRICTION = .295

TERMINAL TRACKING STUDY

GROUP	SIZE	PERCENT
PASS ACADEMY	868	89.3%
FAIL ACADEMY	85	8.7%
NO SHOW OR WITHDRAW	19	2.0%
TOTAL	972	100.0%
STILL ACTIVE IN 2152 (OF PASSES)	789	90.9%
NOT ACTIVE IN 2152 (OF PASSES)	79	9.1%
TOTAL PASSES	868	100.0%
TOTAL STILL ACTIVE	789	81.2%
TOTAL NOT ACTIVE	183	18.8%
TOTAL DEVELOPMENTALS	972	100.0%

FOR THOSE STILL ACTIVE IN 2152 OPTION (N=789):

LAB PHASE COMPOSITE SCORE = 81.2
SUPERVISOR RATING SCORE = 5.3

CORRELATION OF LAB COMPOSITE WITH
SUPERVISOR RATING = .245
CORRECTED FOR RESTRICTION = .334

☆U.S. GOVERNMENT PRINTING OFFICE: 1982 369-606/741 1-3