DOCUMENT RESUME

ED 391 468 IR 017 616

| | |
|---|---|
| AUTHOR | Shlechter, Theodore M.; And Others |
| TITLE | Multimethod-Multisource Approach for Assessing High-Technology Training Systems. |
| PUB DATE | 19 Apr 95 |
| NOTE | 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995). |
| PUB TYPE | Reports - Research/Technical (143) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | Computer Simulation; *Educational Technology; *Evaluation Methods; Improvement Programs; *Institutes (Training Programs); Instructional Effectiveness; *Military Training; Professional Training; *Research Methodology; Skill Development |
| IDENTIFIERS | *Army National Guard; *Multiple Data Gathering Methods |

ABSTRACT

This investigation examined the value of using a multimethod-multisource approach to assess high-technology training systems. The research strategy was utilized to provide empirical information on the instructional effectiveness of the Reserve Component Virtual Training Program (RCVTP), which was developed to improve the training of Army National Guard units. Data was collected from nine units; 14 RCVTP instructors completed standard rating forms regarding the performance of 38 armored force units and 280 training participants completed Likert-scale items regarding their training experience. Data from the different methods indicated that the units further developed their collective tactical skills across the training period. The use of multiple methods and sources was found to provide a better understanding of the RCVTP's effectiveness than could be provided by any single method and source. In addition, this research strategy is a viable approach to evaluating a high-technology based training system in a non-controlled context without the possibility of obtaining either baseline or transfer measures of performance. (Contains 36 references.) (Author/AEF)

# Multimethod-Multisource Approach for Assessing High-Technology Training Systems

Theodore M. Shlechter
U.S. Army Research Institute
Fort Knox, Kentucky

K. Paul Nesselroade, Jr.
University of Louisville

Billy L. Burnside
U.S. Army Research Institute
Fort Knox, Kentucky

David W. Bessemer
U.S. Army Research Institute
Fort Knox, Kentucky

James Anthony
University of Louisville

---

Paper presented at the 1995 Annual Meeting of the American Educational Research Association, San Francisco, CA. Opinions in this paper are those of the authors and do not necessarily reflect those of the U.S. Army Research Institute, Department of the Army or the United States Government.

## ABSTRACT

This investigation examined the value of using a multimethod-multisource approach to assess high-technology training systems. This research strategy was utilized to provide empirical information regarding the Reserve Component Virtual Training Program's (RCVTP's) instructional effectiveness. Observers collected data from nine units; fourteen RCVTP instructors completed standard rating forms regarding the performance of 38 armored force units; and 280 training participants completed Likert-scale items regarding their training experience. Data from the different methods indicated that the units further developed their collective tactical skills across the training period. The advantages and problems with using a multimethod-multisource research strategy for assessing high-technology training systems were then discussed.

# Multimethod Multisource Approach for Assessing High-Technology Training Systems

This paper addresses the problem of evaluating the effectiveness of a high-technology based training system in a non-controlled context without the possibility of obtaining either baseline or transfer measures of performance. Such non-traditional evaluations may become more prevalent for the military training community as the resources to conduct controlled transfer evaluations become more scarce. The high costs in time, money, and personnel associated with traditional transfer evaluations are well recognized by trainers and researchers (Blaiwes & Regan, 1986).

## Evaluation Issues

Selecting Appropraite Data collection Methods. Evaluating a high-technology training program in a non-controlled context without the possibility of obtaining either baseline or transfer measures of performance presents several interesting challenges. As is the case with any evaluation, the primary challenge involves determining the most appropriate method(s) for collecting the data. This issue is even more pronounced for high-technology training systems (such as the Simulation Networking or SIMNET system[1]) which have not been equipped with any device for providing quantitative performance data. Researchers must then collect data through observations (Gound &

[1] See Garvey & Radgowski, 1988 for a detailed description of this system.

Schwab, 1988), questionnaires (Brown, et al., 1988), or instructor ratings (Bessemer, 1991; Shlechter, Bessemer, & Kolosh, 1991).

Each method is potentially problematic. Observational methods are labor intensive, a situation which limits the sample size. These methods may also be contaminated by the obervers' subjectivity and corresponding problems with reliability (P.A. Adler & Adler, 1994). Questionnaires can be tainted by the students' inability to report, accurately, the effects of the training device on their performance. The accuracies of self-reports have been hotly debated by psychologists (e.g., Burnside, 1982; Herrmann, 1982). Instructor ratings may be contaminated by the expectations or biases held by the instructors (Cook & Campbell, 1979). These instrumentation problems are more critical in non-controlled or quasi-experimental designs, which are susceptible to extraneous variables (Cook & Campbell).

One approach for dealing with such extraneous variables is found in Bessemer's (1991) quasi-experimental evaluation of SIMNET. This evaluation consisted of obtaining instructor ratings for 1705 Armor Officer Basic (AOB) students of which 1059 did not receive SIMNET training and 646 did. Multiple regression techniques were used to help remove the effects associated with extraneous variables, e.g., instructor biases in the ratings.

There were two main difficulties with implementing Bessemer's evaluation approach. One, his evaluation involved a large sample size, which might be difficult to obtain in a future of dwindling resources for research and development. Two, statistical adjustments can never fully substitute for experimental controls.

A multimethod evaluation approach has also been suggested as a technique for circumventing the cited limitation with naturalistic evaluations (e.g., Denizen, 1978 as cited in Patton, 1987; Denizen & Lincoln, 1994; Cook & Campbell, 1979). This approach is expected to provide a more in-depth understanding of the phenomenon under study than could be provided by the use of any single evaluation methodology (Denizen & Lincoln). Also, areas of agreement between methods would boost confidence in the data's internal and construct validity (Cook & Campbell). And, Scandal, Money, Grainier, & Hall (1983) have noted that self-report methodologies may serve to strengthen and refine data from other, more generalized approaches toward predicting task performance.

Denizen (1978) has also suggested that naturalistic evaluations should sample data from and/or by a variety of sources. Each source could provide a different perspective regarding the training situation. Observers who are not part of the training process may view a subject's performance differently from an instructor who is part of the process. Perhaps then, a multimethod-multisource approach should be employed when conducting a quasi-experimental evaluation of high-technology training systems.

Sampling Adequate Criterion Measures. A multimethod-multisource design can be useful for helping evaluators to meet another long-standing challenge--sampling adequate criterion measures. As noted by Gagne (1954):

> "The 'criterion problem' has been with us (researchers of training devices and simulators) for a long time." (p. 95)

More recently, Shute and Regian (1993) have noted that sampling adequate criterion measures has been a problem which has plagued evaluations of high-technology training systems, especially those systems designed to help students to become proficient in performing complex tasks. This problem has been manifested, for example, in nearly all previous evaluations of SIMNET's effectiveness (Kraemer & Bessemer, 1987; Bessemer, 1991; Brown, Paschal, & Southard, 1988; TEXCOM Combined Arms Test Center, 1990; Shlechter, Bessemer, & Kolosh, 1991). These studies have focused on measuring differences in the SIMNET-trained and control units' abilities to perform certain standard Army training and evaluation program tasks.

Cognitive psychologists, however, have recently argued that expertise involves more than successfully performing a set of tasks (Collins, Brown, & Newman, 1989; Kraiger, Ford, & Salas, 1993; Patrick, 1992). Experts are better able to perform the same task more quickly than the less advanced students (Kraiger, et al). That is, expertise involves "automatizing" the important skills associated with task performance. Expertise also involves the ability to attend to task cues without too much reliance on instructor prompting and to articulate the reasons for one's actions (Collins et al, 1989; Patrick).

Previous Multimethod-Multisource Evaluations. Multimethod-multisource evaluations of training systems are not a novel idea. Finley, Rhinehelder, Thompson, and Sullivan (1972), for example, used both experimental and field evaluation designs for evaluating the training effectiveness of a naval air traffic control center training device. And, Brehmer & Dorner (1993) have suggested that investigations of computer-simulated microworlds include experimentation and case studies.

A search of several bibliographic data bases and literature reviews (Adams, 1978; Hays & Singer, 1989; Orlansky, Dahlman, Hammon, Metzko, Taylor, & Youngblut, 1994; van Berkum & DeJong, 1991) failed to locate very many naturalistic evaluations of high-technology training systems which employed a multimethod-multisource approach. Adams has noted that evaluations of flight simulators have mainly consisted of controlled transfer studies or equipment ratings by subject-matter experts (SMEs). Adams has argued that both methods are highly flawed.

Furthermore, the more recent literature on training simulations has tended to be: (a) analytic estimates of the system's training capabilities (Burnside, 1990); (b) controlled transfer studies (e.g., McAnulty, 1992; Swezey, Perez, & Allen, 1991); (c) field studies involving multimethods but not multisources (e.,g., Lesgod, Lajoie, Bunzo, & Eggan, 1988); (d) equipment ratings by SMEs (Harrison, Acchione-Noel, Butler, Nantze, & Walker, 1992); (e) cost estimates (see Orlansky et al, 1994); and (f)investigations of the system's fidelity ( see Hays & Singer, 1988). Lesgold et al, for example, used both a "think-aloud" protocol ("What are you now thinking?") and simulated recall methodology ("What did you think?") for their field investigation of the Sherlock troubleshooting systems.

Overview of current study

Objective. This study was thus designed to illustrate the advantages and limitations with using a multimethod-multisource approach for conducting a naturalistic evaluation of a high-technology based training program. The training program evaluated was the Reserve Component Virtual Training Program (RCVTP), which has been implemented at Fort Knox, KY.

Brief Description of the RCVTP. The RCVTP has been developed through congressional funding to improve the training of Army National Guard (ARNG) units. This funding was made available because ARNG units, who are becoming an increasingly important element of post cold-war combat, have limited training resources and time with only 39 days allocated for training per year.

This program's primary goal involved having ARNG units experience National Training Center (NTC)-like missions in a time-compressed manner. Providing NTC-like training in a time-compressed manner involved utilizing the available high technology training simulation systems at Fort Knox, KY. The primary simulation utilized by the RCVTP was the SIMNET system.

Providing NTC-like training for ARNG armored units also entailed developing a structured set of SIMNET training exercises (training tables). This structure consisted of having units perform actions (critical subtasks) associated with specific training objectives and cues. Examples of critical subtasks included: (a) reaching the starting point on time; (b) executing fires when the enemy crosses the trigger line; and (c) conducting displacement as directed.

Approximately one hundred such training tables were created for this training program with each training table designed to be conducted in two hours. One half-hour was spent by units on preparing for the mission, 1 hour on executing the mission, and another 1/2 hour on participating in an after-action review (AAR) of the exercise. (See Shlechter, Bessemer, Nesselroade, & Anthony, 1995) for more detailed information regarding this training program.)

The RCVTP training managers felt that conducting an evaluation with experimental controls would encroach upon their training program. That is, they wanted the training conditions for the RCVTP's formative evaluation to be very similar to the training conditions for the fully implemented program.

Methods used. The methods and sources used to assess the RCVTP were based on the previous SIMNET evaluations. These different assessments consisted of observations by evaluators, instructors' (RCVTP observer/controllers'--O/Cs') judgments of performance, and participants' questionnaire responses. These methods are further delineated in the sections dealing with

Assessment A (observations by researchers), Assessment B (O/Cs' judgments) and Assessment C (questionnaire responses).

Participants sampled. Most of the participants for this evaluation were part of the developmental trials phase of the RCVTP. This phase took place during the Winter and Spring of 1994. All units volunteered to participate in this evaluation.

Assessment A: The Observers' Reports

This assessment was conducted by researchers from the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), who were independent of both the instructional design and training process. To save costs, these observers sampled targeted training units.

Method

Participants. Nine units were observed. These units consisted of three armor companies, two armor company teams, and two armor platoons. All were ARNG units with the exception of one company team and one armor company who were active units stationed at Fort Knox, KY.

Instruments. The RCVTP Training Observation Form was created to measure the different aspects of tactical skills expertise. This instrument allowed the observers to collect data on: (a) time taken for preparations; (b) O/Cs' and units' actions for the training table; (c) time taken to complete the training table; (d) problems encountered during a mission, such as radios not working or the unit's failure to send a report; and (e) coded entries about the AARs. There was also room in this form for making comments about the mission.

The observers also completed five Likert-scaled items dealing with the mission and 22 items dealing with the AAR. A 1-5 scale was used for these items--1 as never and 5 as always--which was based on a scale developed by Kraemer and Wong (1992). These items for the exercises were based on the critical subtasks associated with the RCVTP.

Training the observers. The observers'--two research psychologists and a graduate student intern with the Armored Forces Research Unit at Fort Knox, KY--training consisted of systematically going over a detailed set of instructions. They were also directed to view videotapes of an AAR conducted when the RCVTP training tables were being piloted. Also, two of the observers read the Army manual on platoon tactics (FM 17-15: U.S. Department of the Army, 1987). The third observer was quite knowledgeable with regards to platoon and company tactics.

Data collection procedure. Data were collected for the sampled units by three observers. Because of constraints imposed by the training trial procedures, these observers were rarely able to record data for the same training tables. The sequence of training tables viewed by each observer varied from unit to unit. Observer A, for example, viewed the first three training tables for Unit 1 and the last three training tables for Unit 2. This variation helped to control for possible data biases due to systematic observer differences.

A reliability check was conducted as observers were able to follow the same training tables for one unit. Few discrepancies were found among observers with regards to the performance data.

Scoring procedures. Two judges scored the observational reports based on a predetermined scoring scheme. The few discrepancies found in this scoring were resolved by a discussion between the judges.

## Results and Discussion

Data for training table performance. Kendall (1975) $\tau$ rank-order correlations were computed to determine the existence of any significant trends in the units' exercise time, errors, and coaching scores across successive exercises in their RCVTP training. Alpha-level for the statistical tests done in this evaluation was set at .10. Because of the limited sample size, these analyses also involved combining the data across platoons and companies and across active and ARNG units. And, data for one training table were not recorded (see Table 1).

Table 1.

Means and Standard Deviations of the Units' Time in Min, Error Rates, and Coaching Scores for Successive Training Tables

| Training Tables | | Time in Min | | Error Rate | | Coaching Score | |
|---|---|---|---|---|---|---|---|
| | n | M | SD | M | SD | M | SD |
| First | 9 | 85.22 | 30.40 | 12.89 | 4.81 | 8.44 | 5.27 |
| Second | 9 | 52.00 | 23.04 | 6.11 | 2.93 | 4.22 | 3.03 |
| Third | 8[a] | 40.88 | 10.51 | 5.00 | 2.44 | 3.75 | 2.81 |
| Fourth | 8 | 41.00 | 12.59 | 6.38 | 4.43 | 5.25 | 3.99 |
| Fifth | 7 | 37.57 | 20.33 | 5.14 | 2.79 | 3.42 | 2.14 |
| Sixth | 3 | 32.00 | 10.44 | 1.67 | .58 | 2.00 | 1.73 |

[a] Data missing for one unit.

As shown in Table 1, these units, typically, took less time, made fewer errors, and needed less coaching as their training progressed. Significant negative trends associated with these measures were confirmed by statistical analyses(see Table 2). These trends were not a function of the units' being less likely to finish their later training tables. Units were found to be more likely to complete their fifth training table rather than their first training table. The RCVTP thus seemingly helped these units to develop their collective tactical skills.

Table 2

Weighted Mean τ-Values and Tests of Significance Including All RCVTP Tables

| Variables | *M* [a] | *SD* | *df* | *t* |
|---|---|---|---|---|
| Times | -.574 | .274 | 8 | 6.31** |
| Errors | -.340 | .174 | 8 | 5.86** |
| Coaching | -.206 | .186 | 8 | 3.32** |

[a] negative number indicates a decreasing trend.
** $p < .05$

Questions exist, however, about the generalizeability of the RCVTP's training effectiveness. The previously cited improvement trends could have reflected units' becoming more adept at using the SIMNET system. The observers did feel that units were more disoriented in their first training table than in the second training table, with reported means of 2.89 and 2.00 for training tables 1 and 2, respectively. This difference was statistically significant, $t(8) = 2.10$, $p < .10$. The observers' comments also indicated that most of the coaching was done vis-a-vis the units' problems with: (a) orientation in the SIMNET terrain, (b) use of the SIMNET radios, and (c) formations.

Data for the AARs. These data were problematic. For one thing, poor reliability was found among observers. Also, observers had trouble following and recording the content of these discussions. Finally, fewer units were sampled for these data than for the performance data, because AARs were not always given after each mission.

The AAR data did indicate that SIMNET-related problems were not an issue for these units. Fewer than 1% of their reported comments in any given AAR dealt with SIMNET. Also, observers indicated that the participants rarely asked questions about or made comments about using SIMNET, with an overall mean rating of 1.30 for this AAR summary item.

Summary of this assessment. This assessment did provide a picture of the RCVTP's effectiveness. However, additional evidence based on a larger sample is needed to confirm this assessment's findings. This evidence also needs to be based on performance judgments made by subject-matter experts. Such judgments are described in the next section, Assessment B.

## Assessment B: The O/Cs' Judgments

Fourteen O/Cs provided these data. An O/C typically assessed the performance of four units. Occasionally, two or more O/Cs were identified as working together on an assessment.

### Method

Participants. Data were collected on 38 armored force units. These units consisted of 17 armor platoons, 10 armor companies, 5 scout platoons, and 6 mechanized infantry platoons. Only 5 active units (4 armor companies and 1 scout platoon) were included in this sample. This sample also included the units who were sampled in Assessment A.

Data collection procedure. For each training table completed by these units, the O/Cs indicated in structured rating forms those subtasks in which units needed either to "train to sustain" or "train to improve," representing satisfactory or unsatisfactory performance, respectively.

Scoring procedures and measures. Two independent judges identified subtasks which these units performed at least twice. O/Cs' ratings were then categorized into measures indicating changes in these units' subtasks proficiency as their training progressed. One set of measures dealt with subtask proficiency changes associated with units' initial and final performance of a subtask. These measures were: (a) improve/sustain; (b) sustain/sustain; (c) improve/improve; and (d) sustain/improve. Also tabulated was the total number of "train to improve" and "train to sustain" ratings for these units' initial and final performance of the different subtasks.

Another set of measures involved examining these units' subtask proficiency across training tables. These measures consisted of counting, separately, the number of ratings for each training table which dealt with units': (a) first performance of a subtask (first subtasks) and (b) later performances of the same subtask (later subtasks). First and later subtasks were counted separately, because the former measure provided an indication of these units' subtask proficiency prior to the RCVTP.

Results and Discussion

Wilcoxin signed-rank tests for matched pairs were computed on the initial and final performance rating data.

Data for the initial and final performance measures. As shown in Table 3, a total of 359 subtasks had at least two ratings. Based on these frequencies, the percentage of subtasks with train to sustain ratings increased from 61.8% to 78.6 %. Furthermore, when the subtasks with ratings of train to improve were compared to the subtasks with ratings of train to sustain, a significant majority (74.6%) of them were train to sustain.

These units seemingly thus became more proficient in these subtasks as their training progressed. This observation was confirmed by the data analyses as significantly more subtasks were included in the improve/sustain category as compared to the subtasks included in the improve/improve category. Also, significantly more subtasks were included in the sustain/sustain category than in the sustain/improve category. (See Table 3 for the results of the statistical tests.)

Data analyses also revealed that the armor and mech/scout platoons were more likely to improve than were the armor companies. In subtasks rated train to sustain, for example, the

Table 3

Initial and Final Subtask Rating Counts by Unit Type

| Type of unit | $n$ | Improve-sustain | Sustain-sustain | Improve-improve | Sustain improve |
|---|---|---|---|---|---|
| Armor Platoons | 17 | 46 | 90 | 20 | 12 |
| Mech/scout Platoons | 11 | 20 | 45 | 6 | 6 |
| Armored Companies | 10 | 25 | 56 | 20 | 13 |
| Total | 38 | 91 | 191 | 46 | 31 |

Table 4.

Wilcoxon Signed-Rank Tests of Numbers of Subtasks in Categories Based on First and Last O/C Ratings

| Type of Unit | $n$ [a] | $N$ [b] | $T$ | $Z$ |
|---|---|---|---|---|
| Improve/Sustain versus No Change[c] | | | | |
| All units | 38 | 31 | 115.50 | 2.60*** |
| Armor platoons | 17 | 14 | 15.00 | 2.35** |
| Mech/scout platoons | 11 | 8 | 6.00 | 1.68* |
| Armor companies | 10 | 9 | 19.50 | .35 |
| Sustain/Improve versus No Change | | | | |
| All units | 38 | 34 | 8.00 | 4.95*** |
| Armor platoons | 17 | 16 | 0.00 | 3.52*** |
| Mech/scout platoons | 11 | 8 | 2.00 | 2.24** |
| Armor companies | 10 | 10 | 1.00 | 2.70*** |
| Improve/Sustain versus Sustain/Improve | | | | |
| All units | 38 | 32 | 60.00 | 3.81*** |
| Armor platoons | 17 | 15 | 8.00 | 2.95*** |
| Mech/scout platoons | 11 | 9 | 7.00 | 1.83* |
| Armor companies | 10 | 8 | 7.00 | 1.54 |

[a] Number of units. [b] Number of non-zero differences. [c] No Change category includes both sustain/sustain and improve/improve sets of ratings.
*$p < .10$. **$p < .05$. ***$p < .01$.

armor companies' gain from first to last rating (60.5% to 71.1%) was about half that of the armor and mech/scout platoons' gain (6?.4% to 82.0%).

Data for the first and later subtasks by unit type. As shown in Figure 1, the sustain percentage for these units' first performance of a subtask varied around 59.9%, with only a modest increase across successive tables. This trend suggested that some small generalized transfer effects helped offset an expected decrease in performance when the more difficult subtasks were initially encountered in later training tables.

Their performances for later occurrences of a subtask increased substantially for their third, fourth, and seventh training tables. Small increases were found for their fifth and sixth training tables. These findings provide further evidence that these units became more proficient as the result of practice afforded by the RCVTP. Correspondingly then, the trends found for Assessment A were not just a function of the units learning to use SIMNET.

Summary of Assessment B. This assessment provided additional evidence for the training effectiveness of the RCVTP. Hence, this assessment has thus provided some answers to the questions posed from Assessment A. Questions, however, still remained about the participants' feelings toward this training program.

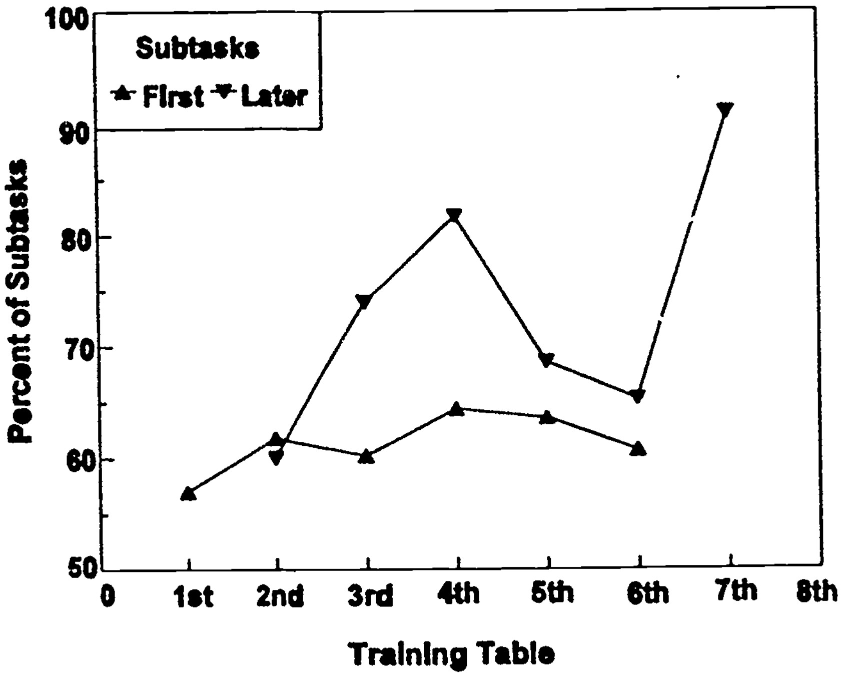


Figure 1. Percent of "first" and "later" subtasks with "train to sustain" by successive training tables.

Assessment C: The Units' Questionnaire Responses

This section is based on the formative assessment of the RCVTP as conducted by the instructional design team.

Method

Participants. Questionnaire data were collected on 280 participants from the developmental trials. This sample thus included participants from Assessments A and B.

Two hundred thirty-nine of these particpants were unit leaders, e.g., company commanders, platoon leaders, and tank commanders. These participants came from: (a) 19 armor companies, (b) 12 armor platoons, (c) 3 scout platoons, and (d) 3 mechanized infantry platoons. And they included 206 ARNG and 74 active soldiers, who were from armor companies.

Instrument. The instrument used for this assessment was a 40-item Likert-scaled self-report questionnaire. The items included the participants' perceptions regarding: (a) their proficiency **before** the training; (b) their proficiency **after** the training; (c) the training benefits of the RCVTP as compared to other SIMNET training experience; and (d) the various aspects of the RCVTP. The scale for these items ranged from seven as the most positive to one as the least positive with four as a neutral point. The participants were also given the opportunity to provide reasons for their answers to the different questions.

Data collection procedure. All ethical guidelines prescribed by ARI and the American Psychological Association were followed when the questionnaire was administered at the end of the participants' training.

Results and Discussion

Data regarding levels of proficiency. A significant difference was found regarding participants' estimates of pre- and post-RCVTP training proficiency levels, $t$ (238[2])=19.55, $p$ < .001. The participants, regardless of unit type, claimed to be more proficient after training (M=5.44) than they were before training (M=3.95).

Also, comparisons between unit types on a difference score (before training estimates minus post-training estimates)

---

[2] Number of participants is fewer than 280 because only the data from unit leaders were analyzed.

revealed a "significant interaction" between unit types and proficiency levels. As shown in Figure 2, leaders from the reserve company units indicated that they improved more than did their active counterparts. This interaction seemingly occurred because ARNG unit leaders claimed to be at a lower level of initial proficiency than did their active counterparts. This training program thus raised the confidence levels of the ARNG armor company leaders to the claimed post-training levels of the active company leaders.

Data regarding the participants' perceptions of the RCVTP.

Participants, regardless of their designation, believed that improvement in their unit's performance was a function of the RCVTP. Means of 5.43 and 5.54 were found for the items dealing, respectively, with improvement as a function of the time in the simulators and the AARs. They, furthermore, indicated that they became more proficient after this training than after their other SIMNET training experiences with a mean of 5.66 for this item. The questionnaire data also provided some insights into the participants' feelings about components of the RCVTP's instructional design. One, they felt that discovery learning did take place with mean scores of 5.70 and 5.75 for items dealing with their AAR comments helping them to improve on the platoon training tables and the company training tables, respectively. Two, these training tables were viewed as becoming more difficult as the training progressed. A mean of 5.55 was found for the item dealing with this issue.

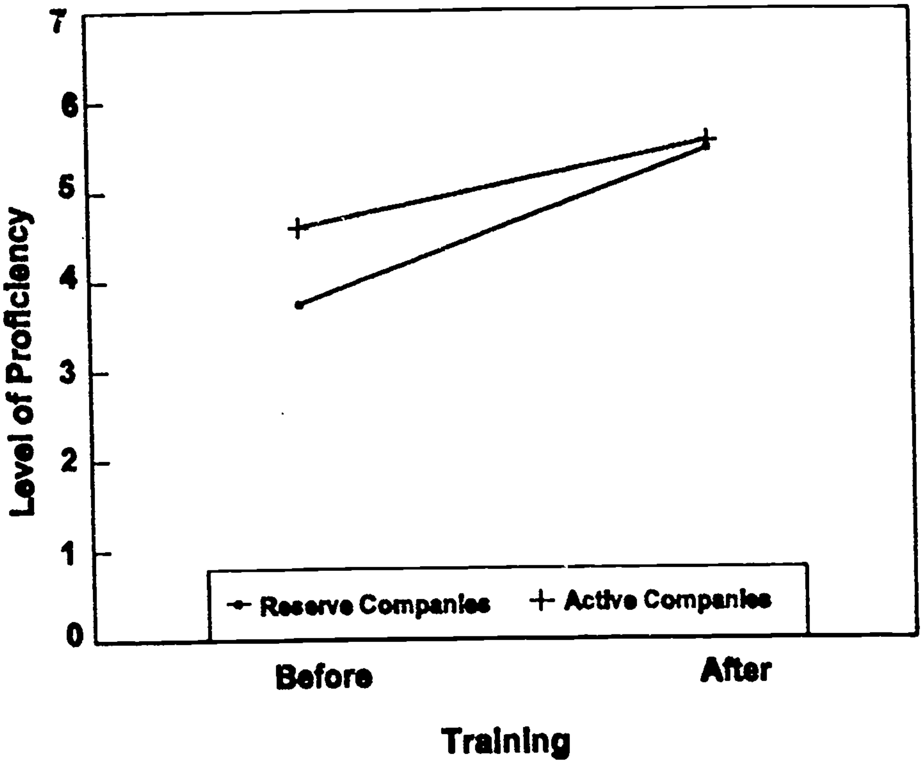


Figure 2 Means of proficiency estimates by unit leaders from reserve components and active armor companies.

Sample of participants' comments. As indicated by the quantitative data, the participants' comments tended to be positive. They were most appreciative of the training opportunity. One participant wrote:

> "We have no opportunity for company level maneuvering at home station and the opportunity for that here is priceless."

Another participant stated:

> "I believe that these missions with the simulators are the most effective training that I have had...I hope that we

receive more SIMNET training in the future."

Their comments also indicated that improvement was a function of the RCVTP. A participant claimed that during the RCVTP the unit was less likely to get lost in the SIMNET terrain than during their previous SIMNET training. Another wrote:

> "(RCVTP was) a very valuable training program. Ability to do a lot of movement in a short time."

There were a few negative comments, however. A participant from an armor company noted:

> "When a unit first arrives I believe that we went from crawl to run (basic to complex), instead of crawl, walk, run. It made it (the RCVTP) a little bit more difficult than (it) should be."

And a unit leader from an armor company wrote that the O/Cs should use their visual aids more during the AARs. These suggestions could help make the RCVTP an ever better instructional program.

Summary of Assessment C. This program's effectiveness has thus been established from the perspective of ARNG users. ARNG units seemingly then would like to utilize this program for their future collective tactical training. This assessment also provided insights into the reasons for the participants' positive perceptions of this program and possible ways of improving the RCVTP.

## Summary and Conclusions

Data from the different methods and sources indicate that the units developed their collective tactical skills across the training period. This evaluation has thus demonstrated the RCVTP's instructional value for helping tactical units to become more proficient.

### Value of Multimethod-Multisource Approach

This evaluation has also further demonstrated the value of employing a multimethod-multisource evaluation strategy for conducting naturalistic evaluations of high-technology based training systems. As stated, each method might have provided problematic data. The observational data, for example, were limited by their small sample size and the exclusion of defensive tables. Areas of agreement among assessments thus provided more valid conclusions than any single assessment method would have provided.

Also, each assessment yielded insights into this training situation from complementary perspectives. As indicated, the observational data reflected the perspective of evaluators who were independent of the instructional design and training processes; however, they were not subject-matter experts. The instructors were subject-matter experts but were part of the instructional process. The questionnaires tapped the users' perspective. Taken then from these different methods and sources, the evidence for RCVTP's effectiveness becomes more compelling.

Each assessment also provided complementary insights into the training situation. The observational data indicated that the participants were apparently able to attend to the cues inherent in the RCVTP tables without too much reliance on instructor prompting. Also, insights into users' feelings about their learning process were provided by the questionnaire data. Product and (some) process outcomes associated with the RCVTP were thus obtained in this evaluation; while other SIMNET evaluations (e.g., Shlechter, Bessemer, & Kolosh, 1991) have only obtained product outcomes.

Obtaining process outcomes provided these evaluators with further confidence in this evaluation's internal validity. Unlike Shlechter et al's (1991) SIMNET evaluation, this evaluation demonstrated that the cited improvements associated with the training system were not an artifact of additional instructor prompting.

Also, quantitative and qualitative data were obtained in this evaluation with the latter providing meaning to the former. As indicated, the participants' comments provided insights into their reasons for wanting to use this program. Adequate criterion measures were thus seemingly sampled in this evaluation.

Problems with this evaluation

These researchers had trouble collecting some processes outcomes, especially those assoicated with the AARS. That is, we are not able to assess the participants' ability to articulate the reasons for their actions. Perhaps, our problems with the AAR data might have been the result of trying to collect too much data. An observer, for example, claimed that the AARs went too fast for him to record all the requested information. Researchers must then not overwhelm their data collectors by having them collect too much data.

We should have also more carefully matched "data source" with the method. For instance, some of these observers professed to having problems judging the units' performance. One observer maintained that the RCVTP instructional personnel had to continually help him with his ratings. These observers, perhaps, were best suited for collecting the more objective data while performance judgments should have been left to the experts.

Closing statement

This investigation has further delineated the advantages of and problems with conducting multimethod-multisource evaluations of high-technology based training systems. As discussed, the use of multiple methods and sources has provided us with a better understanding of the RCVTP's effectiveness than could be provided by any single method and source. Also, this research strategy provides a viable approach to evaluating a high-technology based training system in a non-controlled context without the possibility of obtaining either baseline or transfer measures of performance. As previously stated, such evaluations may become more prevalent as the resources to conduct more controlled evaluations become more scarce.

These authors must finally address the problem of information regarding evaluation techniques. As indicated, it was nearly impossible to find any information on this topic in the different bibliographical data bases (DITIC, ERIC, or PSYCHLIT). Perhaps, a common source delineating the lessons learned from different evaluation techniques is needed. Otherwise, valuable research time may be lost as researchers are continually-"re-inventing the wheel."

References

Adams, J. A. (1978). On the evaluation of training devices. Human Factors 21(6), 711-720.

Adler, P. A. & Adler, P. (1994). Observational techniques. In N. K. Denizen & Y.S. Lincoln (Eds.)., Handbook of qualitative Research (pp. 377-392). Thousand Oaks, CA: Sage Publications

Bessemer, D. W. (1991). Transfer of SIMNET training in the Armor Officer Basic Course (Tech. Rep. No. 920). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 198)

Blaiwes, A. S. & Regan, J. J (1986). Training devices: Concepts and progress. In J. A. Ellis (Ed.). Military contributions to instructional technology (pp. 83-170). New York, Praeger Press.

Brehmer, B & Dorner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of field studies. Computers in Human Behavior 9, pp. 171-184.

Brown, R. E., Pishel, R. E., & Southard, L. D. (1988, April). Simulation networking (SIMNET) preliminary training developments study (PTDS) (TRAC-WSMR-TEA-8-88). (Available from the U.S. Army Armor School, ATTN: ATSB-DOTD-ORA, Fort Knox, KY 410121-5200)

Burnside, B. L. (1990). Assessing the capabilities of training simulators: A method and simulating networking (SIMNET) applications. (Research Rep. No 1565). Alexandria, VA: U.S.

Army Research Institute for the Behavioral and Social Sciences. (AD A226 354)

Burnside, B. L. (1982). Subjective appraisal as a feedback tool (Tech. Rep. No. 604). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A138 873)

Campbell, C. H., Campbell, R. C., Sanders, J. J., & Flynn, M. C. (1994). Methodology for development of structured simulation-based training (Research Product). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. Manuscript in Preparation.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.S. Resnick (Ed.), Knowing, learning and instruction: Essay in Honor of Robert Glaser (pp. 453-459). Hillsdale, NJ: Erlbaum.

Cook, T. D., & Campbell, D, T. (1979). Quasi-experimentation: Design and analyses issues for field settings. Chicago, Rand McNally.

Denizen, N. K. (1978). The logic of naturalistic inquiry. In. N. K. Denizen (Ed.), Sociological methods: A sourcebook. New York: McGraw-Hill.

Denizen, N. K. & Linclon, Y. S. (1994). Introduction: Entering the field of qualitative research. In N. K. Denizen & Y.S. Lincoln (Eds.)., Handbook of qualitative Research (pp. 1-17). Thousand Oaks, CA: Sage Publications.

Finley, D.L., Rhinehelder, R. W., Thompson, E. A., Sullivan, D. J. (1972, August). Training effectiveness evaluation of naval training devices. Part 1. A study of the effectiveness of a carrier air traffic control center training device. Orlando, FL: Naval Training Equipment Center. (AD 751 556)

Gagne, R. M. (1954). Training devices and simulators: Some research issues. The American Psychologist, 95-107.

Garvey, E. E., & Radgowski, T. (1988). SIMNET-D standing operating procedure (Report No. 6929). Cambridge, MA: BBN Systems and Technologies Corporation.

Garvey, E. E., & Radgowski, T. (1988). SIMNET-D standing operating procedure (Report No. 6929). Cambridge, MA: BBN Systems and Technologies Corporation.

Gound, D. & Schwab, J. (1988, March). Concept evaluation program of Simulation Networking (SIMNET) (Final Report, TRADOC TRMS No. 86-CEP-0345, ACN 86299). Fort Knox, KY: U.S. Army and Engineer Board. (Available from Commander, U.S. Arm Armor Center and Fort Knox, AATN: ATZK-DS, Fort Knox, KY 40121-5180)

Harrison, J.A., Acchione-Noel, S., Butler, W.G., Nantze, S.R., Walker, M. (1992). AH-64 combat mission simulator (CMS)/ cockpit weapon and emergency procedures trainer (CWEPT) cost and training effectiveness analysis (CETA). White Sands Missile Range, NM. (Available from Commander, U.S. Army TRADOC Analysis Command, AATN: ATRC-TD, Fort Leavenworth, KS 66027-5200).

Hays, R.T. & Singer, M.J. (1989). Simulation fidelity in training system design: Bridging the Gap between reality and training. New York: Springer-Verlag.

Herrmann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. Psychological Bulletin, 92, 343-452.

Kendall, M. K. (1975). Rank correlation methods (4th Ed.). London: Charles Griffin & Company LTD.

Kraemer, R. E. & Wong, D. T. (1992). Evaluation of a prototype platoon gunnery trainer for Armor Officer Basic Course training (Research Report 1620). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A254 289)

Kraemer, R. E., & Bessemer, D. W. (1987). U.S. tank platoon training for the 1987 Canadian Army Trophy (CAT) competition using a simulation networking (SIMNET) systems (Research Report 1457). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A191 076)

Kraiger, K., Ford, J. K., & Salas, E. (1993). Applications of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology, 78(2), 311-328.

Lesgold, A., Lajoie, S., Bunzo, M.,, and Eggan, G. (1988). Sherlock: A coached practice environment for an electronics troubleshooting job. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

McAnulty, D. M. (1992). Effectiveness of the AH-1 flight and weapons simulator for sustaining aerial gunnery skills. (Research Report No 1612). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A250 810)

Orlansky, J., Dahlman, C. J., Hammon, C. P., Metzko, Taylor, H. L., Youngblut, C. (1994). The value of simulation for training. (IDA Paper P-2982). Alexandria, VA: Institute for Defense Analyses.

Patrick, J. (1992). Training: Research and practice. London: Harcourt Brace Jovanivich, Publishers.

Patton, M. Q. (1987). How to use qualitative methods in evaluation. Newbury, CA: Sage Publications, Inc.

Schendel, J.D., Morey, J.C., Granier, M. J. & Hall, S. (1983). Use of self-assessments in estimating levels of skill retention (Research Rep. No. 1341). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A141 042)

Shlechter, T. M., Bessemer, D. W., & Kolosh, K. P. (1991). Role-playing and the effectiveness of a high technology training simulator (Tech. Rep. No. 938). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A224 913)

Shlechter, T. M. Bessemer, D.W., Nessleroade, K.P., Anthony, J. (1995). An initial evalution of a simulation-based training program for Army national guard units. Alexandria, Va: U.S. Army Research Institute for the Behavioral and Social

A manuscript in preparation.

Shute, V. J. & Regian, J. W. (1993). Principles for evaluating intelligent tutoring systems. Journal of Artificial Intelligence and Education, 4, 245-271.

Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electromechanical troubleshooting skills. Human Factors 33(3), 309-323.

TEXCOM Combined Arms Test Center (1990, August). Close combat tactical trainer (CCTT) force development testing and experimentation (FDTE) (Test Report No. FD 0200). Fort Hood, TX: Author. (Available from Headquarters, Training and Doctrine Command, AATN: ATIS-S, Fort Monroe, VA 23561)

van Berkum, J. J. and DeJong, T. (1991). Instructional environments for simulations. Education & Computing, 6, 305-388.