

Witmer, Bob G., Kristiansen, D. M., US Army Research Institute for the Behavioral and Social Sciences, Fort Knox, Kentucky. (Thurs. A.M.)

A Systematic Approach to Training Program Evaluation

A training program for transitioning tank crewmen from the M60A1 to the M1 tank was evaluated during the operational test for the M1. The training program was evaluated systematically using a series of job aids and data collection forms developed by the Army Research Institute (ARI). The evaluation method used was thorough--lesson plans, training procedures, and testing procedures were all evaluated--yet simple. The data collection forms were completed by mid-level noncommissioned officers whose only experience in evaluation consisted of a three-day workshop. Completed data collection forms were returned to the training analyst, an ARI researcher, who identified deficiencies in the training program from the information appearing on the forms. When soldiers could not pass course embedded tests, the method provided objective information about what had gone wrong during training and how to modify the training program to improve it.

## A SYSTEMATIC APPROACH TO TRAINING PROGRAM EVALUATION

Bob G. Witmer  
Research Psychologist  
US Army Research Institute  
Fort Knox Field Unit

and

D. M. Kristiansen  
Team Chief  
US Army Research Institute  
Fort Knox Field Unit

### Introduction

The Army, in response to new Warsaw Pact weapon system development, new technology in NATO, and searches for economies in force structure, is constantly upgrading its weapon system inventory. Within the past few years, the development of a large number of new weapons systems has been started. Many of these systems are in, or are approaching, the operational testing phase.

Attending each of these systems are training programs designed to transition soldiers to the new equipment and to train newly accessioned soldiers. Transition training programs, and to some extent programs for new soldiers, are developed concurrently with the hardware. The first real opportunity to test the effectiveness and efficiency of transition training comes during operational testing. Historically, little attention has been paid to training during operational tests. Cost and Training Effectiveness Analyses have been long on cost analysis and short on training analysis. The problem has been that a useful, easily administered, high face valid set of procedures for gathering information on training effectiveness and efficiency has not been available.

The lack of a useful set of procedures for measuring training effectiveness and efficiency has resulted in many training problems going undetected. Poor training practices used in the transition training program for a given weapons system are often incorporated in the design of other training programs for that weapons system. Once routine training begins, analysis of the effectiveness of that training is rarely considered. Without rigorous formal evaluation, poor training can be given over and over again producing large numbers of soldiers who operate or maintain weapons systems with considerably less than optimal proficiency.

To ensure that training programs for new weapon systems (or old weapon systems for that matter) are effective and efficient, a methodology was needed whereby factors that adversely affect training effectiveness and efficiency could be identified. Such a method has been developed by researchers at the Fort Knox Field Unit of the Army Research Institute (ARI) and given a field trial during Operational Test III of the M1 tank.

### Training Program Evaluation Methods and Materials

The methodology for assessing training program effectiveness is the elaboration and refinement of a concept developed by Harless Performance Guild, Inc. The concept assumes a training program should be based on a set of explicitly stated training objectives. The training objectives must state the tasks to be learned in performance terms to include the conditions under which the performance must occur and the standards that must be met. The object of the training is to arrange a series of training events that enable soldiers who receive the training to meet the training objectives. The purpose of the test is to ensure that soldiers can perform the tasks stated in the training objectives. The method does not question the adequacy of the training objectives to meet the soldier's training needs. It assumes that the objectives were established through a thorough front-end analysis. The method seeks to answer two questions: "Has the training met its own goals as stated in the training objectives?" and "Is the training as effective and efficient as possible, given the constraints under which it must operate?"

To answer these questions, the method requires that data be collected on the training and testing processes as they unfold. Preliminary information on the training and testing processes is obtained through an evaluative review of the lesson plans. The adequacy of the lesson plans is assessed by evaluating each

lesson plan against a set of preestablished criteria. The criteria basically determine if the training objectives as stated are measurable, if the test as designed is a good measure of whether or not the training objectives were met, and if the lesson plan specifies the necessary training events in sufficient detail to allow the instructor to train the soldiers to meet the training objectives. Examples of the criteria used in evaluating lesson plans are included in Table 1. The detailed procedures for evaluating lesson plans and a complete list of the criteria may be found in ARI Research Product 81-15 (Kristiansen and Witmer, Note 3).

TABLE 1. Sample Items for the Systematic Evaluation of Lesson Plans

1. Does the training objective specify what the soldier must do after having been trained?
2. Does the training objective specify the standards to which the soldier must perform?
3. Are the standards clearly spelled out so that the soldier, the instructor, and a training evaluator can tell the difference between performance at or above standard from performance that is below standard?
4. Do the test items derive directly from the training objective?
5. Are the instructions for administering the test such as to ensure standardization across instructors?
6. Does the lesson plan prescribe how demonstrations should be conducted?
7. Is practice called for in the lesson plan?
8. Does the lesson plan call for practice to take each soldier up to the training standard?

Data on the training process is also collected as the instruction is being delivered. Training observers monitor the training as it is being conducted, recording their observations and comments about the training process on specially designed training observation worksheets. Data is recorded about certain critical training events such as the explanation of unfamiliar terms and concepts, demonstrations, practice activities, and testing for task proficiency. While monitoring training, training observers also record information about the training environment on a worksheet designed for collecting such information. Data collected on the training environment includes information about the availability of training resources (i.e., equipment, materials, personnel) and other factors in the training environment (e.g., noise, temperature, lighting) that might influence training effectiveness. Typically a given block of instruction will include a test to determine if the soldiers can demonstrate task proficiency. A testing observation worksheet is used to record data regarding the conduct of the test to include information about test administration, scoring procedures, and contamination of test results. Sample items from the training observation, training environment and testing observation worksheets are listed in Table 2. Notice that the items listed are in question format so that they can be answered with a simple YES or NO response. Additional information may be obtained by asking training observers to record their comments when training problems are observed. A complete list of the worksheet items and a detailed explanation of how they are used may be found in the observer's job aid (Witmer, Note 4).

TABLE 2. Sample Items by Worksheet for the Structured Observation of Training

Training Observation Worksheet

1. Were soldiers told the training objectives including tasks, conditions, and standards?
2. Did the instructor tell the soldiers how the equipment worked and label the parts?
3. Could demonstrations be seen and heard by all soldiers?
4. Did all soldiers practice?
5. Did the instructor follow the lesson plan?

Training Environment Worksheet

1. Were enough instructors present to provide adequate supervision and assistance?
2. Did the training equipment work properly?
3. Did each soldier receive a copy of handouts or other materials used?
4. Were the weather conditions so uncomfortable that soldiers were distracted from the training?

Testing Observation Worksheet

1. Were soldiers tested on any tasks that were not taught?
2. Were the standards specified in the training objectives used to score test performance?
3. Did the examiner help the soldiers in any way during the test?
4. Were soldiers told what they did right and wrong on the test?

Data collected using items like those in Table 2 are useful in identifying training program deficiencies. When tests given at the end of a block of instruction produce unacceptably high failure rates, the data collected during training are used by the training analyst to determine what changes to make in the training program to reduce the performance deficiency. When test performance is much better or much worse than anticipated, the data collected during testing may identify irregularities in the administration of the test that account for the unexpected test scores. Kristiansen (Note 1) has developed a modifications job aid for assisting the analyst in making recommendations for training program changes from the data collected during training and testing. Kristiansen and Witmer (Note 2) are presently developing an additional job aid, Research Product 81-18, that will provide overall guidelines to the analyst on how to organize and conduct a training program evaluation.

Planning the M1 Transition Training Program Evaluation

Preliminary versions of these job aids were used in conducting the evaluation of the transition training program for the M1 tank at OT-III. The transition training program was designed by the M1 NET team to transition tank crewmen and mechanics from the M60A1 tank to the new M1 tank. Three tank companies were to be transitioned in succession to the new tank. For these companies both individual and collective crew-level skills would be taught. Organizational maintenance training would also be conducted to transition track vehicle and turret mechanics from the M60A1 to the M1 tank. Because the M1 transition training program was new and unproven, there existed a need to evaluate the effectiveness of the program. ARI fulfilled that need by providing a methodology and evaluation materials for determining effectiveness of the transition training.

Considerable preparation and planning preceded the field trial of the TPE job aids and procedures. The planning process was complicated by the fact that several different organizations with different goals and different data needs were involved. Among the organizations that were to have the greatest impact in determining the ultimate shape of the M1 training program evaluation were the TRADOC Combined Arms Test Activity (TCATA), the Office of Armor Force Management and Standardization (OAFMS), the Directorate of Training Developments (DTD) at the US Army Armor Center, the TRADOC Systems Analysis Activity (TRASANA), the Operational Test and Evaluation Agency (OTEA), and the Army Research Institute. Through a series of meetings and other interactions, these agencies jointly made the plans and decisions that would determine how the training evaluation for the M1 OT-III would be conducted. Some of the important issues discussed during these meetings include the following: 1) what data would be collected; 2) how it would be collected; 3) who would collect it; and 4) how the data would be distributed to the various participating agencies. Despite some initial disagreements among the agencies involved, each of these issues was resolved prior to the beginning of the OT-III.

Basically it was decided that two types of data would be collected - training data and performance data. The training data would be collected using the procedures and worksheets (see Table 2) developed by ARI. In addition to the worksheets developed by ARI, several other data collection instruments (e.g., student questionnaire, instructor questionnaire, training aids data sheet) that were of special interest to other agencies were to be used for collecting training data during the M1 OT-III. Performance data on the tests given at the end of each block of instruction were to be recorded for each soldier on individual score sheets. Diagnostic tests designed to measure selected soldier skills on the M60A1 tank just prior to M1 transition training and the same skills on the M1 following transition training were developed by TRASANA and DTD. TRASANA also constructed special score sheets for recording these diagnostic data. Although sampling of training and performance data had been considered, it was decided to collect complete training and performance data by monitoring all training sites.

The training data would be collected by a team of data collectors organized and controlled by TCATA. The data collection team would be composed of a company team chief data collector (captain), three platoon team chief data collectors (lieutenants) and one tank data collector per tank (mid-level noncommissioned officers). Data collection responsibilities were to be shared by the platoon and tank data collectors with the company data collectors acting as supervisor for the other data collectors and reviewing completed worksheets to ensure that the data collectors had responded appropriately to each worksheet item.

Performance data would be collected by the M1 new equipment training team (NETT) instructors for both end-of-block tests and diagnostic tests. However, tank data collectors would monitor the collection of performance data and complete testing observation worksheets describing the testing process for each test given.

Training data were to be collected as the training was being conducted for each block of instruction. Data collectors would be assigned to a particular block of instruction at least 24 hours prior to observing the instruction. The data collectors would observe each part of the instruction as it was being conducted. Guided by the items listed on the worksheets, the data collectors would look to determine whether certain training events (e.g., demonstration, practice, test) occurred satisfactorily and record comments on events that were not satisfactory.

Because of the large amounts of data that were expected to be generated by the evaluation of the M1 transition training program and the diverse data requirements of the agencies involved, a system was devised for handling the data. The data handling system called for completed worksheets and score sheets to be checked for omissions and inconsistencies by the company team chief data collector before being forwarded to the field test center. At the field test center, the worksheets and score sheets were to be reproduced and copies distributed to ARI and other

participating agencies. After copying, the original data forms would be placed in a master file for future reference. The data on the worksheets and score sheets were to be analyzed by ARI as received, and any resulting recommendations for training program changes were to be forwarded by ARI to the NETT through DTD. This system was designed to allow ARI to receive the data, analyze it, and make recommendations to the NETT within 24 hours of the time the training was conducted.

Prior to the beginning of the M1 OT-III ARI obtained copies of the lesson plans for the transition training program for tank crewmen. Each lesson plan was evaluated using the procedures and criteria described in a preliminary version of ARI Research Product 81-15 (Kristiansen and Witmer, Note 3). This job aid for evaluating lesson plans seemed to work well, producing 43 pages of comments and suggestions regarding problems associated with the lesson plans. However, the NETT did not revise the lesson plans as recommended prior to the beginning of the M1 OT-III.

In preparing for the M1 OT-III evaluation, ARI conducted workshops in order to train the soldiers who were to collect the data. In the workshops, the soldiers would learn about objective observation procedures and be familiarized with the observer's job aid and the training evaluation worksheets. During the workshop, each item on training environment, training observation, and testing observation worksheets were explained and relevant examples were given where applicable. After this initial familiarization with the worksheets, the soldiers were required to use the worksheets to record observations on an actual class as a practice exercise. Following practice in using the worksheets, the soldiers discussed their observations of the class and received feedback from the workshop leader.

#### Field Testing the TPE Method and Materials During the M1 OT-III

The M1 OT-III commenced at Fort Hood, Texas in September of 1980. ARI's early involvement in the operational test consisted primarily of training data collectors to use the TPE worksheets to collect training evaluation data. Aside from training data collectors, ARI's role during the M1 OT-III was limited to that of the training analyst. As the training analyst ARI analyzed the training and performance data generated during the OT-III and recommended changes to the training program based on this analysis. ARI's activities during the M1 OT-III provided an excellent opportunity for field testing the TPE method and materials. Analyzing the transition training allowed the ARI researcher to assess the adequacy of the TPE system as a method for identifying training problems. Along with teaching the workshop, performing the analysis also provided information on the ability of noncommissioned officers to use the TPE worksheets to collect the training data.

In conducting the field test of the TPE system, ARI confined its analysis activities to evaluating the effectiveness of the M1 transition training given to tank crewmen. ARI evaluated the effectiveness of this training for each of the three tank companies participating in the M1 OT-III. Data was systematically collected by task on observable training process variables detailed in the observer's job aid (Witmer, Note 4). For most of the training program, training was conducted at multiple sites, which often required as many as thirteen training observers watching the same training simultaneously, but at different sites. The bulk of the training data was collected by the tank data collectors. Contrary to what was planned, the platoon data collectors collected very little data and served primarily as supervisors for the tank data collectors. The company data collector reviewed each completed worksheet and after receiving a complete set of data for a given block of instruction, forwarded the data to the field test center for reproduction and distribution. At the field test center, the ARI analyst received copies of the Observation of Training Events Worksheet, the Training Environment Worksheet, the Observation of Test Events Worksheet and individual test performance score sheets for each block of instruction. Several days often elapsed between the time when the training was conducted and the time when ARI received the data for a given block of instruction. The delay in the receipt of the data most often was

due to inefficient reproduction of the data resulting from problems with both the copying machine and the machine operator. Frequently the copies when received were of such poor quality that it was necessary to go back to the originals to read a comment that was illegible on the copy.

For ease of analysis, the data for each block of instruction were transferred onto summary data forms. The summary data forms showed the total number of occasions on which a worksheet item was judged satisfactory or not satisfactory and included all the comments made regarding a particular block of instruction. The summary data forms allowed the training analyst to quickly identify the training problems for each block of instruction.

For analysis purposes, the analyst also summarized the data from the test performance score sheets. The score sheets recorded individual soldier performance by task for each task tested. Each soldier received either a GO indicating that the task was performed to standard or a NO-GO indicating that the soldier failed to meet the standard. The analyst reviewed the score sheets for each block of instruction to obtain the percentage of soldiers receiving NO-GO's for each task. Tasks for which 20 percent or more of the soldiers tested received a NO-GO were considered to represent performance deficiencies. Possible causes of these deficiencies were identified from the training data recorded on the Observation of Training Events, Observation of Test Events and Training Environment Worksheets. From these causes, the analyst, with the help of the modifications job aid (Kristiansen, Note 1), suggested changes in the training for eliminating the performance deficiencies.

For some blocks of instruction, review of the score sheets did not turn up any performance deficiencies. Nevertheless the observation of test events worksheet was reviewed to determine if there were any irregularities in the testing procedures. When problems were identified in the testing process that could adversely affect the validity of the test, changes to the test were suggested based on the guidance provided in the modifications job aid.

The training analyst summarized his findings and recommendations for each block of instruction in a memorandum to DTD. The memorandum identified the tasks for which performance deficiencies were found and specified changes in the training program designed to correct the deficiencies. The memorandum also identified problems with testing procedures when they existed and suggested changes as appropriate.

Because of a number of problems unique to the M1 OT-III at Fort Hood, ARI memorandums were not forwarded to the NETT until the first company had completed much of its training. Day-to-day changes were not made as planned.

Even though the NETT received the ARI's memorandums following transition training for a given company, there was still time to modify the transition training before training the next company. Upon completion of the transition training for the first company, DTD issued a memorandum instructing the NETT to modify the transition training given to the next two companies so as to eliminate some of the recurring training problems identified by ARI for the first company. Some of the modifications suggested by ARI appeared to have been made prior to conducting transition training for the second and third companies. Evidence that some changes were made comes from the training and performance data collected for the second and third companies. However, changes were not extensive and were not always documented in the lesson plans. Failure to document changes in the transition training from one company to the next precluded establishing direct

relationships between the changes made in the transition training for each block of instruction and performance on the tests given following that block of instruction. However, considerable information was gathered that indirectly demonstrated the usefulness of the TPE methodology in evaluating the effectiveness of training programs.

The TPE methodology provided the largest pool of objective training data that was available during the M1 OT-III. The usefulness of this data is attested to by the fact that virtually every organization involved in evaluating the effectiveness of the transition training made use of this data. The primary users were OAFMS, TACTA and ARI. OAFMS used the data for certifying the readiness of the OT III players as M1 qualified crewmen and mechanics. OAFMS also used the data in certifying the effectiveness of the M1 transition training program. TACTA kept detailed records of all the TPE training data and used the data and the memorandums generated by ARI as input for their own independent training effectiveness analysis. DTD and TRASANA also used the data in conducting a Cost and Training Effectiveness Analysis (CTEA) of the M1 transition training program, although TRASANA relied more heavily on pre- and post-diagnostic test data. All of the organizations involved in evaluating the effectiveness of M1 transition training seemed to find the data useful.

Among the changes made in the transition training for the second and third companies were the addition of demonstrations to some lessons and closer adherence to the lesson plans in conducting the instruction. Evidence that these changes were made came from comments recorded by the data collectors on the worksheets and were verified by informal contacts with the NETT and OAFMS. The number of comments indicating that tasks were not demonstrated or that lesson plans were not followed dropped sharply from the first company to the second and remained at the lower level for the third company. This indicates that the TPE methodology was sensitive to changes made in the training process.

Performance data indicating soldier performance on the end-of-block tests shows improvement from one company to the next. This may be interpreted as indicating increased training effectiveness. For example the proportion of tasks having 100 percent first-time GO rates increased from 24 percent for the first company, to 34 percent and 53 percent for the second and third companies, respectively. While such increases may be due in part to other factors (e.g., reduction of standards for some tasks and the elimination of some task requirements), the trend toward higher first-time GO rates constitutes indirect evidence that the changes made in the training from one company to the next increased training effectiveness and thus supports the usefulness of the TPE methodology responsible for these changes.

Although the data provided by TPE methodology was considered useful by the organizations employing it and the recommendations suggested by ARI based on the data seemed to increase training effectiveness, ARI encountered some problems in using the TPE methodology for evaluating the M1 transition training. The biggest problem came in getting the NCO's who were collecting the data to use the TPE worksheets the way they were designed to be used. The worksheets listed specific items which required the data collectors to observe the training to determine if it met specific criteria described in the observer's job aid and discussed in the TPE workshop. When these criteria were not met, the data collectors were supposed to record a comment detailing what they observed. However, many of the data collectors treated the worksheets as a simple checklist, responding to the items subjectively based on their general impressions of the items rather than the objective criteria specified in the observer's job aid. Furthermore the number of comments recorded on the worksheets were far fewer than were expected based on independent observations of the training by the ARI analyst and others.

Any one of several factors might account for the failure of some data collectors to use the worksheets as the worksheets were designed to be used.



The NCO's collecting the data were tankers who were perhaps more interested in the new M1 tank than in making objective observations about the training and testing processes. For these tankers, the task of collecting training and testing process data may have been considered menial or meaningless, especially in contrast to learning how to operate the new tank. The data collectors were required to complete several other forms in addition to the TPE worksheets; this additional workload may have diminished the amount of effort devoted to the TPE worksheets. On several occasions, the TPE analyst observed that data collectors were standing around in groups of four or five while the training that they should be monitoring was being conducted on the tanks. In these instances, the data collection team supervisors were not supervising and the data collectors were not collecting data. Such incidents and informal discussion with the data collectors indicated that many of the data collectors were not approaching the data collection task seriously. This apparent lack of motivation in the data collectors may have resulted in part from the factors listed above, but may also be due in part to the adverse working conditions. For example, data collectors were often required to collect data all day in wet or cold and windy weather. Often they were called upon to collect data right through (and long after) normal meal hours. Many of the "creative comforts" provided to the participating units were not given to the observers and they were treated as unnecessary by the NETT and the units.

The problems encountered in using the TPE methodology during the M1 OT-III are instructive. They indicate that the TPE analyst must select the data collectors more carefully and personally oversee and control the data collection effort. The number of forms to be completed by any one data collector and the number of hours spent completing these forms must be limited to a reasonable level. In order to ensure the timely flow of data from the data collector to the analyst and from the analyst to the persons responsible for instituting changes in the training program, a direct line of communications should be established from the analyst in both directions. The originals of completed worksheets and score sheets should go directly from each data collector to the analyst. The analyst would then analyze the data and results of this analysis would be made available to other organizations. If other persons or organizations needed the raw data, they would have to obtain copies through the TPE analyst. The recommendations for changes in the training program would then be forwarded directly to a member of the team responsible for making changes in the training program. This person would check into the possibility of making each of the changes suggested by the analyst and would inform the analyst which changes were made and which were not.

In order to reduce the tendency of data collectors to respond subjectively to the items on the TPE worksheets, the worksheets and the items used during the M1 OT-III were modified in the new version of the observer's job aid. The old worksheets required data collectors to make judgments (i.e., OK or Not OK) regarding their observations according to criteria listed for each item. On the revised worksheets, items are worded more precisely and data collectors only need respond objectively with a "YES" or "NO" indicating whether or not the event occurred as described in the item.

In conclusion, the TPE system has proven to be a useful method for evaluating training programs such as the transition training program for the M1 tank. The method provides objective training and testing process data not heretofore available. The TPE system has been well documented in a series of easy-to-use job aids (Witmer, Note 4; Kristiansen, Note 1; Kristiansen and Witmer, Note 2; and Kristiansen and Witmer, Note 3). The quality of these job aids has been much improved through the lessons learned from the M1 OT-III experience, and provide detailed information on how to collect and analyze training data in order to evaluate the effectiveness and efficiency of training programs.

### Reference Notes

1. Kristiansen, D. M. A job aid for modifying ineffective or inefficient training programs (ARI Research Product 81-17) Fort Knox, KY. Fort Knox Field Unit, US Army Research Institute for the Behavioral and Social Sciences, Manuscript submitted for publication, September 1981.
2. Kristiansen, D. M. & Witmer, B. G. Guidelines for conducting a training program evaluation (ARI Research Product 81-18) Fort Knox, KY. Fort Knox Field Unit, US Army Research Institute for the Behavioral and Social Sciences, Manuscript in preparation, October 1981.
3. Kristiansen, D. M. & Witmer, B. G. A job aid for the systematic evaluation of lesson plans (ARI Research Product 81-15) Fort Knox, KY. Fort Knox Field Unit, US Army Research Institute for the Behavioral and Social Sciences, Manuscript in preparation, October 1981.
4. Witmer, B. G. A job aid for the structured observation of training (ARI Research Product 81-16) Fort Knox, KY. Fort Knox Field Unit, US Army Research Institute for the Behavioral and Social Sciences, Manuscript submitted for publication, September 1981.