AD-P003 455

DETERMINING COST AND TRAINING EFFECTIVENESS TRADEOFFS FOR TRAINER DESIGN: TEST OF AN EXPERIMENTAL MODEL

Dr. Ruth A. Wienclaw
Honeywell Inc.
West Covina, CA

Dr. Jesse Orlansky
Institute for Defense Analyses
Alexandria, VA

ABSTRACT

This paper reports the status of an ongoing project to develop a macro model describing the decisions involved in developing training equipment. The purpose of the model is to assist managers in making such decisions by providing information concerning the tradeoffs between the cost and effectiveness of training provided by different configurations and choices of equipment. The goals of the current phase of the study were to determine the feasibility of collecting data to empirically test the model and turn it into a practical tool to be used in making decisions relating to trainer design and development, and to perform a preliminary test of the model.

Results of the field data collection led to the conclusion that the data necessary to test the model can be obtained. However, such measures need to be refined before the model can be turned into a practical tool. The preliminary test of the model performed in this study resulted in no major modifications of the model.

PURPOSE

This paper reports the status of an ongoing project to develop a practical model to assist managers in making decisions concerning training equipment. The model is designed to permit comparisons of the cost and effectiveness of alternative configurations of training equipment. The model will allow managers to make cost/benefit tradeoffs between the various characteristics that may be utilized in training equipment, and give both the military and industry guidelines to justify decisions relating to trainer design.

The purpose of the current phase of the study was two-fold:

(1) To determine the feasibility of data collection for empirical validation of the model, and

(2) To perform a preliminary test of the model.

BACKGROUND

In their 1982 Summer Study on Training and Training Technology, the Defense Science Board concluded that consideration must be given to training effectiveness during the design and development of military training systems. This conclusion was based on the fact that effective training is essential in order to maintain operational readiness (1). In a recent evaluation of operational trainers, the General Accounting Office stated that most training equipment is designed without due consideration to training effectiveness (2). It is essential, therefore, that tools be developed so that the effectiveness of training equipment can be estimated during the design and development phases. In this way, knowledgeable tradeoffs can be made between both the cost and effectiveness of training equipment during the development process.

The Air Force Human Resources Laboratroy concluded recently that no such practical model exists as yet (3). Such a model is necessary to aid both the Armed Services and industry in defining the requirements needed to design training systems that training personnel adequately and at a minimum cost. The current study is part of an ongoing project to help meet this need.

Development of an Experimental Model

A preliminary version of a model was described last year (4). All terms, such as "effectiveness" and "cost" were defined operationally for measurement both at school and on the job; cost included acquisition, operation, and support. A taxonomy was provided to describe the characteristics of a training device. Provisions were also made for specifying relevant characteristics of students, instructors, and training goals. A graphic depiction of the model is given in Figure 1.

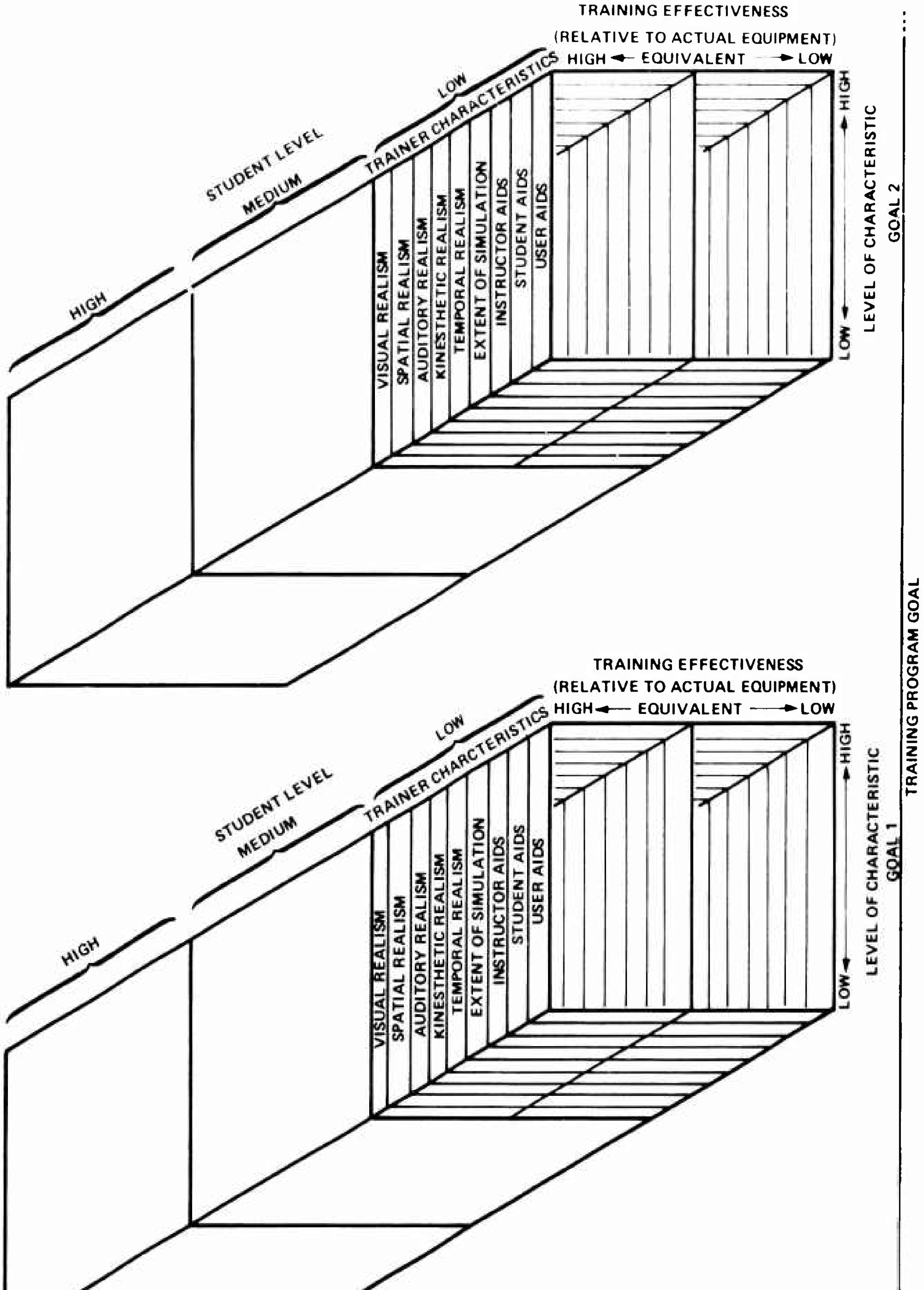

Figure 1. Revised Model for Determining Relative Effectiveness of Two Training Methods

PRELIMINARY TEST OF THE MODEL

Purpose

The purpose of this test was to examine the feasibility of collecting data needed to develop and use the model. The problem was to compare the effectiveness of training of F-16 maintenance technicians when either simulators or actual equipment had been used in training courses. Data were collected at Hill Air Force Base, Utah, and Hahn Air Base, Germany, where simulators were used since August 1979 and August 1981, respectively; data were also collected at Nellis Air Force Base, Nevada, where simulators had never been used.

The F-16 simulators consist of six free-play systems designed to assist in teaching maintenance courses in flight controls, communication, navigation, and electrical systems, and engine start, engine diagnostics and engine run for the F-16 aircraft.

Data Collection Instruments. Two sets of data collection instruments were used. A set of Behaviorally Anchored Rating Scales (BARS) was developed to assess technicians' performance in the field. Instructors, in the role of subject matter experts, were asked to create a series of critical incidents describing behaviors which differentiate between a good technician and a poor one. The incidents focused on specific technician actions closely related to the job, and differentiated between success and failure as a maintenance technician. The incidents were rated by the instructors on a seven-point scale with the scale value of 1 being very poor performance behavior and the scale value of 7 being very high performance behavior. Those incidents with the lowest standard deviations and means closest to 1 and 7 were then placed on a graphic type rating scale to be used as behavioral anchors for the scales.

There are two advantages to using BARS: first of all, the description of the scale points is written in terms that can be easily understood by the raters. Second, since the type of person who developed the scale is also the type of person who uses the scale, the raters have a vested interest in using the scales correctly (5).

The use of the BARS development technique in this study yielded seven specific scales:

(1) Safety: Behaviors which show that the technician understands and follows safety practices as specified in the technical data;

(2) Thoroughness and Attention to Details: Behaviors which show that the technician is well prepared when he arrives on the job, carries out maintenance procedures completely and thoroughly, and recognizes and attends to symptoms of equipment damage or stress;

(3) Use of Technical Data: Behaviors which show that the technician properly uses technical data in the performance of maintenance functions;

(4) System Understanding: Behaviors which show that the technician thoroughly understands system operation allowing him to recognize, diagnose, and correct problems not specifically covered in the technical data and publications;

(5) Understanding of Other Systems Behaviors which show that the technician understands the systems that interconnect with his specific system and can operate them in accordance with the technical data;

(6) Mechanical Skills: Behaviors which show that the technician possesses specific mechanical skills required for even the most difficult maintenance problems; and

(7) Attitude: Behaviors which show that the technician is concerned about properly completing each task efficiently and on time.

The second data collection instrument was a series of questionnaires for students, instructors, and technicians. These questionnaires were used to collect two types of information: (1) demographic information concerning respondents' background, training, and experience, and (2) subjective information such as respondents' attitudes toward training devices in general, and their perceptions and evaluations of the specific device with which they were working.

Data on training effectiveness were collected through student course test scores and Work Unit Code (WUC) information. A new system known as the Consolidated Data System (CDS) has been instituted for the F-16 aircraft that allows for more flexible and responsive maintenance data reporting than was previously available. This system relies on maintenance information recorded by each work center on Air Force Form AFTO 349, "Maintenance Data Collection Record," which is entered into the data base at each wing. The advantage of using the CDS is the ability to aggregate maintenance data in a more usable form and with flexibility as to which information is displayed. The information needed for the current study was which component was worked on (WUC), what action was taken, the time necessary to complete the action, and what work center performed the action.

The purpose was to determine whether maintenance data records, collected routinely, might provide data on the effectivness of maintenance training. In this case the issue was to see if technicians trained with simulators performed differently from technicians not trained with simulators.

Procedure. Data were collected using the three versions of the questionnaire (student, instructor, technician) to gather background data concerning the subjects and their opinions of training courses and devices. The BARS were used to determine the instructors' and supervisor's performance appraisals of those students having previously graduated the courses. This repeated use of the BARS was intended to help determine the validity of such subjective judgments, and to partially ascertain the relationship between judgments of technician performance at the school and the field levels. The distributions of subjects receiving these instruments are given in Tables 1 and 2.

Table 1. Breakdown of Questionnaire Respondents by Base and Status

	Hill	Hahn	Nellis
Instructors	8	15	10
FTD Students	26	13	19
Technicians	38	15	8

Table 2. Breakdown of Performance Assessments by Base and Status

	Hill	Hahn	Nellis
Current FTD Students	17	7	19
Current Technicians	44	13	10
Past FTD Students	0	11	37

The procedure for determining the maintenance productivity of specific work centers started with the choice of the specific work unit codes to be examined. Although there are six F-16 trainers which are of interest in this study, due to the small number of observations associated with the action codes included in several of the WUC's, it was not possible to gather sufficient data to analyze all six of these trainers. Two system-level WUC's were found to have a sufficient number of action code observations to be included. These were 14000, Flight Control Systems, which is applicable to courses in flight controls and instrumentation, and 23000, Turbofan Power Plant, which is applicable to courses in engine diagnostics. Within these system-level WUC's, two component-level WUC's were chosen for further analysis: 14A00, Primary Flight Control Electronics, and 23Z00, Turbofan Power Plant (F-100 engine). These WUC's were chosen for analysis because they were directly related to actions taught on the maintenance trainers and the number of observations was sufficient for analysis.

Results

Correlations between BARS ratings made by supervisors of technician's performance in the field and the ratings made by instructors of the same technician's performance in the school setting are shown in Table 3. The correlations are low, indicating that the success of a technician as measured by the BARS cannot be predicted from his performance in the training course. The only performance measure that shows a statistically significant relationship between school and field performance is the scale measuring "Use of Technical Data" (r = .5). This correlation indicates only a moderate relationship, with 25% (r^2) of the variance of the technicians' scores accounted for by their student scores. These results suggest that such ratings of student performance in school settings do not provide a valid predictor of performance in the field.

A repeated measures analysis of variance (ANOVA) on the BARS data suggested that there is an improvement in performance over time for both types of training (i.e., trainer or actual equipment) after the students graduate and perform maintenance procedures in the field. The one exception to this is the "Understanding of Other Systems" scale (see Table 4). Ratings given to the technicians who were trained using the actual equipment appear to be consistently higher than for those trained using the trainers (Figure 2). The ratings for technicians trained by the two methods, however (once again with the exception of the scale "Understanding of Other Systems"), appear to converge over time. The average length of time between course graduation and supervisor performance rating was three and a half months. This suggests that improvement in performance produced by different training methods dissipates as on-the-job training increases. The goal of devising a more effective training system, therefore, actually becomes one of producing competent technicians faster than by other methods.

Table 3. Correlations Between Performance Ratings (BARS) of Course Graduates as Students and as Technicians

Performance Measure	Pearson r
Safety	0.221
Thoroughness	0.169
Use of Technical Data	0.526*
System Understanding	0.381
Understanding of Other Systems	0.111
Mechanical Skills	0.156
Attitude	0.328

*$p \leq .05$
$N = 18$

Table 4. F Values for BARS Score Improvement Over Time (Repeated Measures ANOVA)

Scale	F Value	p
Safety	5.24	.036
Thoroughness	4.10	.060
Use of Technical Data	6.07	.026
System Understanding	4.16	.058
Understanding of Other Systems	3.22	.092
Mechanical Skills	6.26	.024
Attitude	13.92	.002

Although the differences between ratings of course graduates as students and as technicians are not all statistically significant at the $\alpha = .05$ level, the trends are clear. Statistical significance is difficult to achieve with small samples, even though an underlying trend may indeed exist. It is important, also, to note that the average performance rating for both groups is above the 4.0 midpoint (halfway between the 1.0 minimum and the 7.0 maximum performance ratings) for each of the seven measures. This may be interpreted to mean that both types of training are producing at least satisfactory technician performance.

In the analysis of the WUC's data for WUC 14A00 it was found that the greater the degree of worker training, the better the productivity of the unit (see Figure 3). This was true for 17 out of the 18 data points. Only the remove after cannibalization action showed a slight reversal of this pattern. For WUC 23Z00, the same trend was found. In this comparison, training appears to bear some relationship to productivity (see Figure 4). The two highest trained at 74 percent, were both more productive for each action code than the least trained base, at 59 percent.

Discussion

The first goal of the current study was to determine the feasibility of collecting practical data for use in validation of the model. The work unit code data show promise for being valid on-the-job measures of training effectiveness. This can be seen in the comparisons of performance of technicians trained in FTD courses *versus* those who had not received such formal training. However, the WUC's data as used in the current study need refinement for use as a measure of training effectiveness.

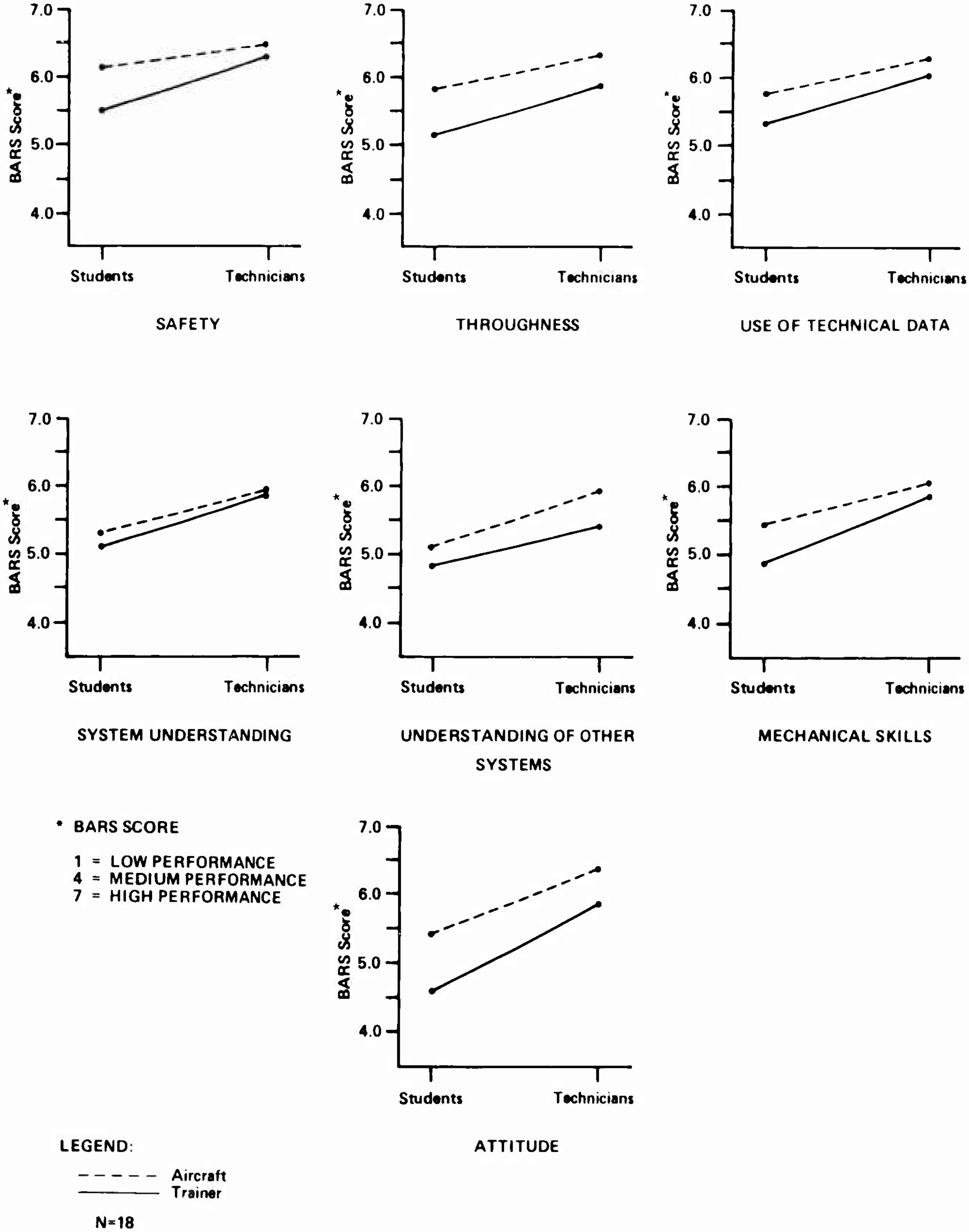

Figure 2. Cell Means for BARS Repeated Measures ANOVA Between Course Graduates as Students and as Technicians

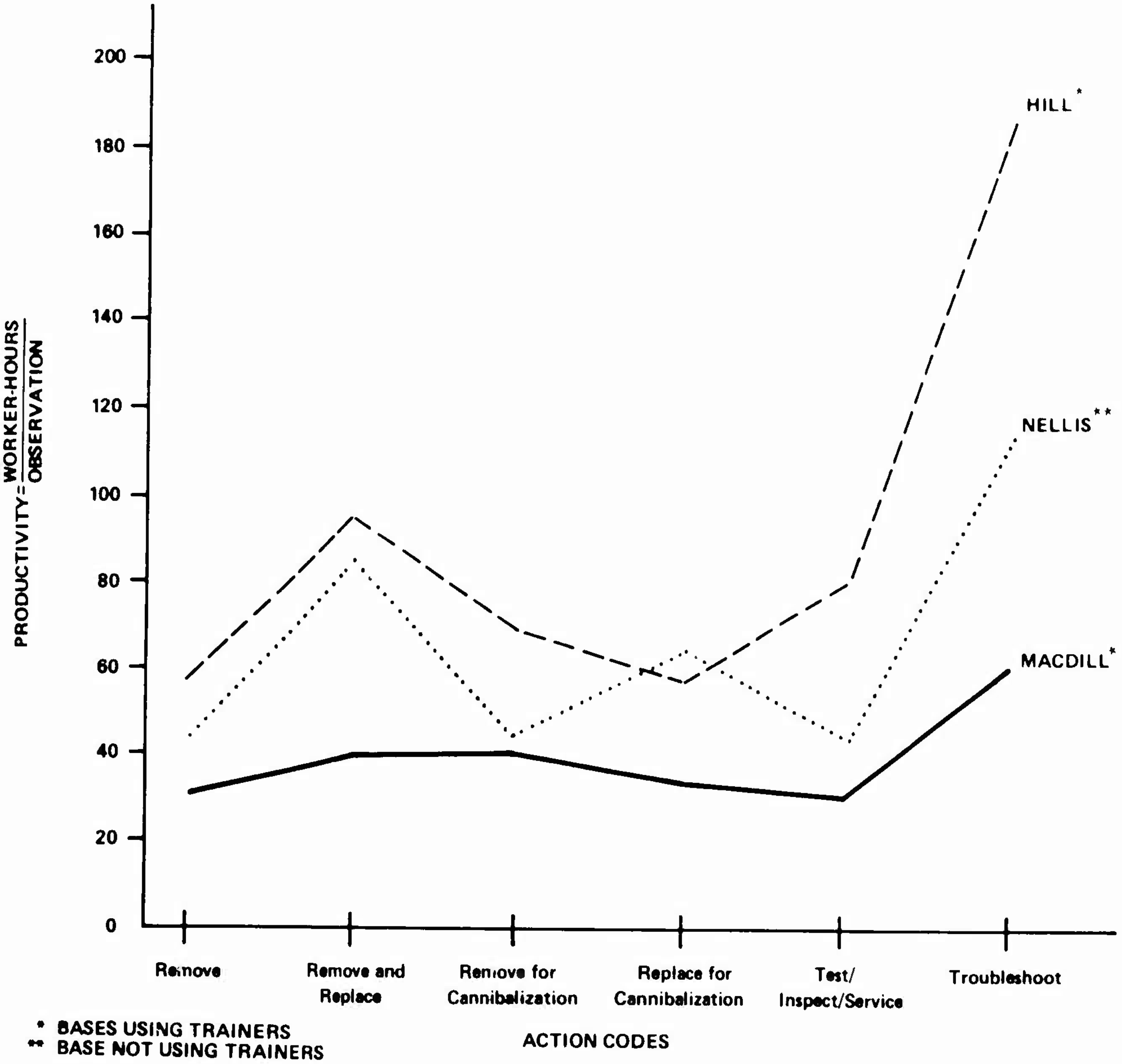

Figure 3. Training: Wing-to-Wing (14A00)

Some problems inherent to the Air Force maintenance data collection system may limit its accuracy as a measure of performance relevant to training. For example, problems with data recording and data entry could lead to biases or inaccurate data analyses. Second, while WUC's that are associated with actions taught using maintenance simulators can be identified, these WUC's tend to be very specific. In the refinement of the WUC's data as measures of field performance it will, therefore, be necessary to take this fact into consideration in order to develop the most useful measure of training effectiveness possible. Finally, in the interpretation of any data based upon WUC's information, one must also consider the environmental influences on maintenance technicians which occur in the field and which may lead to differences in performance between bases that are not due to training (6).

The second goal of the current study was to perform a preliminary test of the training effectiveness model. The BARS data (i.e., rating scales) suggested differences in th performance on the job between students trained on simulators or actual equipment. These differences lessen as on-the-job training increases. However, only ratings on one measure ("Use of Technical Data") at school are correlated significantly with the same measure on the job. Several questions need to be answered. First, do the technicians trained with the trainers eventually perform as satisfactorily as those trained with the actual equipment? Second, is the additional on-the-job training necessary to bring a trainer-taught technician up to

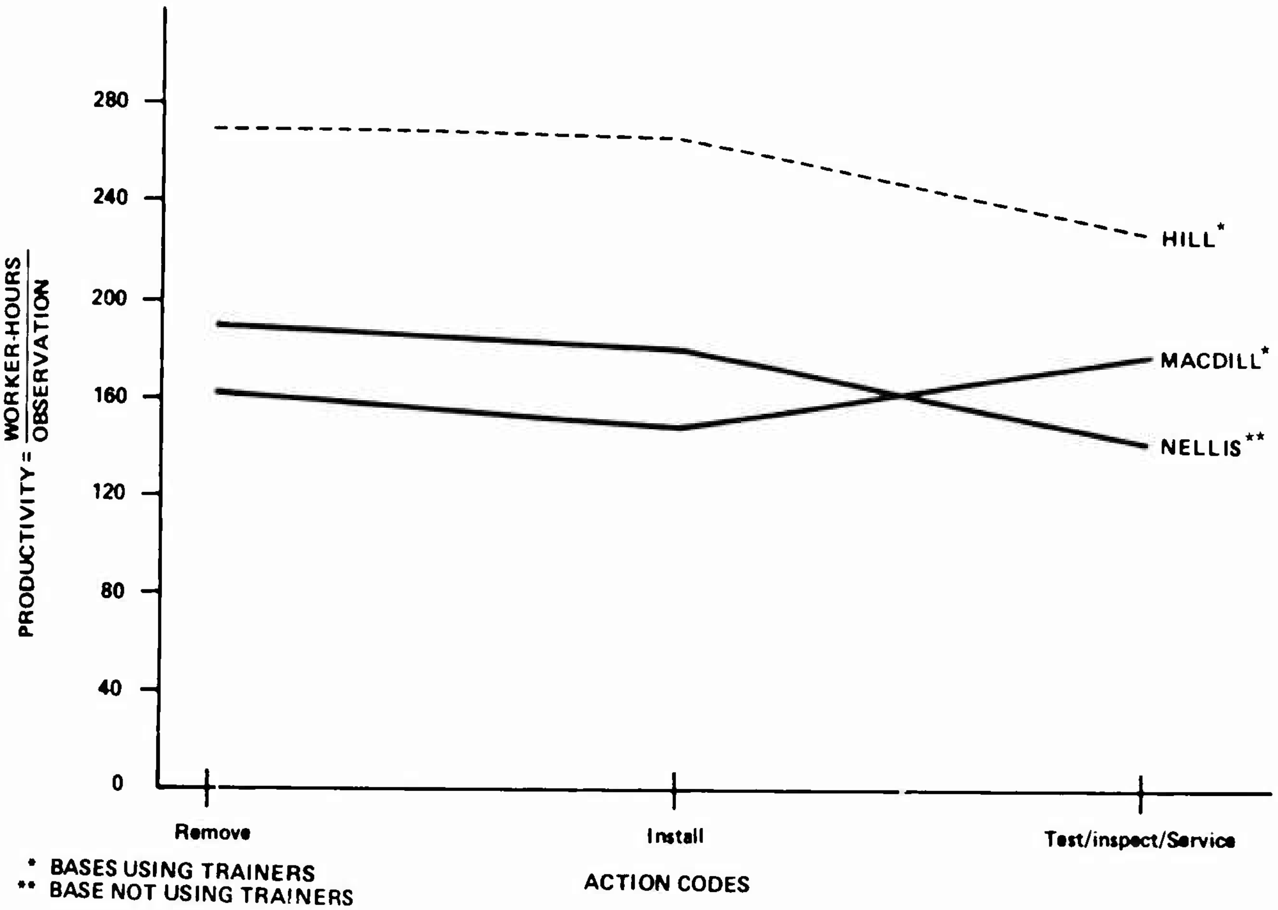

Figure 4. Training: Wing-to-Wing (23Z00)

a satisfactory performance level equivalent to that of actual equipment-trained technicians worth the cost? The answer to these questions can only come from a longitudinal study that controls for confounding variables. Particular care must be taken to control for the length of time subjects have spent in aircraft maintenance, since this is likely to have a strong effect on performance. Those trained on the actual equipment should be compared to those trained on the individual trainers as the data indicate the possibility that some trainers may be doing a better job of preparing technicians than other trainers.

FUTURE DIRECTIONS

Several steps need to be taken to develop a model that can be used to make decisions regarding tradeoffs between cost and training effectiveness. First, a more rigorous validation of the training effectiveness model must be made. The model should be tested with real world data and be modified as required. Second, an overlay for the training effectiveness model must be developed which relates the various design parameters to their costs. Finally, these two sections must be synthesized so that tradeoff equations between cost and training effectiveness can be developed, and the model turned into a practical tool. The outline of these steps is given in Figure 5.

It will also be necessary to quantify the design parameters used in the model so that they can be meaningfully related to the other variables of the model. This requires several steps. First, quantitative scales must be developed for the dimensions of realism and of instructional aids used in the model. This can be accomplished through the use of scaling methods, such as the Coombs Unfolding Technique, which determine the intervals between various points on a qualitative scale, as in the current model. Concomitantly, it will also be necessary to refine the effectiveness measures (i.e., field performance) so that they are as meaningful as possible. When these two steps are accomplished it will then be possible to collect field data on a representative sample of training equipment to determine their configuration in terms of extent and degree of realism and instructional aids, and their resulting effectiveness for a given training goal. These data can then be used to validate and/or refine the model.

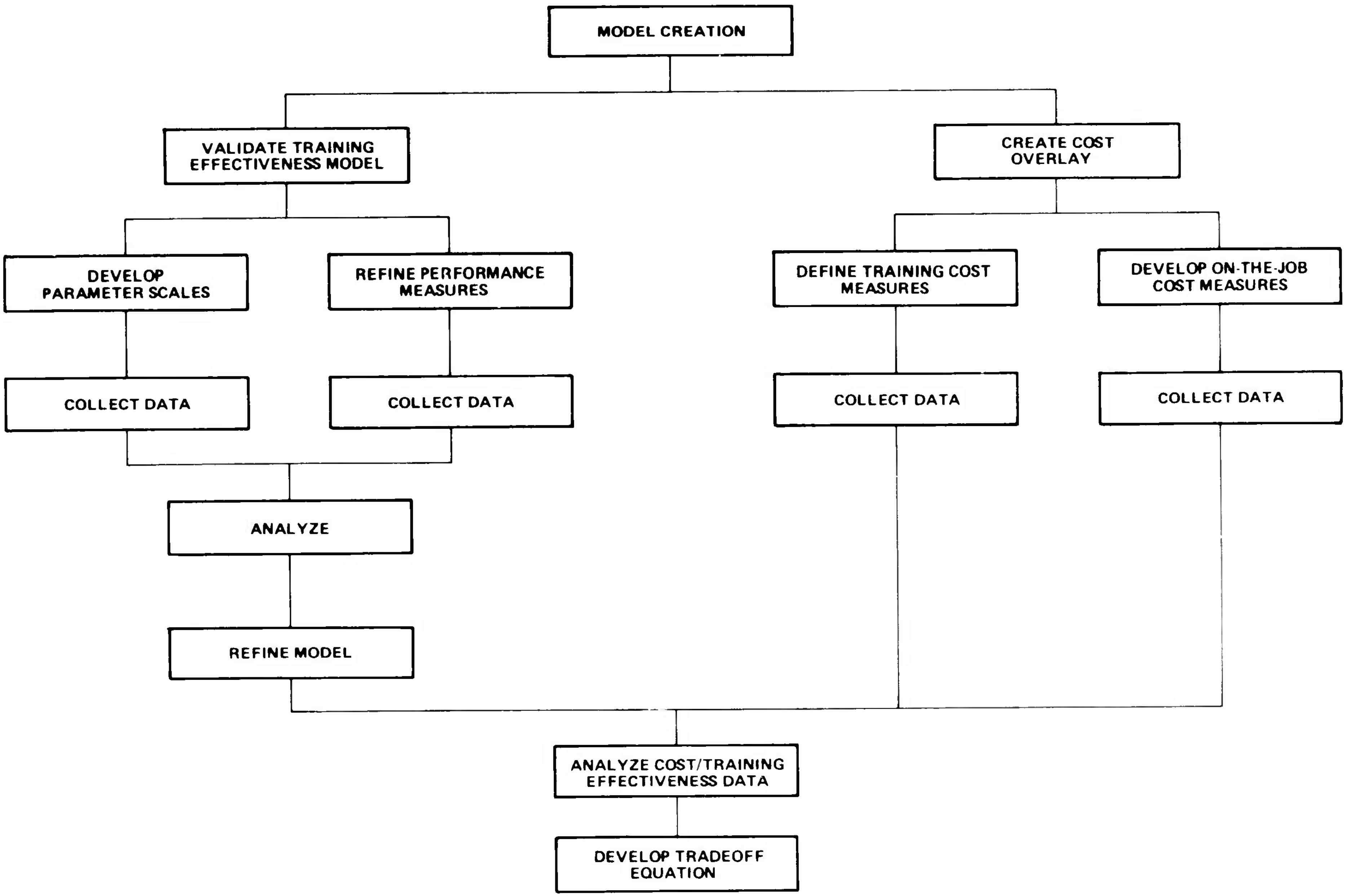

Figure 5. Flow Diagram of Model Validation

The second step in developing the working model into a practical tool is to develop a cost overlay. This requires not only data on the costs of alternative components in a trainer, but also data on the cost of additional on-the-job training necessary for a course graduate to attain minimum proficiency in the field. When this has been accomplished it will then be possible to develop working equations which allow tradeoffs between cost and training effectiveness to be made during the design and development of a training device.

REFERENCES

(1) Report of the Defense Science Board 1982 Summer Study Panel on training and Training Technology, Office of the Undersecretary of Defense for Research and Engineering, Washington, D.C., November 1982.

(2) Stein, K. J., GAO Finds Simulators Used Inefficiently. Aviation Week and Space Technology, 21 February 1983, 118 (8), 72-73.

(3) Pennell, R., & Smith, E. Cost and Traning Effectiveness for Simulator Acquisition. Lowery Air Force Base, CO: Air Force Human Resources Laboratory, Technical Training Branch, August 1982.

(4) Wienclaw, R. A., & Hines, F. E. A Model for Determining Cost and Training Effectiveness Tradeoffs for Training Equipment. Proceedings of the 4th Interservice/Industry Training Equipment Conference, Washington, D.C.: National Security Industrial Association, November 1982.

(5) Smith, P. C., & Kendall, L. M. Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales. Journal of Applied Psychology, 1962, 47, 149-155.

(6) Kane, W. D., Jr. Task Accomplishment in an Air Force Maintenance Environment. Bolling Air Force Base, Washington, D.C.: Air Force Office of Scientific Research, February 1981.

ABOUT THE AUTHORS

DR. RUTH A. WIENCLAW is a Staff Scientist at the Training and Control Systems Operations of Honeywell Inc. She is currently the Principal Investigator on the Training Equipment Effectiveness Study. Dr. Wienclaw holds a Ph.D. in Industrial/Organizational Psychology from Memphis State University and an M.A. in Experimental Psychology from Margquette University. She formerly worked with the Naval Training Equipment Center in Orlando, Florida, and as an Assistant Professor at the University of Central Florida. Dr. Wienclaw has been a private consultant in the areas of training and organizational diagnosis and research for the past eight years.

DR. JESSE ORLANSKY is a member of the technical staff of the Institute for Defense Analyses. He was educated at the City College of New York and Columbia University.