LEVEL II

AIR 2/79-66200-FR

AD A069078

RESEARCH NOTE 79-14

PERFORMANCE EFFECTIVENESS IN COMBAT JOB SPECIALTIES: ADDITIONAL STUDIES

Andrew M. Rose
Tetsuro Motoyama
Ted W. Allen
American Institutes for Research

and

Michael Drillings
Army Research Institute for the
Behavioral and Social Sciences

February 1979

Prepared for: U.S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Prepared by: American Institutes for Research
1055 Thomas Jefferson St., N.W.
Washington, D.C. 20007

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Note 79-14

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Performance Effectiveness in Combat Job Specialties: Additional Studies

5. TYPE OF REPORT & PERIOD COVERED: Final Report

6. PERFORMING ORG. REPORT NUMBER: AIR 2/79-66200-FR

7. AUTHOR(s): Andrew M. Rose, Tetsuro Motoyama, Ted W. Allen, Michael Drillings

8. CONTRACT OR GRANT NUMBER(s): DAHC 19-77-C-0020

9. PERFORMING ORGANIZATION NAME AND ADDRESS: American Institutes for Research, 1055 Thomas Jefferson St., N.W., Washington, D.C. 20007

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Ave., Alexandria, VA 22333

12. REPORT DATE: Feb[illegible]1979

13. NUMBER OF PAGES: 90

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

AIR-2/79-66200-FR ARI RN-79-14

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Hands-On Performance	Performance Effectiveness
Predictor	Duty Position
Combat Readiness	Job Congruence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Army faces a potential problem in the maintenance of trained skills, particularly in the Combat Arms. In peacetime, the soldier's assignments may result in day-to-day activities which are, to varying degrees, different from the tasks he might be responsible for in a combat situation. This discrepancy between what tasks the combat soldier is responsible for and what he actually does as part of his peacetime duty assignment could possibly reduce his effectiveness when required to perform combat-related tasks.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

023 450

#20 (cont'd.)

Given this situation, it would appear likely that soldiers whose on-the-job behaviors (tasks) approximate the behavior (tasks) they will be called upon to perform in combat would be more likely to demonstrate proficiency than their counterparts whose job activities were far removed from their projected combat roles. A two-phase exploratory study of this phenomenon and of its consequences for individual training and evaluation was conducted. The major findings of the first phase were that the performance of some soldiers (primarily those in the 11E driver and loader positions) was predictable given knowledge of their particular background and job experience. Also, predictions were possible in many cases given information between an individual's day-to-day activities and his projected combat tasks. However, these findings were unsystematic: the kinds of predictors that were operating and the strength of their relationships varied for different 11E crew positions and for different tasks within each crew position. Therefore, a second phase of research concerned with unraveling these complexities was required.

During the first phase, personal, Army, and job background information was obtained from approximately 200 11E crewmen. Also collected was information about the congruence between an individual's day-to-day activities and the combat activities associated with the crew position in which he was to be tested. All troops were then given a series of hands-on performance tests.

During both phases of this study, the obtained data were case into several types of analyses in which background and task congruence variables were examined as predictors of four kinds of performance measures: scores from the hands-on performance test, scores from a readministration of that test, SQT scores, and confidence estimates obtained from the soldiers regarding their ability to perform tasks included in the performance test. Four separate sets of analyses were carried out, one for each crew position. Additional analyses were performed to examine the relationships among the performance measures, as well as to evaluate a task taxonomy for its usefulness and method for organizing the obtained results.

The major finding of this study is that different aspects of performance (as measured by confidence estimates, hands-on performance tests, and SQT scores) could be predicted to a limited degree from a relatively small number of variables. These variables primarily reflect a soldier's specific experience in his crew position. However, significant predictions could only be made for individual tasks; predictions varied for different tasks in any given crew position. Thus, the relationship between a soldier's day-to-day activities and his performance on projected combat tasks, while somewhat more systematic than previously found, is still complex and is task- and soldier-specific.

EXECUTIVE BRIEF

Requirement

The Army faces a potential problem in the maintenance of trained skills, particularly in the Combat Arms. Training programs are designed to establish proficiency in a particular MOS, using as a basis the tasks which will be required of a soldier in a combat environment. Yet, in peacetime, the soldier's assignments may result in day-to-day activities which are, to varying degrees, different from the tasks he might be responsible for in a combat situation. This discrepancy between what tasks the combat soldier is responsible for and what he actually does as part of his peacetime duty assignment could possibly reduce his effectiveness when required to perform combat-related tasks.

Given this situation, it would appear likely that soldiers whose on-the-job behaviors (tasks) approximate the behavior (tasks) they will be called upon to perform in combat would be more likely to demonstrate proficiency than their counterparts whose job activities were far removed from their projected combat roles. The Army initiated a two-phase exploratory study of this phenomenon and of its consequences for individual training and evaluation in May of 1977. The major findings of the first phase were that the performance of some soldiers (primarily those in the 11E driver and loader positions) was predictable given knowledge of their particular background and job experience. Also, predictions were possible in many cases given information between an individual's day-to-day activities and his projected combat tasks. However, these findings were unsystematic: the kinds of predictors that were operating and the strength of their relationships varied for different 11E crew positions and for different tasks within each crew position. Therefore, a second phase of research concerned with unraveling these complexities was required.

Procedure

During the first phase, personal, Army, and job background information was obtained from approximately 200 11E crewmen. Also collected was information about the congruence between an individual's day-to-day activities and the combat activities associated with the crew position in which he was to be tested. All troops were then given a series of hands-on performance tests.

During both phases of this study, the obtained data were cast into several types of analyses in which background and task congruence variables were examined as predictors of four kinds of performance measures: scores from the hands-on performance test, scores from a readministration of that test, SQT scores, and confidence estimates obtained from the soldiers regarding their ability to perform tasks included in the performance test. Four separate sets of analyses were carried out, one for each crew position. Additional analyses were performed to examine the relationships among the performance measures, as well as to evaluate a task taxonomy for its usefulness as a method for organizing the obtained results.

Findings

The major finding of this study is that different aspects of performance (as measured by confidence estimates, hands-on performance tests, and SQT scores) could be predicted to a limited degree from a relatively small number of variables. These variables primarily reflect a soldier's specific experience in his crew position. However, significant predictions could only be made for individual tasks; predictions varied for different tasks in any given crew position. Thus, the relationship between a soldier's day-to-day activities and his performance on projected combat tasks, while somewhat more systematic than previously found, is still complex and is task- and soldier-specific.

Utilization of Findings

Troops assigned to duty positions that are not directly related to their presumed combat roles can be expected to show decrements in the performance of combat related tasks. While reassigment of these soldiers may not be practicable, attention must be given to upgrading and maintaining their combat skills through more frequent hands-on practice and through the use of aids such as Soldier's Manuals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (cont'd.)

INTRODUCTION

Personnel whose job specialties lie within the Combat Arms typically receive extensive training for a specific combat role, and are periodically required to demonstrate their continued mastery of combat skills. For example, tank crews in the U.S. Army are required to qualify annually in gunnery by adequately performing a series of live-fire engagements (Table VIII). Likewise, individual crewmen are required to demonstrate their continued proficiency by taking written and hands-on skill qualification tests (SQTs). These demonstrations of combat readiness are required regardless of each crew member's peacetime job assignments; a criterion level of proficiency is demanded whether or not the day-to-day activities that constitute the soldier's normal job approximate the tasks he may be called upon to perform in a combat role.

In many instances, discrepancies exist between a soldier's day-to-day activities and the kinds of tasks for which he must be prepared in a combat environment. Such discrepancies may arise for several reasons. For example, the high costs associated with constant exercising of troops in their combat roles may preclude such practice as a full-time activity. Similarly, during peacetime there are many rather routine but nevertheless essential tasks which must be performed to ensure a garrison's smooth operation; such assignments are often given to combat troops. These discrepancies, between what tasks the combat soldier is responsible for and what he actually does as part of his peacetime duty assignment, could possibly reduce his effectiveness when required to perform combat-related tasks.

Given this situation, it would appear likely that soldiers whose on-the-job activities approximate the activities they will be called upon to perform in combat would be more likely to demonstrate high levels of proficiency than their counterparts whose job activities were far removed from their projected combat roles. Were such a hypothesis confirmed, procedures would need to be developed for enhancing and maintaining the combat-task skill levels of those soldiers who are primarily engaged in non-combat-related tasks during peacetime.

The Performance-Based Training Work Unit of USARIBSS' Individual Training and Skill Evaluation Technical Area has supported an exploratory investigation to answer a limited set of questions related to the hypothesis.* The first phase

---

*Rose, A. M., and Wheaton, G. R. Performance effectiveness in combat job specialties. (AIR 6/78-66200-FR). Washington, D.C.: American Institutes for Research, June 1978.

of the study compared the combat-related performance of soldiers whose peacetime duty assignments approximated their required activities in combat with the performance of soldiers whose peacetime activities diverged from their combat-projected tasks. By intent and design the first phase of the investigation was limited in scope: its major purposes were to explore that a problem existed and to try out data collection and analysis methods. These limited goals were achieved, and several additional potentially important research issues were uncovered. The exploration of these issues constitutes the second phase of the study.

This report will first review the recently completed effort. The review will briefly present the data collection methodology, the analyses previously conducted, and results obtained. Following this review, we will present and discuss the additional research issues which form the focus of the second phase of the research.

The remainder of the report is divided into four sections. The first of these, "The Data Set and Preliminary Analyses", describes the initial activities of this phase of the research and presents the updated and additional information that was incorporated into the data base. Following this, the second section, "Congruence Effects", presents and discusses the series of analyses undertaken to explore the basic congruence-performance issue. This is followed by a section entitled "An Alternative Task Taxonomy", which is concerned with the development and evaluation of a speculative method for organizing the obtained results so as to increase the generalizability of the findings to other MOS categories. The final section is a brief summary of the findings and a discussion of their implications for the Army.

SYNOPSIS OF THE FIRST PHASE OF THE PERFORMANCE EFFECTIVENESS STUDY

Approach

The first activity in the initial phase of the effort was to select a Combat Arms MOS within which to examine empirically the job-discrepancy/performance-effectiveness issue. Conceptual development and elaboration of the measurement concepts and variables was the objective of the second task. Development of a variety of data collection instruments constituted the third task. Included among these instruments were: (1) task checklists for use in developing measures of a class of practice variables, namely, measures of frequency and recency of task performance; and (2) questionnaires designed to tap into variables presumably reflecting job congruency as well as other background information potentially predictive of performance. The fourth major activity involved the selection of those combat-related tasks on which estimates of performance effectiveness would be obtained. In the fifth task the various instruments were used to collect the indicated predictor and criterion performance data. Analysis of the obtained data comprised the sixth task.

Selection of MOS

Several criteria were considered in selecting an MOS on which to focus, including availability of soldiers, within-MOS combat-role variability, and variety of day-to-day activities of incumbents. As a result, the 11E Armor Crewman MOS was chosen for study and a battalion of 11E troops at Fort Carson, Colorado, was scheduled to participate in the data collection effort.

Selection and Elaboration of Variables

An attempt was made to assess the troops involved in this study along a fairly broad range of potential predictor variables. The intent was basically twofold: first, to determine the variables that predicted proficiency in combat-related tasks for the specific sample of troops involved in this study; and second, to measure some of the variables that might affect performance for a wider range of Army jobs.

There were basically four classes of independent variables: Personal Background, Army Background, Job Background, and

Congruence.* The first group of variables of potential interest involved what the soldier "brings" to the Army. These Personal Background variables were initially viewed as potential moderating or intervening variables for the prediction of proficiency.

Six personal background variables were selected for examination. These were:

- Level of education,
- Aptitude test score (measured as the mean of the test scores reported for a given soldier's aptitude test battery),
- Height,
- Weight,
- Handedness, and
- Age.

Two additional variables were selected that were not strictly personal background as described above, but were hypothesized to potentially play the same moderating or intervening function. These concerned a soldier's general attitude toward the Army and his present job. The soldiers were asked:

- "Do you like your present job?", and
- "Do you plan to reenlist?"

Both were presented as yes-no questions.

The second group of variables involved a soldier's Army Background. These variables were included for exploratory purposes--the sample of 11E soldiers might yield an indication of the types of relationships which existed, and whether these relationships changed for different positions or jobs.

---

*For more details regarding the rationales for variable inclusion and units of measurement, the reader is referred to Rose and Wheaton (op. cit.).

Five variables were chosen as indicators of Army Background. These were:

- Years in active military service,
- Rank,
- BCT: whether or not a soldier had gone through a normal Army progression, including Basic Combat Training,
- AIT: whether or not a soldier had completed Advanced Individual Training, and
- Relevance of other Army training.

The third set of variables was concerned with a soldier's Job Background. These variables were selected to indicate a particular soldier's history and experience with respect to the 11E MOS.

The five variables selected were:

- Time in 11E MOS,
- 11E skill level,
- Time in crew,
- Number of different positions held during tank gunnery programs, and
- Number of Table VIII's, regardless of position held.

The fourth set of independent variables was selected to measure various aspects of job-position correspondence. In essence, these Congruence variables attempted to assess what a soldier *does* on a day-to-day basis rather than what his assigned combat position would be. Each of these variables was tied to the particular duty position in which the soldier was to be tested. These congruence variables were viewed as the primary independent variables of this study; it was hypothesized that the relationship between these variables and performance would: 1) indicate if there was a potential problem for the Army due to the job-position mismatch; 2) define the general nature and scope of the problem, and 3) provide an indication of the nature of the relationships and possible mediating circumstances.

Eight Congruence variables were selected for exploration. These were:

- Time in duty position,
- Duty position skill level,
- Duty position congruence (whether the soldier's duty position was the same as his position for the current performance test),
- Position training congruence (whether the soldier had held his tested position during tank gunnery programs),
- Table VIII position congruence,
- Time in tested position,
- Hours per day in tested position, and
- Platoon Sergeants' ratings.

The final two predictor variables were concerned with each soldier's particular practice experience for the hands-on performance tests of this study. For each task on which he would be tested, the soldier was asked to indicate how often he practiced it and the last time he had performed the task.

Data Collection Instruments

Two, paper-and-pencil, group administered instruments were developed to obtain detailed information on the background and task congruence variables that were to be used as predictors of combat-related task performance. These were, respectively, the Background Information Questionnaire (BIQ) and the Job Information Survey (JIS).

Background Information Questionnaire. The BIQ consisted of two parts. The first, comprised of some 17 questions, required short fill-in answers or multiple-choice responses. The intent of this first set was to develop information about troops' personal background (e.g., age, height, weight, level of education), Army background (e.g., years in service, rank), job background and "congruence" (e.g., job title, position assigned to for the research, time spent in that position). The second part of the BIQ was designed to elicit confidence estimates from troops concerning how well they could perform the tasks which they would subsequently be tested on. While not exactly equivalent, the tasks presented for rating were in essence composites of those included in the subsequent hands-on tests and covered in the JIS instrument.

Job Information Survey. The JIS was designed to provide data on practice effects, including how frequently and how recently troops had practiced or performed those combat-related tasks of interest in the study. A given soldier was to respond only to that portion of the JIS related to the crew position to which he had been assigned for the study. Opposite each task listed under the appropriate crew position in the JIS, the soldier placed one checkmark in the column that best described how frequently he performed it, and another in a column that indicated when he had last done it.

Hands-On Performance Tests and Measures

The primary source of criterion data was a set of hands-on performance tests relating to each crew position in the M60AIAØS tank. The tasks, collectively known as the Tank Crew Skills Test, had been developed by HumRRO under contract to the USARIFU at Fort Knox, Kentucky, and represented an adaptation of earlier work* on the M48A5 tank system. The purpose of the test was to portray the readiness of individual crewmen, focusing in particular on specific aspects of performance in need of remedial training or practice. As a consequence, the scoring procedures were oriented around a step-by-step appraisal of performance within any one task. This made it meaningful to portray performance in terms of the proportion of steps passed or failed within the task and to indicate precisely which steps had been failed. At a slightly more general level it was possible to describe performance on the overall task as successful (GO) or as needing some type of remediation (NO GO). Finally, one could characterize performance in terms of the proportion of all tasks that were GO. Since the tests were being administered to tank crewmen at Fort Carson, Colorado, as part of a readiness and remediation study, arrangements were made to support that effort and to gain access to the resulting performance data.

*Harris, J. H., Osborn, W. C., & Boldovici, J. A. Reserve component training for operating and maintaining the M48A5 tank. ARI Technical Report TR-77-A14 (NTIS: ADA043057), 1977.

Data Collection

Subjects. A total of 216 troops from the 2nd Battalion, 34th Armor, 4th Infantry Division, garrisoned at Fort Carson, Colorado, took part in the data collection effort. These soldiers represented the 2nd Battalion's full complement of 54 four-man crews. The soldiers were organized according to a Battalion Battle Roster which designated the crew to which they would belong and the crew position (i.e., driver, loader, gunner, or tank commander) which they would occupy during subsequent testing. The data collection effort of the present study focused on these assigned positions. The Battle Roster was primarily comprised of troops from the battalion's three line companies, each of which fielded 17 crews, and secondarily from its Headquarters Company which furnished three additional crews.

Procedure. Data collection was accomplished under the direction of U.S. Army Research Institute Field Unit (USARIFU) and HumRRO personnel from Fort Knox, Kentucky as part of a larger research effort involving tank crew gunnery. These individuals established liaison with the 2nd of the 34th, monitored development of the subject samples, trained test administrators, supervised collection of the hands-on performance test data, and administered background information questionnaires and job information survey instruments. USARIBSS personnel participated in the planning and orchestration of the effort while staff from AIR focused on troops' completion of the questionnaires and surveys and on the retrieval of additional background data from the soldiers' personnel records.

Data were collected during two waves, each lasting approximately one week. Half of the sample of 216 troops was tested during each week. Approximately 36 troops (i.e., nine crews) were processed on each of two days. In general, troops first responded to the BIQ, completing it for the crew position to which they had been assigned. They then reported to the hands-on performance testing location. One to three days after testing, troops completed the JIS instrument which provided frequency and recency of practice data about the hands-on tasks which they had performed earlier in the week.

Completed questionnaires, test booklets, release forms, surveys, etc., were organized on a subject-by-subject basis to create an initial raw data file structure. With the aid of this information three additional steps were taken. First, soldiers whose files were incomplete were followed up in an attempt to reduce missing data. Second, release forms were taken to the Enlisted Personnel Records Section where personnel folders were examined in order to extract specific kinds of background and aptitude data. Third, for all tested troops, brief interviews were arranged with their platoon sergeants. These supervisory personnel described and provided insights into the normal duty activities of the troops comprising the study sample on an individual-by-individual basis. They also provided estimates of task criticality that were to be used as criteria for selecting a subset of tasks for intensive analysis. Essentially the same data collection procedures were used in the second wave which occurred two weeks later.

Data Analysis

Data collected in the field were subsequently scored, transferred to coding sheets, and punched onto cards for computer processing. The resulting data base was of formidable size and complexity even when attention was limited to those variables of direct relevance to the current study. Some form of predictor or criterion data was available on 200 soldiers; these 200 cases clustered naturally into four smaller and independent crew position data bases. Instances of missing data were frequent. Within any given data base there were as many as 28 potential predictor variables which could be related to as many as 27 hands-on performance tests; finally, three different kinds of dependent criterion measures were available, including proportion-of-steps-passed and GO/NO-GO measures on each test and a more global measure representing the proportion of tasks within each crew position that were GO.

To resolve the complexities inherent in the data base and to reduce the amount of analysis to manageable size, the decision was made to treat each crew position separately, viewing the four sets of analyses as replications of the basic task congruence and performance effectiveness theme. Within each replicate a three-pronged analytical strategy was pursued. First, a principal components factor analysis with rotation to a varimax criterion was used to explore the dimensionality of the personal, Army, and job background predictor variable set. The outcome of this analysis was to be used to specify a potentially smaller set of less redundant variables with which to work. Second, a comparable analysis of the criterion tests was undertaken within each crew position, the dependent variable being proportion of steps passed within each test. Again the objective was to reduce the set of 100 tasks to a more reasonable size. Given predictor and criterion variable sets, the third step was to explore their interrelationships by means of multiple-regression techniques.

The results of these analyses are presented and discussed in full in a recent Technical Report (Rose & Wheaton, op. cit.). The following section briefly describes some of the more pertinent findings.

Results

Descriptive Information. As the first step in the data analyses, descriptive information was generated for the four crew positions on each of the background and congruence variables. The general "picture" of a driver as indicated by the personal and Army background information was of a high-school-educated 21-year old who had been in the Army for a little over two years, had gone through BCT and AIT and had achieved a rank of E3 or E4. With regard to his job background, he had held the title of 11E20 for approximately one and a half years, had worked in a crew slightly longer, had held one position during training, and had been tested on Table VIII once. The same picture was generally true for loaders, except it appeared that some of them had "cycled through" the driver position. Thus, as compared to drivers, loaders had been in the Army for almost one year longer, had a slightly higher rank, had been in the 11E MOS and in crew longer, and half of them had gone through more than one position during training and Table VIII testing.

In addition to being older and having more education, the gunners were more experienced than either drivers or loaders. They had achieved a higher rank and skill level and had held the 11E MOS for more than two years. They had been in a crew longer and had gone through previous training and testing cycles in two different positions.

Tank commanders were substantially different from the other positions on several variables. They were better educated, older, had been in the Army much longer (seven years versus two to three for the other positions), and had achieved higher ranks than the other positions. Also, they had a higher probability of reenlisting and liked their job more than the other positions.

Congruence Variables. The position-job match variable is the purest but least sensitive congruence variable, since the question is simply whether or not a soldier is performing his job in a position to which he was assigned for the research. All crew positions (with the exception of loaders) were filled by 80-87% "appropriate" personnel. It is difficult to say whether this is a high proportion, relative to the range of Army-wide combat positions; furthermore, this is only a gross measure, since it does not assign a congruence score to any particular referrent set of tasks for a position. On the other hand, the 80-87% figures might be artificially high for 11E's compared to other Army MOSs due to the fairly restricted range of positions in the MOS. The relatively low position-job match score for the loaders (58%) may indicate that loaders have a fairly narrow and delimited range of tasks unique to the position. Thus, during peacetime, it is more likely that loaders will perform other non-specialized crew duties; or perhaps for the purposes of this research, the loader position was viewed as least demanding on unexperienced personnel who were tapped to fill out crews by manning this position.

The two questions which asked whether or not the soldier had gone through a training cycle in his tested position and whether or not he had gone through a Table VIII in that position produced highly similar responses across the four positions. About one half of the soldiers responded positively to each question, with the only exception being the tank commanders' Table VIII experience. Over 80% said that they had gone through the Table VIII as a tank commander. It is possible that these training and Table VIII numbers are not representative of 11E's Army-wide in that these particular companies had recently completed a fairly extensive (but unrelated to this study) training and testing program.

The hours-per-day variable indicated that loaders and gunners spent less time on their position-related tasks than did drivers and tank commanders. This finding is consistent with the hypothesis that loaders and gunners are restricted by the nature of their jobs: loaders cannot practice loading, nor can gunners practice firing, due to resource constraints.

Intercorrelations of Background and Congruence Variables. The next stage of the analyses was designed to examine the relationships and interrelationships among the background and congruence variables. There were two purposes for this step: first, the zero-order correlations could be of interest in their own right; and second, as mentioned above, it was necessary to resolve the complexities inherent in the data base and to reduce the data to a manageable size. Given that the general analytic design was a regression paradigm, it was important to carefully select independent variables. One selection strategy was to screen a large set of variables and delete redundant and irrelevant ones (with respect to performance prediction). This could be done from an examination of the intercorrelation matrixes and the factor analysis of these matrixes.

The net result of these analyses was to eliminate several variables as predictors of hands-on performance. The variables omitted from the regression analyses were: aptitude test scores mean, height, weight, handedness, "like-job," and "reenlist" from the Personal Background variables, and BCT, AIT, and relevance of other training from the Army Background variables. Two others dealing with secondary MOS were also eliminated because they were found to have zero variance in the sample.

Recency and Frequency of Task Performance. The final two predictor variables were recency and frequency of task performance. They were treated separately not only because they varied within crew position on a task-by-task basis, but also because they represented conceptually distinct aspects of congruence, worthy of study in their own right.

In summary, the recency and frequency data ran the gamut across the 100 tasks comprising the hands-on readiness test battery. Many tasks had not been practiced by appreciable proportions of each crew position sample. This lack of practice might be due to troops who were assigned crew positions for the study that they did not normally occupy or that they had occupied for a relatively short period of time. At the other extreme were tasks on which performance was relatively frequent and which had been practiced quite recently. With the exception of drivers, however, who had substantial practice in driving, the most frequently performed tasks tended to deal with preparation for operations and maintenance. This lack of intensive practice on more tactical tasks is presumably characteristic of combat troops in peacetime.

Hands-On Performance Tests. Because of the exploratory thrusts of the effort and the need to conserve resources it was decided that during the first phase, a subset of tasks within each of four crew positions would be examined rather than analyzing the full set of 100 tests. Accordingly, an analytical strategy was devised and implemented for pinpointing those tasks having initial promise in addressing the task congruency and performance effectiveness hypothesis.

The first step in selecting criterion tasks was to score performance on each of the 100 in terms of the proportion of steps (scorable units) successfully completed. These scores were used in the initial analyses in order to maximize the yield of diagnostic information about each task and to promote variability in troops' performance scores, the latter being an important consideration in trying to create an optimal set of conditions under which to explore the hypothesized relationship between task congruence and performance.

Descriptive analyses run on the tasks comprising each crew position revealed several interesting findings. Performance levels were characteristically high (ranging from a mean on one driver task of .57 to a mean of .99 for a commander task) and showed relatively little variability among soldiers (e.g., variances ranged on individual tasks from .0009 to .25). The latter statistic revealed much more consistent performance in the sample of troops than had been anticipated, based on the number of "fills" in the study, the range of time in duty position, the fact that many troops reported not having performed several of the tasks during the previous year, and, when they had done so, the varied patterns of practice which seemed to typify the sample.

In order to select specific tasks to focus on in each crew position, the factor-analytic approach used earlier to screen predictor variables was again pursued. The proportion-of-steps-passed performance data were intercorrelated for all tasks within each crew position, and the four intercorrelation matrixes were then examined by means of a principal components solution, with rotation of factors to a varimax criterion. Other considerations for the final selection of tasks revolved around their judged criticality (the more critical being more desirable), the number of steps involved, and on their variances (greater variance being an asset in the eventual regression analyses). A sample of 19 tasks, or roughly 20% of the total set, was selected for examination in the subsequent regression analyses. Five tasks were used to represent drivers, six came from the loader set, four were drawn from gunner tasks, and four were used to represent the tank commander position.

Zero-Order Validities. As the final step in preparation for the multiple-regression analyses, the zero-order validity (correlation) coefficients between the predictor and criterion variables within each crew position were computed. The predictors were comprised of background, congruence, and practice variables. Criterion measures were proportion of steps passed, proportion of troops acquiring a GO on each task, and the proportion of all tasks within a position on which a GO was received.

With respect to the proportion of steps passed by task, relatively few highly significant ($p < .01$) relationships were obtained. Those relationships which did exist were confined primarily to the driver and loader positions. In particular, a number of significant relationships were found in the latter position. These involved job background, congruence, and Army background predictor variables. This same general pattern was repeated with respect to the task and overall GO proportions. Subsequent analyses were devoted to exploring these relationships for multiple as opposed to single predictor variables.

Multiple Regressions. Given the general exploratory nature of this first phase of the study, the large number of potential predictor variables, and the relatively small (statistically) sample of soldiers, the selection of an adequate regression technique posed some problems. After considering several alternatives, the "all possible subsets regression" procedure was selected. This analysis estimates regression equations for "best" subsets of predictor variables and does extensive residual analysis. Best is defined in terms of the sample R-squared, adjusted R-squared, or Mallows' C_p.

Turning now to an examination of the results, there were several fascinating and somewhat unexpected findings. The first and most critical issue is whether the entire set of variables actually predicted a significant proportion of variance in performance. The answer was clearly yes. Thus, despite the grossness of the scoring rule (i.e., overall proportion of tasks for which the soldier passed all steps, regardless of the number of steps in any particular task or substantial differences in difficulty among tasks), there was predictable variance for each crew position.

The second and equally critical issue was whether the congruence variables were related to performance, this being the basic experimental hypothesis of the research effort. If, for example, none of these variables (singly or collectively) predicted hands-on task performance, the initial supposition that a potential problem exists for peacetime soldiers in non-combat positions would be unjustified. However, it was apparent from the results that the congruence variables were involved and actually seemed to be the strongest predictors. Therefore, there was clear support for the project's major thesis: there is a strong relationship between a soldier's day-to-day activities and his proficiency on combat-related tasks. Furthermore, the nature of this relationship is in the logical direction: generally, the higher a soldier's congruence scores, the better his performance. A third aspect of the results was unexpected: frequency and recency of practice were, for the most part, nonpredictive of this performance measure.

Another intriguing aspect of the results was the pattern of predictors across crew positions. There were some similarities; for example, the most common predictor variable for all positions was on-the-job experience in duty position, measured either as months in Duty MOS or as months in tested position. However, there were some striking dissimilarities. For example, different background variables were important for different positions; furthermore, some of these changed in sign between positions. The most likely explanation unfortunately is the most difficult to demonstrate support for--that is, the differing demands on the soldiers. Some task performances are best predicted by a soldier's overall background and experience, others by specific job experience. Each job is a combination of a variety of different tasks; hence no generalizations (beyond those made above) are warranted.

Summary

The critical findings were that different variables predicted performance for different crew positions and for different tasks within positions. Furthermore, different scoring rules resulted in different sets of predictors. Thus, these results simply were not systematic enough to permit the development of generalizable findings.

However, during the course of the investigation a number of potentially important research issues were uncovered that related to the basic thrust of the study but which could not be pursued satisfactorily within the first phase of the effort. As summarized below, these were addressed during Phase II.

1. Effects of Task Congruity on Combat Task Performance. During Phase I, this central issue was addressed in terms of a small number of selected hands-on readiness test scores. During Phase II, the scope was broadened to encompass the more extensive information about performance that was available. Specifically, the task congruity hypothesis was reexamined using nearly all of the 100 readiness tasks that were administered.

2. Additional Performance Data. The hands-on performance tasks used in the present investigation are relatively new instruments and as such have not been studied or evaluated in detail. Given the availability of data from a second administration of these tests to many soldiers in the original sample, it would be important to compare their performance on both administrations. This analysis would indicate the extent to which scores are subject to variations in test administration, an artifact of potential concern. Also, related analyses could be conducted on the SQT performance data that were available. By extending the analyses to SQT data it would be possible to determine not only whether different peacetime duty activities are related to varying levels of proficiency on combat-related tasks but also to determine whether such relationships actually extend to skill qualification itself. The existence of task congruity effects on qualification, if demonstrated, has important and far-ranging consequences for the structuring of duty activities and refresher training.

3. Confidence Information. The soldiers' own estimates of their ability to perform different combat-related tasks could be compared to their actual levels of performance. If such ratings were indeed accurate reflections of subsequent

performance it would be possible to use them as proxy measures of a unit's combat readiness that pinpoint specific requirements for kinds and amounts of remedial training. If demonstrated to be valid estimates of performance, these ratings could also be used as dependent variables in an examination of the task congruity hypothesis. That is, these ratings could be evaluated as surrogate performance measures in an investigation of the congruence variables.

4. Task Taxonomy. One particularly intriguing aspect of increasing the number of tasks examined would be the possibility of uncovering systematic effects due to task "type". The relatively small number of tasks used in the first phase did not include sufficient samples of (for example) "procedural", "judgmental", and/or "cognitive" tasks so that relationships could be investigated. It is possible that a "task type" variable could untangle some of the inconsistencies among performance predictors observed in the limited data set studied in Phase I.

THE DATA SET AND PRELIMINARY ANALYSES

This chapter serves three basic functions: First, it presents the initial activities undertaken in the attempt to clarify the unsystematic relationships found in Phase I. Second, it presents the updated and new information in the data set; this information was used as input to the remaining analyses presented in this report. Finally, it discusses several issues not directly related to the congruence-performance question, that nevertheless are important in their own right. Such issues include an examination of the relationship between Readiness Test performance and performance on a readministration of the test, the relationship between soldiers' confidence estimates of their ability and actual performance, and the rationale for and development of an approach to improving the quality of the performance data which could incorporate practically the entire set of performance tasks administered during Phase I.

Background and Congruency Variables

One possible reason for the unsystematic relationships between background and congruency variables and performance as found in Phase I was the quality of the predictor variables. Analyses conducted in Phase I suggested that some of these variables might have been redundant or were inadequate reflections of what they were originally intended to measure. Furthermore, inconsistencies in individual soldiers' records (presumably based on misinterpretations of specific questions) could have led to the unsystematic results. Therefore, as an initial activity in Phase II, the existing data base was reexamined and edited.

The first minor editing consisted of rescaling of two variables: the Level-of-Education variable and the Years-Served variable. For the former, the six-point scale previously used was modified slightly to better reflect ordinal points in a soldier's educational progress. For the latter, Years Served was converted into months in an effort to increase the variability of scores (and thereby possibly increasing their potential value for prediction). As a result of the change in the Level-of-Education variable, there were minor changes in the descriptive statistics for each crew position as compared to the data presented in the Phase I Technical Report. The means and variances for drivers, loaders, and gunners increased slightly, while the mean and variance for tank commanders decreased slightly. The conversion of Years Served into months essentially multiplied the previous scores by 12. The original raw data, expressed in terms of months, were used.

The second minor editing consisted of deleting several variables from the data set. It was decided that these variables, due to their distributional properties and uninterpretability in previous analyses, were simply adding "noise" to the predictor variables; relationships might become clearer if they were omitted. Thus, the variables of Aptitude Scores Mean, Handedness, BCT, AIT, and Relevance of Other Training were dropped from the data set.

A final major alteration of the data set was prompted when a reexamination of the Tank Commander file revealed that 12 of the 59 tank commanders were officers. This fact had produced some "artifacts" in the original analyses. For example, since officers do not have an MOS designation in the EPMS, such variables as 11E Skill Level and Months in Primary MOS had no meaning, and were erroneously (or correctly) scored as zeros. In addition, officers follow a different career progression than enlisted personnel; it would be difficult at best to attempt generalizations regarding tank commanders based on background and experiential information from a group of officers. Thus, it was decided to delete these 12 officers from the file. In an effort to guage the potential impact of this deletion, descriptive statistics for the background and congruence variables for officers and non-officers were computed. Furthermore, for those variables which had variance, t-statistics to test mean differences were calculated.

Results of these t-tests showed significant differences between officers and non-officers on eight variables. Officers were younger (25 years old vs. 28 years old for non-officers), had a higher level of education (roughly, college graduates vs. high school graduates), had been tank commanders for a shorter amount of time (approximately 4 months vs 7.3 months), had been in a crew for less time (0.9 years vs. 2.5 years), had performed fewer Table VIII's (approximately once vs. 2.7 times), and spent fewer hours per day performing tank commander duties (one hour per day vs. 4.2 hours). Furthermore, on the average, officers liked their jobs more and planned to "reenlist" more frequently than non-officers. All other background and congruence variables showed non-significant differences. Thus, the deletion of the officers clearly changed the descriptive statistics for the tank commanders and potentially had an impact on the correlations between background and congruence variables and the criterion measures. These effects will be discussed in appropriate sections of this report; however, all succeeding analyses (including tank commanders) conducted in Phase II used the reduced (non-officer) data set.

In order to examine if any of these revisions in the data set had a major effect, intercorrelations among the background and congruence variables were computed, along with means and standard deviations for each variable within each crew position. These data were then compared with Tables 3A through 3D and 4A through 4D in the previous Technical Report. With respect to the descriptive data, the results were clear: with the exceptions noted above (i.e., Level of Education, Months Served, and the various Tank Commander variables), and a drop in the Duty Congruence score from 80% to 67% for drivers, the editing of the data base did not substantially change either the means or variances of any variable. Therefore, rather than present four lengthy tables of descriptive information, Tables 3A through 3D of the previous Technical Report can be assumed to accurately reflect the characteristics of the troops in this study.

The "new" intercorrelation matrixes were carefully compared to the previous matrixes. As could be expected, minor differences appeared. Formal statistical tests for comparison of correlational patterns, while available, were judged to be inappropriate for the small sample sizes and non-normal distributions of most of the variables. Again, rather than reproduce Tables 4A through 4D of the previous Technical Report with minor revisions, it can be assumed that these matrixes adequately express the pattern of relationships among the background and congruence variables.

Criterion (Performance) Data

Another potential cause of the unsystematic relationships found during Phase I (and discussed in the previous Technical Report) was the nature of the obtained performance data. For example, most of the Readiness Test items had highly skewed distributions; many had little or no predictable variance; and different scoring rules resulted in different sets of significant predictor variables. As mentioned in the review above, it was decided to conduct extensive analyses on a few selected tasks rather than on the entire set of 100 potential criterion tasks. The tasks selected were chosen to maximize predictability: for the most part, they were the tasks with the most variability, least skewness, and where the different scoring rules did not change their relative difficulty. However, this approach had one major drawback: it sacrificed a large proportion of potentially revealing information contained in the performance data associated with the remainder of the tasks.

Another approach to improving the quality of the performance data which does not involve ignoring substantial portions of the data base was adopted for Phase II. This approach consisted of pooling tasks; that is, to combine several similar individual tasks into larger homogeneous groups. Typically, pooling of variables ameliorates distributional abnormalities, decreases skewness, and increases variance compared to individual items. Thus, if a rational basis could be found for grouping all of the tasks into smaller sets, the possibility for clarifying the nature of the relationship between congruence and combat readiness would be substantially improved.

There were many possible bases for establishment of composite groups of criterion tasks. One such rationale is explored in a later section, where a task taxonomy based on considerations of learning, retention, and practice requirements is developed. Another rationale is for tasks to be grouped on the bases of similarity of equipment involved and similarity of functional requirements of tasks. For example, an obvious grouping of driver tasks is to combine all tasks involved in installing and operating the periscopes (i.e., all tasks involving the same piece of equipment); another obvious group of driver tasks is the combination of tactical driving tasks in non-engagement situations (e.g., crossing a vertical obstacle, crossing a ditch, ascending and descending steep grades, etc.). This group is a "functional" group, all tasks involving actual driving. This basis for grouping tasks has one further practical advantage: during Phase I data collection, troops were requested to estimate their confidence in their ability to perform certain tasks. The tasks that they rated were, in fact, composites formed precisely according to the above rationale. Thus, by adopting this basis for task grouping, it would be possible to straightforwardly explore the relationship between a soldier's confidence and his actual performance.

The reconstituted task groups and the correspondence between these groups and the tasks in the Readiness Test are shown in Table 1 for each crew position. As can be seen, 20 of the 23 Driver tasks fall nicely into five groups. These groups include servicing and operating the periscopes, starting up the tank, preparing to fire, driving in non-combat situations, and tactical combat driving. Three Readiness Test items which did not conveniently fit any of these groups were excluded for most of the remaining analyses, principally because drivers did not provide confidence estimates for these tasks. For the remaining positions, 20 of the 27 Loader tasks were collapsed into 11 groups; and 21 of the 26 Tank Commander tasks were collapsed into 10 groups. Thus, 78 of the original 99 Readiness Test items were included in the reconstituted groups.

Given this restructuring of the criterion performance data set, several new scores had to be created for each soldier. Thus, not only were performance scores for each task group computed, but also new recency and frequency averages were calculated. Table 2 shows the descriptive data of Readiness Test performance for each task group. It was decided that the "proportion of steps passed" would be used as the

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups

DRIVER TASKS

Task Group	Readiness Test Items
1. Remove M27 and install M24 periscopes, perform before operations checks and services on the periscopes, and place M24 periscope into operation.	1. Remove M27 periscope. 2. Perform before operation checks and services on the M24 and M27 periscopes. 3. Install M24 periscope. 4. Place M24 periscope into operation.
2. Perform operations previous to starting tank engine, start tank engine and place tank in motion.	6. Start tank engine. 7. Perform before operation checks and services on tank engine and transmission oil levels. 8. Place tank in motion.
3. Perform driver's procedures preparing to fire main gun.	10. Perform prepare-to-fire procedures.
4. Driving a tank in tactical situations, including driving in neutral steer, crossing obstacles and ditches, driving up and down steep hills, and driving through water obstacles.	11. Operate tank in neutral steer. 12. Cross a vertical obstacle. 13. Cross a ditch. 14. Ascend a steep grade. 15. Descend a steep grade. 16. Drive over a water obstacle.
5. Driving in tactical situations, including performing evasive maneuvers, driving into defilade position, and driving in response to fire commands during target engagement.	17. Perform evasive maneuver upon "enemy" contact by following tank commander commands. 18. Perform evasive maneuver upon "enemy" contact by beginning evasive maneuver on own initiative. 19. Drive into defilade position upon "enemy" contact. 20. Drive in response to fire command during target engagement.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

DRIVER TASKS (cont'd.)

Task Group	Readiness Test Items
5. (cont'd.)	21. Drive to a halt in response to fire command. 22. Lock brakes during a stationary engagement.

Readiness Test Items not included in Task Groups.

5. Perform before operation checks and services on gas particulate unit.
9. Check track tension.
23. Acquire targets.

LOADER TASKS

1. Check and adjust track tension.	2. Check track tension. 3. Adjust track tension.
2. Boresight an M219 machinegun mounted on a tank.	7. Boresight M219 machinegun.
3. Stow main gun rounds in tank.	8. Stow main gun rounds in tank.
4. Perform prepare-to-fire procedures from loader's station.	9. Perform prepare-to-fire procedures.
5. Load main gun in response to fire command; change type of loaded round.	10. Load main gun in response to fire command. a. Battlesight with SABOT loaded. b. Main gun not loaded. c. Main gun loaded with SABOT but some other ammunition is called for.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

LOADER TASKS (cont'd.)

Task Group	Readiness Test Items
6. In response to main gun misfire, perform "rotate round" procedure and unload procedure.	12. Rotate round in misfire procedure. 13. Unload misfired main gun round.
7. Apply action to reduce stoppage of M219 machinegun.	14. Apply immediate action to reduce stoppage in M219 machinegun.
8. Maintain the M219 machinegun including unloading, removal from tank, disassembly, inspection, assembly, checking operation, and mounting on a tank.	15. Unload M219 machinegun. 16. Remove M219 machinegun from tank. 17. Disassemble M219 machinegun. 18. Inspect M219 machinegun. 19. Assemble M219 machinegun. 20. Check operation of M219 machinegun. 21. Mount M219 machinegun on tank.
9. Maintain the main gun breechblock, including removal, disassembly, assembly, and installation.	23. Remove main gun breechblock. 24. Disassemble main gun breechblock. 25. Assemble main gun breechblock. 26. Install main gun breechblock.

Readiness Test Items not included in Task Groups.

1. Perform before operation checks and services on tank engine and transmission oil levels.
4. Check operation of M3 heater.
5. Prepare tank for boresighting.
6. Check boresight alignment of main gun.
11. Ready M219 machinegun in response to fire command.
22. Load M219 machinegun.
27. Read replenisher tape.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

GUNNER TASKS

Task Group	Readiness Test Items
1. Perform gunner's procedures preparing to fire main gun.	4. Perform prepare-to-fire procedures.
2. Prepare gunner's telescope for operation.	6. Prepare telescope for operation.
3. Prepare gunner's periscope for operation.	7. Prepare periscope for daylight operation.
4. Prepare azimuth indicator for operation and operate elevation quadrant.	8. Prepare azimuth indicator for operation. 9. Operate elevation quadrant.
5. Boresight gunner's telescope.	10. Boresight telescope.
6. Boresight daylight sight and IR sight of gunner's periscope during daylight.	11. Boresight daylight periscope. 12. Boresight IR periscope during daylight.
7. Boresight the M219 machinegun.	14. Boresight the M219 machinegun.
8. Zero the main gun.	15. Zero the main gun.
9. Zero the M219 machinegun.	16. Zero the M219 machinegun.
10. Respond correctly to tank commander's fire commands, including actions in main gun engagements and coax engagements.	17. Main gun engagement--gunner fires from a stationary tank at a stationary target. 18. Main gun engagement--gunner fires from a tank moving to a halt at a moving target. 19. Coaxial machinegun engagement--gunner fires from a moving tank at a stationary target. 20. Main gun engagement--gunner fires from a tank moving to a halt at a stationary target.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

GUNNER TASKS (cont'd.)

Task Group	Readiness Test Items
11. Apply immediate action in case of main gun failure to fire (misfire) and unload misfired main gun.	21. Apply immediate action in the case of a misfire. 22. Unload misfired main gun.

Readiness Test Items not included in Task Groups.

1. Check operation of M3 heater.
2. Charge manual elevation system.
3. Place turret into power operation.
5. Prepare tank for boresighting.
13. Boresight the searchlight using the alternate method.
23. Acquire targets.

TANK COMMANDER TASKS

Task Group	Readiness Test Items
1. Perform main gun prepare-to-fire procedures from the TC position.	2. Perform prepare-to-fire procedures.
2. On the M85 machinegun, dismount from tank, disassemble, assemble, and mount on tank.	4. Dismount M85 machinegun. 5. Disassemble M85 machinegun. 7. Maintain, clean, and inspect M85 in preparation for firing. 8. Assemble M85 machinegun.
3. Load and clear M85 machinegun.	3. Load and clear M85 machinegun.
4. Maintain, clean, and inspect M85 in preparation for firing and after firing.	6. Maintain, clean, and inspect M85 after firing.
5. Perform TC duties to prepare rangefinder for operation.	11. Prepare rangefinder for operation.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

TANK COMMANDER TASKS (cont'd.)

Task Group	Readiness Test Items
6. Determine range to target with rangefinder.	12. Determine range to target with rangefinder.
7. Boresight rangefinder.	13. Boresight rangefinder.
8. Boresight M85 Machinegun.	15. Boresight M85 machinegun.
9. Zero the main gun, the M219 machinegun, and the M85 machinegun.	16. Zero main gun. 17. Zero M219 machinegun. 18. Zero M85 machinegun.
10. Announce fire commands for different types of engagements including main gun, M219 machinegun, and M85 machinegun.	20. Main gun engagement--tank commander fires while moving to a halt at a stationary target. 21. Main gun engagement--gunner fires while tank is stationary at a stationary target. 22. Main gun engagement--gunner fires while tank is moving to a halt at a moving target. 23. Coaxial machinegun engagement--tank commander fires from a moving tank at a moving target. 24. Coaxial machinegun engagement--gunner fires from a moving tank at a stationary target. 25. Caliber .50 engagement--tank commander fires while moving to a halt at a moving target. 26. Main gun engagement--gunner fires while moving to a halt at a stationary target.

Table 1. Correspondence Between Readiness Tests and Composite Criterion Task Groups (cont'd.)

<u>Readiness Test Items not included in Task Groups.</u>

1. Check operation of M3 heater.
9. Mount M85 machinegun.
10. Prepare tank for boresighting.
14. Boresight searchlight using the alternate method.
19. Acquire targets.

Table 2. Descriptive Data: Readiness Test Performance[1]

Performance Test[2]	DRIVER Test n ≈ 42		DRIVER Retest n ≈ 13		LOADER Test n ≈ 35		LOADER Retest n ≈ 6	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	.69	.35	.60	.40	.80[3]	.37	—	—
2	.83	.20	.66	.30	.64	.44	.82	.29
3	.84	.27	.62	.33	.82	.31	.98	.06
4	.89	.12	.61	.19	.65	.39	.65	.34
5	.93	.14	.53	.32	.80	.35	.93	.03
6	—		—		.69	.40	.93	.03
7	—		—		.63	.40	.63	.49
8	—		—		.79	.34	.97	.04
9	—		—		.79	.39	.95	.06
10	—		—		—		—	
11	—		—		—		—	

Performance Test	GUNNER Test n ≈ 49		GUNNER Retest n ≈ 19		TANK COMMANDER Test n ≈ 44		TANK COMMANDER Retest n ≈ 23	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	.86	.22	.41	.34	.87	.21	.74	.30
2	.74	.33	.33	.31	.95	.16	.89	.22
3	.90	.17	.45	.33	.98	.06	.92	.21
4	.76	.29	.29	.25	.94	.22	.91	.29
5	.83	.29	.57	.32	.92	.21	.89	.23
6	.79	.31	.40	.30	.99	.04	1.0	.00
7	.94	.25	.55	.44	.93	.23	.83	.27
8	.90	.21	.66	.29	.95	.17	.93	.11
9	.87	.22	.51	.37	.96	.13	.82	.31
10	.95	.07	.88	.13	.97	.05	.87	.25
11	.89	.18	.93	.18	—		—	

1. Scores are proportion of steps correct for each task group.
2. Group numbers correspond to the groups listed in Table 1.
3. N for this test = 19.

performance index rather than the "task GO/NO GO" measure, for the basic reason that the latter measure would give differential weights to individual items. For example, the adoption of a GO/NO GO rule for the task groups would require that an entire group of, say, six items would be scored as NO GO if only one item were failed. The same group NO GO score would be assigned to a group consisting of only one item. Furthermore, the proportion of steps passed measure, since it has a higher number of potential values, is more likely to demonstrate substantial variability than any GO/NO GO rule. Therefore, for each composite task group, a proportion-of-steps-passed score was computed by dividing the number of steps passed in the group by the total number of steps taken in the group.

As mentioned above, given these new task groupings, it was necessary to recalculate the recency and frequency variables. [It should be noted that the recency variable was measured along a six-point scale ranging from no practice (0) to most recently practiced within the last two months (5), and the frequency variable was measured along a four-point scale ranging from no practice in the past year (0) to 10 or more times in the past year (3).] Table 3 shows the mean recency and frequency of practice for each of the task groups. These numbers were computed by summing the ratings for all tasks in a given group and dividing by the number of tasks in each. Thus, the values presented in the table are comparable to the data presented in the previous Technical Report.

Additional Criterion (Performance) Data

As remarked previously, the Readiness Test employed in this study is a relatively new instrument and has not been studied or evaluated in detail. The unsystematic relationships found in Phase I might have been partially due to the novelty of the test from the perspective of the test administrators, inconsistent interpretations of the scoring rules, inadequacy of instructions, or a host of other factors influencing the test administration. Therefore, in an effort to clarify some of these potentially biasing influences, the decision was made to incorporate additional performance data into the data set. Two types of information were made available: the results from a readministration of the Readiness Test to a subset of the troops, and the Skill Qualification Test (SQT) scores from many of the soldiers involved in the Phase I data collection. Given these additional data, several questions could be addressed. For example, the relationship between the first and second administrations of the Readiness Test could be examined in order to possibly detect any systematic variations in administration and to provide an initial estimate of the

Table 3. Descriptive Data: Recency and Frequency of Performance for Composite Task Groups

Task Group*	Driver n = 42		Loader n = 42		Gunner n = 51		T. Commander n = 44	
Recency								
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	2.79	2.00	3.14	0.83	4.06	1.66	4.29	1.45
2	4.45	1.11	1.30	1.79	3.98	1.79	4.00	1.33
3	2.98	2.27	2.29	2.24	3.98	1.81	3.07	2.05
4	3.62	1.61	2.44	2.25	3.69	1.76	3.44	1.96
5	2.42	1.72	1.43	1.75	3.76	1.96	4.24	1.71
6	—		1.41	1.90	2.71	1.68	4.64	1.16
7	—		1.33	1.86	2.76	2.07	3.83	2.08
8	—		2.95	1.93	3.63	2.00	2.72	2.19
9	—		3.37	2.12	2.49	2.18	3.01	1.78
10	—		—		1.84	1.50	1.63	1.68
11	—		—		1.86	1.75	—	
Frequency								
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	0.87	0.72	1.68	1.25	1.76	1.01	1.50	.90
2	2.61	0.77	.79	.98	1.49	.99	0.37	0.83
3	1.28	1.16	1.31	1.26	1.55	1.05	1.41	1.00
4	1.81	0.98	1.31	1.14	1.40	0.88	1.52	.95
5	1.20	0.97	0.88	1.10	1.35	1.02	1.55	.95
6	—		0.63	0.90	0.92	0.78	1.95	.99
7	—		.98	1.20	1.16	.95	1.37	1.05
8	—		1.50	1.16	1.34	.96	1.11	.95
9	—		1.39	1.07	1.08	1.02	1.02	0.74
10	—		—		1.05	0.98	1.07	1.07
11	—		—		0.79	0.76	—	

*Task group numbers correspond to the groups listed in Table 1.

reliability of the test. Furthermore, an examination of the interrelationships among the three criterion measures (Readiness Test 1, readministration, and SQT) would supply information regarding the validity of the performance tests. Finally, the relationships among the various criterion measures and the various predictor variables could be examined to further clarify the nature of the effect of congruence on performance effectiveness. These topics will be discussed in the following sections.

Readiness Retest. A total of 130 troops were administered the Readiness Test on the second runthrough. Of these 130, 101 troops were contained in our data file. Unfortunately, only 61 of these troops took the retest in the same crew position as in the first administration. These 61 troops consisted of 13 drivers, six loaders, 19 gunners, and 23 tank commanders. For these troops, descriptive statistics were obtained, again using a "proportion of steps passed" scoring rule. These data are shown in Table 2, along with the comparable first administration. In addition, where the number of troops was sufficient, correlations and paired t-tests between corresponding task groups were computed.

Descriptively, Table 2 shows that the retest scores were lower than the first administration and, in general, had higher variabilities. Loaders are the exception to this rule; however, since there were so few loaders, it is clearly inappropriate to consider this group as representative of loader performance in general. Thus, the loader group retest data were excluded from further analyses. In fact, of the 26 remaining tasks (excluding the loaders), the retest means for 15 were significantly lower (two-tailed t-test; $p < .05$) and nine others were lower (but not significantly so). One of the two remaining tasks (Tank Commander Task Group 6) was virtually identical for the two administrations, and the final task (Gunner Task Group 11) was nonsignificantly higher on the retest. Similarly, 20 of the 25 tasks (Tank Commander Task Group 6 had no variance) showed increases in variance on the retest.

Correlations between the first and second administrations were uniformly low for all task groups. In fact, only one task (Tank Commander Task Group 7) showed a significant ($p < .05$) test-retest correlation. In an attempt to explicate this unexpected finding, bivariate scatterplots of the two administrations were created for each task group. These plots revealed that the low correlations were primarily caused by a few extreme values: the general pattern was for a few troops who had performed very well on the first administration to score extremely low on the retest. Statistically, it is the case that outliers (points far from the mean) affect variance and regression weights (i.e., correlations) more than those near the mean. Combined with the small sample size, those few extreme values acted to substantially lower the obtained correlation coefficients. Discussion of the implications of these results for the possible biases in administration and Readiness Test reliability will be postponed until further data (the within-test correlations) are presented below.

SQT Performance. Selected SQT data were available for 137 troops involved in this study: 23 drivers, 24 loaders, 51 gunners, and 39 tank commanders. There were three SQT skill level tests administered (corresponding to Skill Levels 2, 3, and 4). Given the fact that for purposes of this study some troops were "out of position", the mapping of crew positions onto SQT level was not isomorphic. Thus, all 23 drivers, took SQT 2; 20 loaders took SQT 2, 2 took SQT 3 and 2 took SQT 4; 45 gunners took SQT 2 and 6 took SQT 3; 12 tank commanders took SQT 2, 13 took SQT 3, and 14 took SQT 4. Three scores were available for SQT 2 and 3: the number of NO GO tasks for the hands-on component (SQT-T), the number of NO GO tasks for the written component (SQT-W), and, for some of the troops, the number of items (within tasks) NO GO for the hands-on components (SQT-I). Since the SQT 4 test is entirely written, there are no SQT-4T or SQT-4I scores.

Descriptive data for these test scores are shown in Table 4. Although possibly not germane to this project, a number of aspects of this table warrant attention. First, practically all troops performed remarkably well on the hands-on components of whichever SQT they took. This is probably due to the extensive preparation typically undergone for these career-relevant tests. On the other hand, performance on the written components of all SQT's was surprisingly low, with most troops failing almost half the tasks. Apparently, troops cannot prepare as extensively for the written items as they can for the hands-on items, even though the SQT Notices describe the written tasks in some detail. The issue of preparation for performance testing should be kept in mind when viewing the Readiness Test data; it was assumed that (at least for the first administration) troops were tested "cold": that is, without the benefit of specific preparation for the test.

Discussion of the interrelationships among the criterion performance data will be postponed until the remaining data set, the confidence estimates, is presented.

Confidence Estimates. The final information which was incorporated into the data set was the aforementioned confidence estimates. Each soldier was asked to rate his ability [on a scale of one ("can't perform at all") to four ("can perform very well")] to perform each of the criterion task groups. These ratings are presented descriptively in Table 5 for each position. As can be seen, confidence estimates tended to be lower for the two junior positions (driver and loader);

Table 4. Descriptive Data: SQT Performance

	Driver			Loader			Gunner			T. Commander		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD
SQT 2I[1]	15	2.67	3.22	14	2.29	2.87	30	2.27	3.62	8	1.13	1.36
SQT 2T[2]	23	0.96	1.26	20	0.90	0.97	45	1.02	1.22	12	0.67	0.78
SQT 2W[3]	23	13.57	4.25	20	14.50	3.85	45	12.20	5.14	12	11.92	4.01
SQT 3I	--			2	1.00	1.41	2	6.00	4.24	6	4.50	7.20
SQT 3T	--			2	1.00	1.41	6	1.67	1.21	13	0.77	0.83
SQT 3W	--			2	25.00	5.66	6	25.67	4.55	13	19.69	7.60
SQT 4W[4]	--			2	17.00	7.07	--			14	26.79	6.34

1. Number of NO GO steps for the hands-on component.
2. Number of NO GO tasks (maximum = 10) for the hands-on component.
3. Number of NO GO tasks for the written component; SQT 2 has 30 tasks, SQT 3 has 45 tasks.
4. The SQT 4 does not have a hands-on component.

Table 5. Descriptive Data: Confidence Estimates

Task Group*	Driver n = 43		Loader n = 40		Gunner n = 50		T. Commander n = 44	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	2.26	1.11	2.95	1.11	3.04	.87	2.84	.78
2	3.33	.89	2.52	1.06	3.02	.74	3.27	.66
3	2.51	1.12	3.07	1.21	3.02	.75	3.25	.84
4	2.88	1.16	2.77	1.21	3.02	.89	3.20	.70
5	2.49	1.08	2.82	1.20	3.24	.80	2.86	.85
6	—		2.95	1.15	2.55	1.08	3.32	.83
7	—		2.62	1.13	2.88	.44	2.93	.95
8	—		2.72	1.20	3.40	.78	2.91	.86
9	—		3.05	1.01	2.88	.96	3.14	.82
10	—		—		3.24	.94	3.23	.68
11	—		—		3.46	.71	—	

*Task group numbers correspond to the groups listed in Table 1.

these two positions also tended to have more variability in their estimates than did the two senior positions.

It will be recalled that these confidence estimates could potentially be valuable as surrogate criterion variables; that is, if it could be argued or demonstrated that these estimates are related to actual performance, they could be used as "purer" (uncontaminated by specific test factors) criterion variables. In order to assess the value of the confidence ratings, two pieces of information will be presented. First, the within-criterion intercorrelations, and second, the zero-order correlations between the confidence estimates and performance test scores will be shown.

Criterion Task Intercorrelations. The purpose for this analysis is to determine two features of the criterion task set. First, is the internal consistency of each criterion set. This feature could presumably reveal any "bad" items (analogous to item analyses in standard test construction) or, if the pattern of internal correlations differed for two presumably equivalent test forms, any systematic biases in test administration. The second feature is the between-measures relationships (e.g., the relationship between Readiness Test scores and SQT scores).

Data pertaining to the internal consistency issue are presented in Table 6. This table shows the within-test intercorrelations for the task groups for each crew position. The most striking feature of this table is that for confidence estimates, with minor exceptions, all of these intercorrelations are highly significant ($p < .01$) for all crew positions. It would appear that these troops do not discriminate among the task groups when they rate their own ability; if a soldier judged that he was good or poor for one task, he judged that he was good or poor for all the others. This could be a function of the "completeness" of training for these positions and the perhaps restricted range of tasks sampled. That is, if a soldier received training or practice in (for example) driving, the tasks presented for ratings constitute a minimum, necessary, set. Similarly, if a loader knows "how to load", the nine task groups presented are those uniformly included in training. Another less likely explanation is that by collapsing tasks into groups, we have unknowingly created groups that are inconsistent with the troops' view of their job or how their job training and practice is organized. For example, the organization of gunner task groups (see Table 1) is primarily an "equipment" organization, with individual groups for operation of the periscope, machinegun, and telescope. If, however, the "true" organization of gunner's training or practice is functional (for example, a training unit or work detail is "general maintenance" where the gunner

Table 6. Intercorrelations[1] Within Criterion Tasks and Confidence Ratings

A. DRIVERS

		Task Group[2] 1	2	3	4	5
Confidence Estimates (N = 43)	1	—	39	56	54	53
	2		—	66	54	60
	3			—	70	71
	4				—	83
	5					—
Readiness Test (N = 40)	1	—	57	45	*24*	42
	2		—	75	*13*	41
	3			—	*15*	55
	4				—	42
	5					—
Retest (N = 13)	1	—	*62*	*56*	*48*	*29*
	2		—	*50*	*12*	*48*
	3			—	*21*	*16*
	4				—	*49*
	5					—

B. LOADERS

		Task Group 1	2	3	4	5	6	7	8	9
Confidence Estimates (N = 40)	1	—	63	71	72	79	80	62	74	44
	2		—	75	67	72	75	68	80	64
	3			—	87	84	91	83	87	63
	4				—	93	87	88	86	55
	5					—	92	86	93	60
	6						—	81	90	64
	7							—	87	63
	8								—	65
	9									—
Readiness Test [N = 17 (Group 1); N ≈ 30 for remainder]	1	—	*50*	53	*36*	64	58	59	*14*	*02*
	2		—	54	76	*31*	55	61	*34*	43
	3			—	57	54	*38*	*44*	*11*	*12*
	4				—	*40*	54	55	49	45
	5					—	72	68	68	54
	6						—	91	65	84
	7							—	56	80
	8								—	68
	9									—

1. Pearson product-moment r's rounded to nearest hundredth; decimals omitted. Correlations with significance levels $\leq .01$ are indicated by boldface; correlations in italics are nonsignificant.
2. Task group numbers correspond to the groups listed in Table 1.

Table 6. Intercorrelations Within Criterion Tasks and Confidence Ratings (cont'd.)

C. GUNNERS

Task Group

		1	2	3	4	5	6	7	8	9	10	11
Confidence Estimates (N = 50)	1	--	74	74	51	61	51	54	65	55	47	48
	2		--	81	55	68	58	56	58	55	49	49
	3			--	52	72	50	59	66	58	47	41
	4				--	62	46	49	63	48	36	37
	5					--	63	69	82	70	55	56
	6						--	37	54	36	37	27
	7							--	62	91	57	42
	8								--	63	64	66
	9									--	51	44
	10										--	60
	11											--
Readiness Test (N = 47)	1	--	44	38	41	45	39	59	57	32	-08	29
	2		--	53	65	43	55	32	39	21	-04	03
	3			--	62	65	72	50	72	53	09	22
	4				--	55	57	47	56	33	24	38
	5					--	54	42	59	56	04	27
	6						--	34	50	55	02	34
	7							--	85	76	11	01
	8								--	61	15	17
	9									--	-03	17
	10										--	24
	11											--
Retest (N = 19)	1	--	59	44	24	61	46	40	61	58	45	12
	2		--	73	36	58	63	42	63	65	46	23
	3			--	66	65	78	60	71	57	50	30
	4				--	58	39	35	41	28	22	-08
	5					--	61	52	83	71	45	29
	6						--	53	69	45	22	32
	7							--	63	58	27	25
	8								--	80	51	12
	9									--	60	32
	10										--	42
	11											--

D. TANK COMMANDERS

Task Group

		1	2	3	4	5	6	7	8	9	10
Confidence Estimates (N = 44)	1	--	36	24	10	49	30	46	33	58	25
	2		--	42	48	15	22	18	33	36	33
	3			--	70	50	59	61	61	49	55
	4				--	40	41	44	61	35	34
	5					--	72	85	62	59	46
	6						--	74	43	65	53
	7							--	59	64	53
	8								--	48	48
	9									--	48
	10										--
Readiness Test (N = 42)	1	--	-05	42	-01	69	23	59	29	22	21
	2		--	12	88	02	-04	-08	44	-06	13
	3			--	10	60	18	41	01	-08	16
	4				--	13	-04	-05	71	02	29
	5					--	10	70	30	13	44
	6						--	14	-05	10	24
	7							--	03	14	11
	8								--	06	37
	9									--	11
	10										--
Retest (N = 23)	1	--	62	22	42	39	--	11	26	42	43
	2		--	-08	89	03	--	-16	-04	38	10
	3			--	-05	79	--	42	26	46	15
	4				--	18	--	-11	-13	33	19
	5					--	--	53	35	38	52
	6						--	--	--	--	--
	7							--	32	26	33
	8								--	44	32
	9									--	22
	10										--

services all of these pieces of equipment), gunners would necessarily give the same confidence rating across our groups. [The one task group that does not have uniformly high correlations is Tank Commander Group 2, which is assembly, disassembly, and mounting of the M85 machinegun. It is difficult to argue from this single exception, but it seems that this case indicates that tank commanders were able to discriminate this task from the others; hence, they were not treating these confidence estimates as a single, uniform measure of their ability.]

The intercorrelations among the Readiness Test tasks are likewise predominantly significant ($p < .01$), with tank commanders as an exception. This exception is most probably due to a ceiling effect, as there was insufficient variance in tank commander performance to allow for significant correlations (see Table 2). The tank commander correlations that are significant do, however, follow a somewhat logical pattern: task groups 2, 4, and 8 (which significantly intercorrelate) are all concerned with servicing the M85, while task groups 1, 3, 5, and 7 (which likewise intercorrelate significantly) can all be interpreted as combat preparation tasks.

For the other crew positions, the uniformly high intercorrelations support the same sorts of speculations made concerning the confidence ratings: the Readiness Test does not seem to discriminate among the task groups. [Most of the exceptions (e.g., Gunner Task 10) are probably due to lack of variance in the performance data.] This was essentially what was previously found when a similar analysis was performed during Phase I (see Table 7 of the previous Technical Report). In one sense, these uniformly high intercorrelations are disappointing; it was hoped that the restructuring of the test items would lead to measures of different aspects of performance. These hoped-for differences could then presumably have different correlates in the predictor variable set. Similarly, the uniformly high correlations do not provide useful information regarding specific items or biases in test administration. Based on these data, it is still possible that either the Readiness Test is a completely homogeneous sampling of tank crew skills, or that test administrators formed a generalized opinion about each soldier (e.g., deciding that the soldier was a good gunner), and scored the entire test based on this generalized opinion, or that existing tank training precisely fits performance requirements (e.g., more training and practice are provided for more difficult tasks).

The intercorrelations for the Readiness Retest add some confusion to the above speculations. For example, the intercorrelations for drivers are all nonsignificant. This, of course, might be due to the small N's in these computations (since the absolute values are relatively high); or, if viewed in isolation, this result could be interpreted as independent tests of driver's skills. This latter interpretation is clearly inconsistent with previous results, and thus leads to the suspicion that the retest was somehow different from the first administration (at least for drivers). Given that performance on the retest was lower than on the first administration (see Table 2), it is possible that the retest was a more accurate reflection of driver performance. However, there is simply no confirmatory evidence for this hypothesis, especially given the previous discussion of test-retest relationships.

Within-test correlations for the SQT tests suffer from the problem of small N's. Nevertheless, some of these correlations were significant ($p < .01$): for drivers, SQT 2T was related to SQT 2I ($r = .88$, $N = 15$), and SQT 2W ($r = .48$, $N = 15$); for loaders, SQT 2I was related to SQT 2T ($r = .90$, $N = 20$); for gunners, SQT 2I was related to SQT 2T ($r = .86$, $N = 30$) and SQT 2W ($r = .43$, $N = 30$); and for tank commanders, SQT 2I was related to SQT 2T ($r = .93$, $N = 8$) and SQT 3T was related to SQT 3W ($r = .63$, $N = 6$). Thus, in all cases where a computation was possible, the two hands-on scores were related to each other, while the relationships between the hands-on components and the written components were less obvious.

The final issue to be discussed is the interrelationships among the criterion measures. Some of these have been discussed previously, namely the Readiness Test-Retest relationship. The other comparisons--SQT vs. Readiness Test, Confidence Ratings vs. Readiness Test, Retest, and SQT, will be discussed below. [SQT vs. Retest correlations were calculated just for gunners due to the small number of troops in the other positions for which both scores were available.]

SQT vs. Readiness Test. Although this is a critical comparison in terms of assessing the value of the Readiness Test as a valid measure of combat effectiveness, such a comparison must be viewed with extreme caution. Troops, in general, have extensive training for the SQT. They are issued SQT notices which describe exactly what they will be tested on, training for the test is typically an organized activity at the Company level, and SQT scores have a significant impact upon a soldier's career progression. On the other hand, the Readiness Test was administered with little or no preparation on the part of the troops--they did not know what they would be tested on prior to administration. Furthermore, the troops knew that their performance would not have any impact on their careers. Given these considerations, it is not surprising that SQT performance uniformly did not correlate with test performance. The only exceptions were three significant correlations between SQT 2W and the Retest task groups 8, 9, and 10 for gunners (where $N = 18$). In one sense, this lack of a relationship could be viewed as a severe criticism of the tank training system: in essence, the argument would be that in order to meet the established standards of performance (i.e., the SQT), troops must be given extensive preparation; there is no correspondence between their performance if sampled without such preparation and "true" (SQT) measures of combat effectiveness. However, such a sweeping generalization is clearly inappropriate: first, as we have seen, Readiness Test performance is (on an absolute scale), quite good, especially when one considers the lack of preparation. Second, there is no a priori relationship between the content of the two tests. [Such a mapping was attempted, based on both the task communality of the two tests and on the basis of a step-by-step comparison of tasks that were in common. Especially in the latter comparison, the lack of correspondence led to the conclusion that the tests were clearly different in content.] And finally, there is no evidence that minimum acceptable preparation (as would be the case in a combat environment) for the Readiness Test would not have produced more significant relationships.

Confidence vs. Performance Tests. Correlations bearing upon these comparisons are shown in Table 7 as a function of crew position. Considering SQT performance first, the only significant relationships appear for the SQT 2W scores for gunners. [The negative signs are in the "right" direction, since the SQT scores are number of items failed.] The same arguments made above concerning the SQT-Readiness Test relationships apply here as well; the most parsimonious explanation is that the two sets of data measure different things. It would have been interesting if soldiers had been asked to rate their confidence in their ability to perform the individual SQT items.

In general, there are significant relationships between confidence estimates and Readiness Test performance for loaders. The uniform nature of these significant correlations, rather than a strictly diagonal form, reinforces the notion that both the confidence ratings and performance tests do not discriminate among the reconstituted task groups. There is a suggestion of a more direct correspondence for drivers: at least for two task groups (numbers 1 and 5), significant correlations were obtained between confidence estimates and Readiness Test performance. However, the general lack of significant relationships for the other positions is somewhat disappointing. It would be hard to argue that soldiers accurately estimated their performance capabilities. However, it should be noted that the soldiers were not provided with detailed descriptions of the tasks; thus, it might be that the actual Readiness Test items were different from the way soldiers conceptualized them.

Table 7. Zero-Order Validity[1] Between Confidence Estimates and Performance Test Scores and SQT Scores

A. Drivers

		Readiness Test Performance[3]					Retest Performance[4]					SQT Scores[5]		
		1	2	3	4	5	1	2	3	4	5	2I	2T	2W
Confidence	1	**58**	18	16	12	27	21	-19	-05	-28	-62	-17	-05	-09
Estimates	2	25	20	12	33	30	35	-06	44	23	-27	07	-02	-22
For Task	3	**44**	29	22	32	33	23	-24	15	07	-42	-31	-26	-38
Groups[2]	4	**53**	32	24	15	17	36	-03	09	-14	-36	07	-03	01
	5	**59**	36	24	29	**38**	40	-09	18	-05	-23	07	04	-02

1. Pearson product-moment correlations, rounded to nearest hundredth; decimals omitted. Correlations with significance levels ≤ .01 are indicated by boldface.
2. Task group numbers correspond to the groups listed in Table 1.
3. N ≈ 40. In all cases, Readiness Test and Retest performance scores are proportions of steps passed in each task group.
4. N ≈ 13.
5. N = 15 for 2I, N = 23 for 2T and 2W. SQT scores are the number of items NO GO (see text).

B. Loaders[6]

		Readiness Test Performance[7]									SQT Scores[8]		
		1	2	3	4	5	6	7	8	9	2I	2T	2W
	1	44	36	**61**	**48**	**68**	**76**	**68**	**40**	**50**	33	09	04
Confidence	2	07	35	21	34	31	**52**	**45**	**43**	**46**	52	**62**	-13
Estimates	3	50	37	**49**	**43**	41	**70**	**63**	30	**45**	39	32	05
For Task	4	**57**	31	**47**	42	**48**	**72**	**71**	36	**55**	57	36	-07
Groups	5	**43**	34	**43**	**49**	**54**	**78**	**74**	**39**	**58**	48	31	-08
	6	35	32	**48**	**47**	**54**	**76**	**70**	34	**55**	46	41	-01
	7	55	34	42	40	34	**63**	**59**	22	**40**	58	30	-10
	8	39	41	**42**	**48**	**43**	66	**62**	33	**48**	47	36	-16
	9	29	**64**	**43**	**62**	25	**60**	**55**	37	**53**	48	34	-20

6. N's for loader retests were insufficient to compute correlations (see text).
7. N = 17 for task group 1; N ≈ 31 for the remaining task groups.
8. N = 13 for 2I; N = 18 for 2T and 2W.

Table 7. Zero-Order Validity Between Confidence Estimates and Performance Test Scores and SQT Scores (continued)

C. Gunners

		Readiness Test Performance[9]											Retest Performance[10]											SQT Scores[11]		
		1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	21	2T	2W
	1	20	11	14	15	17	14	14	17	-04	-07	14	-43	-28	-28	-22	-16	-06	-10	06	-17	-44	-36	-21	02	-36
	2	30	09	14	13	19	-01	16	18	-05	-05	09	-38	-33	-23	-15	-09	-04	-10	14	-18	-40	-43	-13	16	-31
Confidence	3	37	08	35	18	40	30	34	48	30	-01	14	-47	-51	-47	-33	-40	-40	-11	-18	-26	-31	-42	04	21	-32
Estimates	4	30	24	29	15	17	09	26	31	02	07	06	-43	-02	-11	-17	-41	-03	11	-16	-10	-25	-22	-04	02	-31
For Task	5	39	10	23	10	22	16	22	26	08	-09	06	-38	-26	-27	-36	-33	-05	-21	-03	-10	-26	-25	-29	05	-51
Groups	6	27	20	17	-03	26	13	12	10	-07	-02	-02	20	-10	-36	-29	-11	-27	-20	05	-05	-09	-40	-09	06	-33
	7	26	-05	-01	14	01	-03	20	14	-03	07	11	-46	07	25	11	25	26	18	25	20	-11	27	-18	00	-49
	8	33	-04	19	-01	09	11	27	26	-03	07	05	-34	-35	-41	-44	-31	00	-35	-17	-13	-36	-18	-22	07	-42
	9	12	-10	-08	09	-08	-12	08	00	-06	06	15	-52	00	15	05	13	17	10	23	10	-22	-01	-18	02	-53
	10	25	-16	06	08	08	-07	18	19	03	31	19	-17	-23	-19	-16	00	13	-22	11	-04	-14	-22	-27	04	-25
	11	12	-24	-06	-21	-13	-20	-02	-12	-06	-03	04	-40	-42	-41	-06	-10	-20	-35	-20	-14	-45	-34	-42	-08	-21

9. N = 43.
10. N = 17.
11. N = 26 for 2I, N = 41 for 2T and 2W.

D. Tank Commanders

		Readiness Test Performance[12]										Retest Performance[13]										SQT Scores[14]				
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	2T	2W	3T	3W	4W
	1	04	16	21	16	-01	19	05	-04	07	05	-19	-04	-21	04	-12	—	23	-18	-11	16	35	-10	-60	-69	-58
	2	00	38	16	45	01	06	-19	25	08	19	03	17	-18	13	-05	—	09	-03	-01	10	53	-25	29	07	-10
Confidence	3	18	06	28	04	18	05	10	03	14	31	12	02	53	05	50	—	38	13	20	18	13	27	43	-24	-27
Estimates	4	13	07	18	06	06	04	-06	05	14	42	29	07	73	08	71	—	51	19	36	30	08	11	37	-15	15
For Task	5	04	10	22	09	20	-02	18	-05	15	22	-10	-12	39	00	39	—	43	15	16	15	44	-12	-16	-51	-67
Groups	6	11	23	26	21	18	-13	01	02	14	21	03	-15	59	-08	48	—	34	01	12	00	22	-23	08	-37	-03
	7	09	11	35	11	29	-01	19	-03	16	24	-14	-18	46	-13	36	—	46	08	16	09	17	-05	-14	-60	-52
	8	09	13	18	10	06	-02	-07	04	15	39	-19	-23	47	-20	35	—	36	07	10	13	27	-03	-30	-53	-34
	9	15	20	34	18	14	02	05	01	04	05	-13	-14	10	-13	07	—	05	-11	-11	17	66	-55	08	-69	-48
	10	36	26	55	25	42	06	18	13	11	30	00	00	04	11	23	—	15	-05	-12	35	56	-41	-23	-61	15

12. N = 41.
13. N = 23.
14. N = 12 for all tests. N's for 2I and 3I scores were insufficient to compute correlations.

In summary, these data do not support the contention that confidence estimates could be used as a direct surrogate of Readiness Test performance. However, there is good support (based on their internal consistency and the unsolved questions of the validity and reliability of the Readiness Tests themselves) for the retention of these estimates as a quasi-independent criterion measure. If meaningful predictions could be made of this measure from the predictor variables, the basic relationship between congruence and performance would be clarified.

CONGRUENCE EFFECTS

The relationships between levels of job congruence and effectiveness on combat-related tasks were investigated by means of two different strategies. First, as in the Phase I analysis of the smaller set of criterion tasks, zero-order validities were estimated. These estimates were obtained by computing the correlations between the background, congruence, and practice variables and the various criterion measures of performance effectiveness. At issue was whether those congruence and background variables related to effectiveness measures would generally remain the same across tasks and crew positions.

In the second or "case study" approach, judges assigned subjects to different categories representing varying levels of presumed job congruence. Assignments were made by considering and subjectively weighting a small subset of the variables describing each soldier's background and job congruence. For each criterion task, the categories of soldiers were ranked to represent predicted levels of effectiveness. The strength of the relationships between congruence and effectiveness were explored by analyzing the resulting contingency tables. This approach essentially replaced the "best subsets" multiple regression analyses conducted in Phase I, which were difficult to interpret and were based on a fairly large number of independent variables relative to the rather small sample sizes on which data were available.

Zero-Order Validities

The relationships between the independent and criterion variables were examined separately for each of the four crew positions. In each case the personal background, Army and job background, congruence, and practice variables were correlated with a variety of criterion effectiveness measures. The latter included estimates of effectiveness for each composite criterion task (i.e., proportion of steps passed and confidence ratings) as well as summary performance measures relating to the readiness tests (i.e., proportion of steps passed, and proportion of tasks on which a GO was received) and to Skill Qualification Tests. The results are presented below for each crew position.

Drivers. Significant ($p < .01$) zero order validities were, with but two exceptions, confined to the first composite task dealing with maintenance and services on the periscopes. The two exceptions are a negative relationship between height and tactical driving performance (Task 4, $r = .41$, $p < .01$)

and a positive correlation between months in position during the past year and starting the engine (Task 2, $r = .44$, $p < .01$).

As shown in Table 8, the strongest relationships with performance on Task 1 involve variables primarily from the Army and job background and the congruence categories. Better performance is associated with soldiers who have been in crews for longer periods of time, have served in more crew positions during training, and have participated in more Table VIIIs. The fact that they have participated in training programs and Table VIIIs specifically as drivers is important. Also associated with better performance is frequent practice of this task, as well as Task 2 (starting the engine), longer time in the 11E primary MOS, and the stated intention to reenlist.

These relationships appear generally consistent and logical given that this composite task (of the five tested) is in fact performed least frequently, the one on which soldiers achieve the lowest mean performance, and relative to the other four, is the one possessing the largest variance. The same types of relationships between the independent variables and performance on Task 1 are essentially mirrored in the overall proportion-of-steps-passed measure (HOP) based on all five composite tasks.

The absence of similar relationships for the other driver tasks may be essentially due to reasonably high congruence. These tasks are practiced more frequently, performance is higher (ranging from a mean of .83 to .93), and there is relatively little variance in the test scores (SD ranging from .12 to .27). Roughly two-thirds of the sample are in fact drivers who, during their average ten-and-a-half months in that position, have been able to practice these tasks fairly frequently (more so than the periscope task).

No significant ($p < .01$) zero-order validities are associated with the retest performance data subsequently obtained for a rather small subset ($n=13$) of the original driver sample. No significant relationships are obtained with the overall proportion-of-tasks-GO measure. Finally, with respect to SQT scores, only the soldier's rank correlates significantly with performances ($r=.61$, $p < .01$, for SQT 2I; $r=.48$, $p < .01$ for SQT 2W).

Relationships between the various independent variables and soldiers' estimates of their ability to perform the driver criterion tasks are presented in Table 9. The results differ dramatically from those presented for the readiness test performance measures.

Table 8. Zero-Order Validities* Between Background/Congruence/ Practice Variables and Proportion of Steps Passed by Drivers on the Periscope Composite Task

(Readiness Test Data)

Personal Background	Weight	—
	Height	—
	Age	—
	Education	—
	Like Job	—
	Reenlist	45
Army & Job Background	Months Served	—
	Rank	—
	Months in Primary MOS	37
	Skill Level-Primary MOS	—
	Years in Crew	**46**
	Number Positions Held in Training	**51**
	Number Table VIIIs	**52**
Congruence	Months in Duty MOS	—
	Skill Level-Duty MOS	—
	Test Position-Job Title Match	—
	Position Training	**46**
	Position Table VIIIs	**46**
	Months in Position Past Year	—
	Hours per Day in Position	—
	Platoon Sergeant Ratings	—
Frequency of Practice	Task 1	41
	Task 2	37
	Task 3	—
	Task 4	—
	Task 5	—
Recency of Practice	Task 1	—
	Task 2	—
	Task 3	—
	Task 4	—
	Task 5	—

*Correlations rounded to the nearest hundredth with decimal points omitted. Correlations with significance levels $\leq .001$ are indicated in bold face; others are significant in the range $\leq .01$, $> .001$.

Table 9. Zero-Order Validities* Between Background/Congruence/ Practice Variables and Driver Confidence Estimates for Five Composite Tasks

		TASKS 1	2	3	4	5
Personal Background	Weight	*43*	--	--	--	--
	Height	*36*	--	--	--	--
	Age	--	--	--	--	--
	Education	--	--	--	--	--
	Like Job	--	--	--	--	--
	Reenlist	46	--	--	--	--
Army & Job Background	Months Served	--	--	--	--	--
	Rank	--	--	46	--	--
	Months in Primary MOS	*36*	*43*	*41*	53	*43*
	Skill Level-Primary MOS	--	--	60	56	55
	Years in Crew	*43*	54	61	71	72
	Number Positions Held in Training	--	46	48	55	62
	Number Table VIIIs	--	--	*44*	45	49
Congruence	Months in Duty MOS	46	--	48	51	*41*
	Skill Level-Duty MOS	--	--	47	*36*	*44*
	Test Position-Job Title Match	--	*43*	50	*41*	55
	Position Training	50	45	62	54	62
	Position Table VIIIs	52	--	60	51	48
	Months in Position Past Year	45	49	62	66	64
	Hours Per Day in Position	--	*44*	*36*	*37*	--
	Platoon Sergeant Ratings	--	*45*	*43*	*39*	--
Frequency of Practice	Task 1	66	*45*	50	49	*43*
	Task 2	*43*	60	52	*43*	50
	Task 3	49	*38*	62	*45*	49
	Task 4	--	56	49	52	56
	Task 5	*36*	*40*	58	49	62
Recency of Practice	Task 1	64	48	56	59	49
	Task 2	--	62	53	--	49
	Task 3	57	59	77	50	61
	Task 4	--	71	54	*42*	55
	Task 5	--	--	46	--	*43*

*Correlations rounded to the nearest hundredth with decimal points omitted. Correlations with significance levels ≤ .001 are indicated in bold face; others are significant in the range ≤ .01, > .001.

The caution that must be sounded when inspecting the correlations shown in Table 9 is in assuming that the confidence estimates are in fact valid surrogates of the performance data. As reported earlier, there is some evidence that the two measures are related in the sample of drivers, particularly with respect to the periscope composite task ($r=.58$, $p < .001$). The lack of particularly strong relationships between expressions of confidence and levels of performance on the other four tasks must be tempered by the fact that there is such small variance in the hands-on performance measures.

The more conservative line of interpretation addresses the factors upon which drivers base their estimates. For example, as shown in Table 9, a driver's estimate of his ability to perform a combat-related composite task has virtually nothing to do with his personal background. Rather, it depends substantially on those background, congruence, and practice variables that reflect relevant experience such as: years in crew, number of crew positions held in training programs and Table VIIIs, particularly as a driver; months of experience in the position; and frequency and recency of practice. While we are reluctant to view these relationships as indicative of those which might have been obtained had greater variance in test scores been found, it is interesting to note that the independent variables correlating with the Task 1 performance criterion measure also correlate with the confidence estimates.

Loaders. Zero-order validities for the performance data on each of nine loader composite tasks are presented in Table 10. Performance on the first task (checking and adjusting track tension) does not correlate significantly with any of the independent variables. Presumably the failure to demonstrate strong correlates of performance on this task is a function of the relatively small subset ($n=19$) of loaders who were tested on it. On each of the remaining eight loader composite tasks, background and congruence variables are found that do covary with the proportion-of-steps-passed measure, many of them being particularly strong correlates of performance.

As was true for the driver sample, variables describing the personal background of loaders do not correlate significantly with performance. Instead, performance tends to vary as a function of such background variables as skill level of primary MOS, years in crew, and number of positions held in training programs. Among the congruence variables, those that significantly correlate with performance on more than one task include duty MOS skill level, having held the loader position in training programs, and hours per day spent in

Table 10. Zero-Order Validities* Between Background/Congruence/Practice Variables and Proportion of Steps Passed by Loaders in Nine Composite Tasks

(Readiness Test Data)

		1	2	3	4	5	6	7	8	9	Prop. Go	H.O.P.
Personal Background	Weight	--	--	--	--	--	--	--	--	--	--	--
	Height	--	--	--	--	--	--	--	--	--	--	--
	Age	--	--	--	--	--	--	--	--	--	--	--
	Education	--	--	--	--	--	--	--	--	--	--	--
	Like Job	--	--	--	--	--	--	--	--	--	--	--
	Reenlist	--	--	--	--	--	--	--	--	--	--	--
Army & Job Background	Months Served	--	--	--	--	--	--	--	--	--	--	--
	Rank	--	41	--	--	--	--	--	--	--	--	--
	Months in Primary MOS	--	--	--	--	--	--	--	--	--	--	--
	Skill Level-Primary MOS	--	--	51	42	47	**63**	**58**	43	**56**	44	**61**
	Years in Crew	--	43	--	48	--	45	43	--	--	--	45
	Number Positions Held in Training	--	--	--	--	41	**57**	**52**	**49**	--	**50**	**53**
	Number Table VIIIs	--	--	--	47	--	--	--	--	--	--	38
Congruence	Months in Duty MOS	--	--	--	--	--	--	--	--	--	--	--
	Skill Level-Duty MOS	--	--	45	--	--	**56**	51	--	48	--	**55**
	Test Position-Job Title Match	--	--	--	--	--	--	42	--	--	**48**	36
	Position Training	--	--	--	41	--	42	--	47	--	**47**	43
	Position Table VIIIs	--	--	--	--	--	42	--	--	--	--	--
	Months in Position Past Year	--	--	--	--	--	52	--	--	--	**48**	40
	Hours Per Day in Position	--	--	--	--	--	50	51	--	--	**52**	43
	Platoon Sergeant Ratings	--	--	--	--	--	--	43	--	--	--	--
Frequency of Practice	Task 1	--	--	46	--	**61**	**73**	**65**	--	**61**	**49**	**61**
	Task 2	--	--	--	--	--	--	--	--	--	--	--
	Task 3	--	--	--	49	--	45	--	--	--	--	39
	Task 4	--	--	--	45	--	**57**	**60**	--	44	43	**47**
	Task 5	--	--	--	--	--	40	42	--	--	39	--
	Task 6	--	--	--	--	--	--	--	--	--	--	--
	Task 7	--	--	--	--	--	41	41	--	--	37	--
	Task 8	--	--	--	--	--	48	--	--	--	--	--
	Task 9	--	--	42	--	--	**52**	45	--	--	--	--
Recency of Practice	Task 1	--	--	50	--	**64**	**77**	**66**	--	**62**	44	**60**
	Task 2	--	--	--	--	--	--	--	--	--	--	--
	Task 3	--	--	--	--	--	**56**	44	--	45	--	41
	Task 4	--	--	--	--	--	**64**	**56**	--	**51**	39	**48**
	Task 5	--	--	--	--	--	--	--	--	--	--	--
	Task 6	--	--	--	--	--	--	--	--	--	--	--
	Task 7	--	--	--	--	--	--	--	--	--	--	--
	Task 8	--	--	--	--	**62**	**64**	**52**	--	--	45	**50**
	Task 9	--	--	48	--	48	**65**	**54**	--	**53**	--	**52**

*Correlations rounded to the nearest hundredth with decimal points omitted.
Correlations with significance levels $\leq .001$ are indicated in bold face; others are significant in the range $\leq .01$, $> .001$.

loader activities. For the most part, the recency and frequency with which each task has been practiced do not relate to the performance measure for that task, a finding consistent with the general pattern of results obtained for the driver sample. Retest data were available for too few loaders ($n = 6$) to determine whether these same general patterns of validities characterized performance during the second administration of the readiness tests.

The large number of correlates significantly associated with performance on the individual loader composite tasks is reflected in the summary performance measures as shown in Table 10. The correlates with the overall proportion-of-steps measure (HOP) and the proportion-of-tasks-GO measure (Prop Go) are many and quite strong, reflecting the positive relationship between relevant experience as a loader and successful performance of loader tasks. The obtained correlates are presumably as strong as they are because job congruence is definitely an issue among the loaders in our sample. Only 58% reported that they were routinely functioning as loaders, having done so for an average of 10.83 months. With respect to SQT 2 scores, the longer a loader has worked in that position (Duty Time) the higher is his hands-on performance score expressed either as the number of steps GO ($r=.72$, $p < .01$) or the number of tasks GO ($r=.58$, $p < .01$). A finding that does not make intuitive sense is that the hours-per-day measure correlates negatively with SQT 2 performance expressed as the number of hands-on tests on which a GO is received ($r=-.68$, $p < .01$). An explanation for this relationship is not readily apparent.

Two other findings are of interest. Overall performance, especially HOP, correlates significantly ($p < .01$) with reported frequency and recency of performance on Task 1 (check and adjust track tension), Task 4 (prepare to fire procedures), Task 8 (maintaining the machinegun), and Task 9 (maintaining the breechblock). (In the latter two cases, the frequency correlates are significant at $p < .018$.) These four tasks represent the spectrum of activities for which a loader is responsible. The other five composites also deal with either main gun firing, or the M219 weapon subsystem. The second interesting finding specific to loaders is the number and the strength of the correlations with performance on Tasks 6 and 7. Both of these represent emergency procedures to be carried out in the event of main gun or machinegun misfires, and unlike the other tasks, cannot be considered as fairly normal or routine loader activities.

Correlations between the various independent variables and loader confidence estimates are shown in Table 11. As in the driver data, virtually all of the background, congruence, and practice variables reflecting experience with the composite tasks correlate significantly with estimates of

Table 11. Zero-Order Validities* Between Background/Congruence/Practice Variables and Loader Confidence Estimates for Nine Composite Tasks

		TASKS 1	2	3	4	5	6	7	8	9
Personal Background	Weight	—	—	—	—	—	—	—	—	—
	Height	—	—	—	—	—	—	—	—	—
	Age	—	—	—	—	—	—	—	—	—
	Education	—	—	—	—	—	—	—	—	—
	Like Job	—	—	—	—	—	—	—	—	—
	Reenlist	—	—	—	—	—	—	—	—	—
Army & Job Background	Months Served	—	—	—	—	—	—	—	—	—
	Rank	—	—	—	—	—	—	—	—	—
	Months in Primary MOS	—	—	—	—	—	—	—	—	—
	Skill Level-Primary MOS	**58**	**47**	**56**	**67**	**62**	**65**	**51**	**51**	*43*
	Years in Crew	*41*	**54**	**51**	**53**	**57**	**60**	**48**	**58**	*44*
	Number Positions Held in Training	**65**	**47**	**56**	**65**	**66**	**61**	**64**	**62**	*42*
	Number Table VIIIs	*41*	**47**	*45*	**52**	**49**	**50**	*45*	**46**	*37*
Congruence	Months in Duty MOS	—	—	—	—	—	—	—	—	—
	Skill Level-Duty MOS	**52**	*47*	**56**	**61**	**59**	**60**	**51**	*48*	—
	Test Position-Job Title Match	*38*	—	**48**	**56**	**56**	**50**	*39*	**48**	—
	Position Training	**48**	—	—	*43*	**48**	*38*	**49**	*44*	—
	Position Table VIIIs	*37*	—	*39*	**48**	*43*	*39*	**48**	—	—
	Months in Position Past Year	*44*	*43*	*46*	*42*	*46*	**48**	—	**48**	—
	Hours Per Day in Position	*45*	*44*	*46*	*45*	**50**	*42*	—	**49**	—
	Platoon Sergeant Ratings	*38*	*39*	*41*	*42*	*47*	**48**	**52**	**52**	*46*
Frequency of Practice	Task 1	**77**	**49**	**60**	**59**	**67**	**68**	*45*	**60**	—
	Task 2	*44*	*43*	*40*	**48**	**47**	*44*	**52**	**50**	—
	Task 3	**64**	**50**	**62**	**60**	**66**	**65**	**48**	**62**	—
	Task 4	**56**	**53**	**62**	**59**	**61**	**62**	**54**	**61**	*45*
	Task 5	**51**	**47**	**53**	**61**	**55**	**46**	**52**	**55**	—
	Task 6	*38*	*42*	*44*	**49**	*44*	*41*	*44*	**48**	—
	Task 7	**49**	*46*	**58**	**66**	**56**	**47**	**56**	**54**	—
	Task 8	**59**	**59**	**70**	**64**	**61**	**67**	**60**	**64**	—
	Task 9	**60**	**52**	**64**	**57**	**57**	**62**	**58**	**60**	*33*
Recency of Practice	Task 1	**73**	—	**59**	**58**	**63**	**68**	*43*	**52**	—
	Task 2	*38*	—	—	—	—	—	—	—	—
	Task 3	**59**	—	**52**	**51**	**59**	**55**	*38*	**53**	—
	Task 4	**54**	—	**53**	**55**	**59**	**54**	*46*	**50**	—
	Task 5	**54**	*43*	**50**	**57**	**50**	**50**	*38*	*45*	—
	Task 6	**48**	—	*39*	**47**	*42*	*44*	—	*41*	—
	Task 7	**51**	**47**	*44*	**54**	**50**	**49**	*33*	*45*	—
	Task 8	**69**	*42*	**63**	**60**	**62**	**67**	*33*	**57**	—
	Task 9	**69**	—	**60**	**55**	**59**	**66**	—	**51**	*40*

*Correlations rounded to the nearest hundredth with decimal points omitted. Correlations with significance levels ≤ .001 are indicated in bold face; others are significant in the range ≤ .01, > .001.

one's ability to perform each task. Again, the personal background variables are not strong correlates of stated confidence. The earlier caveat about the confidence estimates as valid surrogates of performance must again be voiced. However, as reported earlier, highly significant ($p < .01$) correlations between confidence and performance were obtained on five of the nine tasks (#3, #5, #6, #7 and #9) and in the other four instances they were significant at slightly lower levels ($p < .05$).

Gunners. Zero-order validities were estimated for the 11 different composite tasks that were used to characterize gunner performance. With respect to the proportion-of-steps-passed performance measure, there are virtually no significant relationships with variables representing personal background, Army and job background, congruence, or frequency and recency of practice. In the entire set of 473 validity coefficients, only eight are significant. One represents a positive correlation between platoon sergeant ratings and performance on the prepare-to-fire-procedures task (Task 1, $r=.35$, $p < .01$). Another indicates a negative relationship between gunner weight and performance during preparation of the telescope for operation (Task 2, $r=-.34$, $p < .01$). The six remaining coefficients essentially represent negative relationships between frequent and recent practice on Task 11 (main gun misfire procedure) and performance on periscope tasks (#3 and #6) and a machinegun task (#9). The alternative performance measures are consistent with this same general pattern. None of the independent variables correlates meaningfully with the retest proportion-of-steps-passed data, or with the overall proportion-of-tasks-GO or HOP measures. The failure to find significant correlations between Army and job background and congruence variables and performance is at odds with findings reported for drivers and loaders.

While correlations between the independent variables and performance on the composite readiness tests are infrequent and haphazard, those obtained when the SQT measures are examined are more consistent. At first glance, however, the obtained relationships are distinctly counterintuitive. Performance on SQT2, as measured by the number of hands-on tests in which a GO is received, is lower the longer gunners have been members of crews ($r = -0.36$, $n = 43$, $p = .009$), the more months they have served ($r = -.44$, $n = 44$, $p = .001$), and the longer they as 11Es have served in the gunner slot ($r = -.38$, $n = 44$, $p = .005$).

One speculation about the inverse relationship between experience and SQT performance derives from the fact that 87% of the gunner sample is comprised of troops whose normal job is that of a gunner. They have been in that position for an average of 10.63 months and have been in the service and have functioned in a tank crew for over two years. As such they are relatively congruent with the gunner's job, a possibility that may be reflected by their high levels of readiness test performance (means range from .74 to .95 with half of the tasks in excess of .86) and relatively low variability (SDs range from .07 to .33). It is conceivable that these kinds of experienced troops may have devised ways of performing their tasks not strictly in accordance with SOP. Such an approach to performance of the hands-on tasks in the SQT might have been pursued by the more experienced and "savvy" gunners. However, because of the strict attention to scoring of steps and sequence of steps in the SQTs, their performance would have suffered. Some support for this notion comes from the readiness retest data. We have been advised that this second administration was scored much more stringently with steps and sequences of steps having to be performed precisely in accordance with specified standards. As a consequence gunner performance dropped from a mean of .85 on the first administration to a mean of .54 steps passed on the retest.

Correlations between gunner confidence estimates and the independent variables are, unlike those for drivers and loaders, largely insignificant. Gunner estimates of confidence are essentially unrelated to personal background, Army and job background, or congruence variables. On the other hand, frequency and recency of practice relate significantly to confidence estimates for a number of the tasks. Frequency of practice is related to confidence estimates for Task 4 ($r = .45$, $p < .001$), Task 5 ($r = .36$, $p < .01$), Task 7 ($r = .34$, $p < .01$) and Task 10 ($r = .34$, $p < .01$). Recency of practice relates to Task 5 ($r = .36$, $p < .01$), Task 8 ($r = .37$, $p < .01$), Task 9 ($r = .35$, $p < .01$) and Task 10 ($r = .38$, $p < .01$). The finding that the confidence which gunners have in their performance of specific hands-on tasks is related to reported frequency and recency of practice parallels the results for drivers and loaders.

For gunners the substitutibility of confidence data for performance criteria is more problematic than for the other two crew positions examined thus far. The two measures are significantly related on only one of the composite tasks (Task 3, $r = .35$, $p < .01$) although on two others (Tasks 8 and 10) substantial positive correlations are obtained ($p < .05$).

Tank Commanders. Results of the zero-order validity analyses of tank commander criterion data resemble those obtained for gunners, particularly with regard to the proportion-of-steps-passed measure. Out of 410 possible correlations (10 composite tasks and 41 independent variables) only three are significant. Tank commanders who report that they like their job perform better when servicing the M85 machinegun (Task 2, $r = .36$, $p < .01$) and maintaining it (Task 4, $r = .43$, $p < .001$). The third significant validity coefficient reflects the fact that participation in gunnery training programs specifically as a tank commander is related to performing main gun prepare to fire procedures (Task 1, $r = .33$, $p < .01$). The fairly substantial amount of readiness retest data available for tank commanders reflect the same paucity of significant correlations. Performance during preparation of the rangefinder is inversely related to one's weight (Task 5, $r = -.46$, $p < .01$). The more months during the past year that one spent engaged in tank commander activities the better was his performance during prepare to fire procedures (Task 1, $r = .57$, $p < .001$).

Proficiency expressed in terms of the summary performance measures covaries with liking one's job, months spent as an 11E, months actually working as a tank commander, and whether one has occupied this crew position in training programs and Table VIIIs. The seven significant ($p < .01$) correlations with the HOP and Proportion GO measures range from $r = .31$ to $r = .36$. Correlations with SQT scores are in a few instances significant, but involve rather small subsamples of tank commanders. For example, on SQT2 the proportion of steps passed on the hands-on tasks is related positively to rank ($r = .86$, $n = 8$, $p < .003$). Interestingly, the proportion of SQT2 hands-on tasks on which tank commanders receive a GO is inversely related to the months one has served as a tank commander and one's participation in training programs while occupying that position (both $\underline{r}$s $= .69$, $n = 13$, $p < .01$). These results resemble those obtained for gunners on the same set of hands-on tasks; the same speculation applies. The longer one has been a tank commander the more likely he is to have developed idiosyncratic ways of performing many of his tasks.

Significant correlations between tank commander confidence estimates about performance on the 10 readiness tests and various independent variables are shown in Table 12. The obtained correlations resemble those reported for drivers and loaders more than they do those few relationships established for gunners. Among the personal background variables, whether one likes his job is related to confidence estimates on four of the ten readiness tasks. The bulk of the significant relationships, however, involve Army and job background as well as congruence variables. Similarly, strong relationships exist between the confidence expressed about performance of a given task and the frequency and recency with which that task has been practiced. These relationships in particular are similar to those found in the two more junior crew positions. As in the case of the gunners, caution must be used in

Table 12. Zero-Order Validities* Between Background/Congruence/ Practice Variables and Confidence Estimates for Tan Tank Commander Composite Tasks

		TASKS									
		1	2	3	4	5	6	7	8	9	10
Personal Background	Weight	--	--	--	--	--	--	--	--	--	--
	Height	--	--	--	--	--	--	--	--	--	--
	Age	--	--	--	--	--	--	--	--	--	--
	Education	--	--	--	--	--	--	--	--	--	--
	Like Job	--	37	42	--	--	--	--	--	**46**	32
	Reenlist	--	--	--	--	--	--	--	--	--	--
Army & Job Background	Months Served	--	--	--	--	--	--	--	--	--	--
	Rank	--	--	--	--	34	--	--	--	--	34
	Months in Primary MOS	36	--	--	--	--	--	--	--	--	--
	Skill Level-Primary MOS	--	--	--	--	--	--	--	--	--	--
	Years in Crew	--	32	35	--	--	--	--	--	--	--
	Number Positions Held in Training	--	--	--	--	32	38	--	--	34	--
	Number Table VIIIs	38	--	--	--	31	--	--	--	--	--
Congruence	Months in Duty MOS	39	--	--	--	36	31	--	39	37	--
	Skill Level-Duty MOS	--	--	--	--	--	--	--	--	--	--
	Test Position-Job Title Match	--	--	--	--	--	--	--	--	--	--
	Position Training	--	--	33	--	34	--	--	32	36	**46**
	Position Table VIIIs	**45**	--	--	--	37	--	32	33	--	--
	Months in Position Past Year	33	--	35	--	--	**40**	--	--	--	--
	Hours per Day in Position	--	--	38	35	--	--	--	33	--	--
	Platoon Sergeant Ratings	--	--	--	--	--	--	--	--	--	--
Frequency of Practice	Task 1	**39**	--	--	--	33	--	35	--	--	--
	Task 2	--	35	37	**46**	--	--	--	33	--	--
	Task 3	--	--	**49**	**47**	32	--	--	**40**	--	37
	Task 4	35	--	37	**43**	--	--	--	34	--	--
	Task 5	34	--	35	--	34	34	33	--	--	--
	Task 6	--	--	--	--	--	--	--	--	--	--
	Task 7	--	--	37	--	39	35	**45**	36	34	38
	Task 8	--	34	35	39	36	--	--	38	--	32
	Task 9	--	**42**	38	35	34	--	--	**45**	--	--
	Task 10	37	--	--	--	35	--	--	35	33	--
Recency of Practice	Task 1	--	--	--	--	--	--	--	--	--	--
	Task 2	--	--	--	33	--	--	--	33	--	--
	Task 3	--	--	**40**	--	--	--	--	--	--	--
	Task 4	--	--	--	36	--	--	--	--	--	--
	Task 5	--	--	34	--	38	**48**	35	**43**	33	--
	Task 6	--	--	--	--	--	32	--	--	--	--
	Task 7	--	--	**41**	--	**46**	**46**	**52**	32	--	--
	Task 8	--	--	32	**42**	--	--	33	38	--	--
	Task 9	--	--	--	--	--	--	--	38	--	--
	Task 10	--	--	--	--	--	--	--	--	--	--

*Correlations rounded to the nearest hundredth with decimal points omitted.
Correlations with significance levels ≤ .001 are indicated in boldface; others are significant in the range ≤ .01, > .001.

considering the import of these relationships. The connection between confidence and performance is weak, a significant correlation being obtained on only one task (Task 2, $r = .39$, $p < .01$). The lack of relationships, however, may be due at least partially to the nature of the tank commander readiness performance data. Scores obtained in both the first and second administrations are very high on the average, and exhibit very small variance. The lack of correlations with these data, either with respect to the independent variables or the confidence estimates, must be considered in this light.

Summary. Examination of the zero-order validities reveals that variables describing experience, whether drawn from the Army and job background or the congruence categories, relate to levels of task performance in the more junior crew positions. Similar relationships are not obtained in the two more senior positions. An implication of these findings, and one supported by the relatively high performance levels of gunners and tank commanders and the proportion of both these samples who were occupying their normal crew positions and who were experienced at doing so, is that both of these samples are too homogeneous with respect to experience to reveal the congruence effects of interest. The driver and loader samples contain more "fills" and less experienced personnel, factors which should give rise to congruence effects if such truly exist. Were other, more heterogeneous samples of gunners and tank commanders investigated, it is predicted that their zero-order validities would more closely approximate the patterns associated with the drivers and loaders sampled in the present study.

In all cases except the gunner sample, job experience and practice variables relate to soldiers' confidence in being able to perform hands-on tasks. Based on the more junior crew members, there is evidence that suggests the confidence estimates do in fact relate to levels of proficiency.

Two-Way Frequency Table Analyses

In Phase I of the current effort relationships among several predictor variables, considered simultaneously, and selected criterion variables were explored by means of specialized multiple regression techniques. The results of these analyses varied among the four crew positions and from task to task within each crew position. The lack of more consistent findings may in part have been due to the use of these sophisticated statistical procedures on the relatively small sample sizes available for analysis.

In the current phase of research a rather different strategy was pursued in the attempt to determine whether several experience or congruence variables might relate to performance and/or confidence criterion measures. In essence, a case study approach was used in which all of the background information reported in the Background Information Questionnaire was carefully reviewed for each soldier. The question initially addressed was whether a brief narrative prepared for each soldier might not permit a global (but multidimensionally determined) judgment about his job congruence or experience. For example, subject #5, a driver is:

> "A 22-year old, 5'10", 145 lb., right-handed soldier with less than a high school diploma. He has been in the service for 31 months and has attained the rank of an E5. He has been an 11E20 for one month and has a duty MOS of 11E20. His present job title is that of a driver which he has held for 19 months. He has no secondary MOS but has had down-range driving experience, tank tables, and classes. He has been a driver in gunnery programs and has fired two Table VIIIs in that position. He has been in a tank crew from two to three years and during the past year has spent 240 days doing driver kinds of activities on an average of four hours a day. He likes his job and plans to reenlist. He will be tested as a driver."

As another example, subject #9, also in the driver sample, is:

> "A 19-year old, 5'7", 150 lb., right-handed soldier who is a high school graduate. He has been in the service for 17 1/2 months and has attained the rank of an E3. He has held primary and duty MOSs for the past 12 months of 63C10 and during that time has had the job title of track mechanic. He has no secondary MOS and has had no special gunnery training or experience with Table VIIIs. He's not been a member of a tank crew. During the past year he has spent 0 days as a driver. He doesn't like his job and won't reenlist. Although he normally repairs tracks he will participate in the research as a driver."

Based on these two and a number of similar cameos developed for each crew position, the idea emerged of locating soldiers along a number of branches in a tree-diagram. The branches were defined by the following binary variables:

1. The soldier is or is not being tested in the same position as his job title;
2. He does or does not have an 11E Primary MOS;
3. He does or does not have an 11E Duty MOS;
4. He has or has not participated in special gunnery training programs;
5. He has or has not crewed in Table VIII(s);
6. He has or has not been a member of a tank crew.

Using the branching strategy described above, it was possible to uniquely classify each soldier in one and only one terminal node. And in spite of the large number of nodes theoretically available, most soldiers in our samples fell into a relatively small number of categories. For example, it was possible to locate drivers in six basic categories. These ranged from the most congruent or experienced drivers (the yes, yes, yes, yes, yes, yes group) to the least congruent or experienced group (the no, no, no, no, no, no group). Incidentally, driver #5 depicted above fell into the former group, while driver #9 fell into the latter category.

Further attempts were made to differentiate among soldiers within a particular congruence category based on several experimental variables. These included:

1. Months in Primary MOS,
2. Duty Time,
3. Hours per Day,
4. Job time, and
5. Years in a crew.

Although this effort took place in the absence of information about zero-order validities it is interesting to note that these variables related to criterion measures for a number of tasks, and crew positions.

The results of the categorization process are depicted by crew position in Table 13. The categories are defined in terms of the binary and experiential variables described above. The number of soldiers in each sample that fall into a particular category are also indicated. Category assignments were made by one judge and reviewed by a second judge. The few instances in which disagreements arose were resolved jointly. In passing it should be noted that the levels-of-experience distinctions made in Table 13 were position-specific in terms of absolute amounts of Duty Time and Prime Time. As such they are post hoc descriptions and are not meant to imply specific levels of experience to be addressed in future work.

Table 13. Congruence Categories

	Drivers	Loaders	Gunners	Tank Commanders
1. Tested in appropriate crew position; 11E Primary and Duty MOS; In gunnery programs and Table VIIIs.				
1a) Very Experienced (PT > 24 mos.; UT > 12 mos.)	n = 6	n = 6	n = 8	n = 10 (PT > 48 mos.; UT > 32 mos.)
1b) Experienced (PT > 18 mos.; UT > 18 mos.)	n = 8	n = 5	n = 6 (UT > 11 mos.)	n = 9 (PT > 12 mos.; UT > 12 mos.)
1c) Inexperienced (PT < 24 mos.; UT < 12 mos.)	n = 6	n = 6	n = 22 (PT < 31 mos.; UT < 10 mos.)	n = 15 (UT < 9 mos.)
2. Same as Category 1 but no Table VIII experience.				
2a) Experienced (PT > 24 mos.)	n = 4	n = 0	n = 3 (PT > 23 mos.; UT > 8 mos.)	n = 1
2b) Inexperienced (PT < 24 mos.)	n = 1	n = 0	n = 7 (PT < 32 mos.; UT < 4 mos.)	n = 1
3. Same as Category 1.				
3a) Now occupying more senior crew position	n = 0	n = 4	n = 0	n = 0
3b) Now occupying lower or other position	n = 0	n = 4	n = 0	n = 0
4. Same as Category 2 but no special gunnery training.	n = 5	n = 3	n = 4	a) Experienced n = 1 b) Inexperienced n = 5
5. Tested in "appropriate" crew position; Non 11E Primary MOS but 11E Duty MOS. They are occupying the crew position in which they will be tested but have not been in special training programs or Table VIIIs.	n = 2	n = 1	n = 0	n = 1
6. Not tested in appropriate crew position. 11E Primary and Duty MOS but not active members of tank crew.				
6a) Sheridan personnel	n = 2	n = 0	n = 0	n = 0
6b) Other jobs	n = 3	n = 0	n = 0	n = 0
7. Same as Group 6b, but Non 11E Duty MOS.	n = 0	n = 1	n = 2	n = 0
8. Same as Group 6 but Non 11E Primary MOS, Non 11E Duty MOS. They have performed in the crew position at one time.	n = 0	n = 4	n = 0	n = 0
9. Not tested in appropriate crew position. 11E Primary MOS; 11E Duty MOS being tested in a totally new position.	n = 0	n = 0	n = 0	n = 2
10. Not tested in appropriate crew position. Non 11E Primary MOS; Non 11E Duty MOS; No special gunnery training programs; No Table VIIIs, not a crew member.				
10a) Turret Mechanics	n = 4	n = 5	n = 0	n = 0
10b) Track Vehicle Mechanics	n = 3	n = 2	n = 0	n = 0
10c) Other	n = 1	n = 1	n = 0	n = 0
	N = 45*	N = 42	N = 52	N = 45

*N = number of subjects with BIO data.

In the next stage of the analysis three judges considered each of the composite readiness tasks for each crew position in detail. They then ordered the categories of soldiers within each sample so as to conform to predicted levels of proficiency on each task. Categories of soldiers high in the ordering were predicted to be more proficient than those lower in the ordered groups. When distinctions in probable performance were not judged to be evident, groups were treated as ties and pooled. Predicted orders were developed independently for each individual task. They are presented in Table 14.

Using these rankings and criterion performance and confidence data, two-way frequency tables were developed for each task that related the ordered categories to the criterion data. One set of such tables was developed for the confidence estimate criterion data (four levels); another set was developed for the readiness test proportion-of-steps passed measure. For purposes of examining these latter data one performance category was used to represent perfect performance (1.00); the remaining scores on each task were split into two or three categories, the attempt being made to obtain an appreciable frequency in each category, and to have roughly comparable frequencies across the performance categories representing less than perfect (1.00) performance. An example of a two-way frequency table relating congruence to performance is presented in Table 15 for the driver task (Task #1) dealing with the preparation of periscopes.

The two-way frequency tables were analyzed in terms of a variety of measures of predictability, correlation, and association. BMDP program P1F (1977)[1] was used to explore the relationship between levels of congruity (represented by the rank-ordered categories of soldiers) and levels of the criteria (scaled confidence and performance data). Three statistics were evaluated including Somer's D, the Spearman rank-order correlation (ρ), and the Pearson χ^2. Somers' D provides the best test of the congruency hypothesis. It ranges from 1.0 to -1.0 and as a predictive measure represents the probability of similar ranking on two indices among cases with different values for the predictor index. Its significance (departure from zero) may be evaluated by means of a t statistic with approximately .4N degrees of freedom (Brown and Benedetti, 1977).[2] The Spearman ρ was used to express the relationship between the two variables as a correlation. χ^2, although not

1 BMDP Biomedical Computer Programs P-Series 1977. Dixon, W.J. & Brown, M.B. (Eds), University of California Press, 1977.

2 Brown, M.B., & Benedetti, J.K. Sampling behavior of tests for correlation in two-way contingency tables. J. Amer. Statist. Assoc. 72, 1977, 309-315.

Table 14. Categories of Soldiers Ordered in Terms of Predicted Performance

TASK*

	1	2	3	4	5	6	7	8	9	10	11
Drivers	1a**	1a, 1b, 1c }	1a	1a	1a	--	--	--	--	--	--
	2a	2a, 2b }	1b, 2a	1b, 2a, 10c	1b	--	--	--	--	--	--
	1b	4, 10a }	1c	1c	1c	--	--	--	--	--	--
	10a		2b, 4	2b, 4	2a	--	--	--	--	--	--
	1c, 2b, 4	5	5	5	2b	--	--	--	--	--	--
	5	10c	6a	6a	4	--	--	--	--	--	--
	6a, 6b	10b	6b	6b	6a	--	--	--	--	--	--
	10b, 10c	6a, 6b	10a, 10b, 10c	10a, 10b	5	--	--	--	--	--	--
					6b	--	--	--	--	--	--
					10a, b, c	--	--	--	--	--	--
Loaders	1a, 1b	1a	1a, 1b }	1a	1a, 1b }	1a	1a	1a	1a	--	--
	10b	1b	1a, 3a }	1b	1c, 3a }	1b	1b	1b	1b	--	--
	3a	1c		1c	4, 3b }	3a	3a	1c	1c	--	--
	1c	3a	4, 3b, 5 }	3a	5, 7, 8 }	1c	1c	3a	3a	--	--
	4, 3b, 5	4	7, 8 }	4		3b	3b	4	4	--	--
	7	3b		3b	10a, 10b, 10c	4	4	3b	3b	--	--
	8	5	10a, 10b, 10c	5		5	5	5	5	--	--
	10a	7		7, 8 }		8	8	7	7	--	--
	10c	8		10a, 10b }		7	7	8	8	--	--
		10a, b, c		10c }		10a, 10b, 10c	10a, 10b, 10c	10a	10a	--	--
								10b, 10c	10b, 10c	--	--
Gunners	1a, 1b	1a, 1b	1a, 1b	1a	1a	1a	1a, 1b	1a	1a, 1b	1a	1a
	2a	2a	2a	1b	1b	1b	2a	1b	2a	1b	1b
	1c, 2b, 4	1c, 2b, 4	1c, 2b, 4	2a	2a	2a	1c, 2b, 4	2a	1c, 2a, 4	2a	2a
	7	7	7	1c	1c	1c	7	1c	7	1c	1c
				2b	2b	2b		2b		2b	2b
				4	4	4		4		4	4
				7	7	7		7		7	7
Tank Commanders	1a	1a	1a, 1b }	1a, 1b }	1a	1a, 1b }	1a	1a	1a	1a	
	2a, 4a }	1b, 2a }	2a, 4a }	2a, 4a }	1b	1c, 2a }	1b	1a, 2a }	1b, 2a }	1b	
	1b, 5 }	4a, 5 }	5	5	2a, 4a }	2b, 4a }	2a, 4a }	4a, 5 }	4a, 5 }	2a, 4a }	
	1c	1c	1c, 2b }	1c, 2b }	5 }	4b, 5 }	5 }			5 }	
	2b, 4b	2b, 4b	4b }	4b }	1c		1c	1c	1c	2b, 4b	
	9	9	9	9	2b, 4b	9	2b, 4b	2b, 4b	2b, 4b	9	
					9		9	9	9		

* Refer to Table 1 for task descriptions.
** Refer to Table 13 for definitions of categories.
Groups listed on the same line or those bracketed were predicted to be equivalent.

Table 15. Two-Way Frequency Table
Relating Driver Congruence and Performance

TASK 1

Driver Categories	4 (1.0)	3 (.825-.99)	2 (.35-82)	1 (0-.34)
	Levels of Performance*			
1a	4	1	1	0
2a	1	2	1	0
1b	3	1	2	0
10a	2	0	1	1
1c, 2b, 4	3	1	5	3
5	0	0	1	1
6a, 6b	0	1	1	2
10b, 10c	1	0	0	3

*Performance expressed as the proportion of steps passed.

a particularly powerful test of the hypothesis under consideration in this research, was also computed. When necessary, because of small expected cell frequencies, χ^2 was recomputed on a table in which adjacent rows or columns were collapsed. Collapsing occurred automatically and continued until the smallest expected frequency reached or exceeded 1.0. The results of these analyses are described below for each position.

Drivers. A significant relationship between congruence and hands-on performance, measured in terms of the proportion of steps passed, exists only for Task #1, the "preparation of periscopes". The rank ordering of the eight levels of congruence correlates significantly with performance ($\rho = .522$, $df = 40$, $p < .01$). The correlation is in the predicted (one-tail) direction. Somers' D is also significant ($D = 404$; $t = 4.158$, $df = 16$, $p < .0005$), indicating that a significant predictive relationship exists; i.e., prediction of level of performance from knowledge about congruence. The χ^2 for the collapsed table (four levels of congruence) is not significant by conventional standards. No significant relationships exist for the four remaining tasks.

Results for the confidence data are presented in Table 16. In each instance there is evidence that the two variables are related, and more importantly, that congruence bears a significant predictive relationship to confidence estimates. These results, together with those for the performance data, parallel the findings presented earlier in the zero-order validity analyses.

Loaders. Results for the loaders are presented in Table 17 for both the performance and confidence criterion data. Again the striking feature of these results is how the significant relationships on several of the tasks mirror the zero-order validity findings.

Performance is significantly related to congruence on six of the nine tasks. The exceptions are interesting. In Task 1 ("check and adjust track tension") differentiations were made among nine levels of congruence and between two levels of performance (perfect and less than perfect). The 19 loaders having data were mapped onto this structure. Since 13 of these performed perfectly (and another two nearly so), relatively few cases were available with which to explore poorer performance. On the other hand, perhaps judges more experienced with the subject matter would not have made such fine differentiations among loader categories. In any event, across the several different kinds of analyses, few correlates of performance on this task have been found.

Table 16. Two-Way Frequency Table Analyses: Driver Confidence

Task	Dimensions*	D	ρ	χ^2	Collapsed Dimensions
1	8x4	.522[a]	.688[b]	26.402[b]	(4x4)
2	5x4	.362[c]	.313[c]	3.403 N.S.	(3x2)
3	8x4	.542[a]	.688[b]	19.438[c]	(6x3)
4	8x4	.463[a]	.605[b]	23.061[c]	(6x3)
5	10x4	.599[a]	.768[b]	41.51[b]	(6x4)

*Levels of congruence by levels of performance.

a) $p < .0005$

b) $p < .01$

c) $p < .05$

Table 17. Two-Way Frequency Table Analyses:
Loader Performance

Task	Dimensions*	D	ρ	χ^2	Collapsed Dimensions
1	8x2	.161	.261	2.86	(3x2)
2	10x3	.141	.203	11.11	(5x3)
3	3x3	.196	.232	1.06	(3x2)
4	8x4	.295[c]	.372[c]	11.22	(5x3)
5	2x3	.636[c]	.438[b]	12.32[b]	(2x3)
6	10x3	.474[a]	.659[b]	36.11[a]	(7x3)
7	10x3	.353[b]	.527[b]	14.75[c]	(7x2)
8	11x4	.275[c]	.357[c]	14.89[c]	(4x3)
9	11x4	.287[b]	.425[c]	9.83	(4x3)

Loader Confidence

Task	Dimensions*	D	ρ	χ^2	Collapsed Dimensions
1	8x4	.349[b]	.472[b]	28.61[b]	(6x3)
2	9x4	.395[b]	.528[b]	29.33[b]	(5x4)
3	8x4	.552[a]	.606[b]	19.80[a]	(6x3)
4	8x4	.410[a]	.575[b]	27.50[b]	(6x3)
5	2x4	.859[b]	.625[b]	20.00[a]	(2x4)
6	9x4	.428[a]	.654[b]	31.34[b]	(8x3)
7	9x4	.391[b]	.529[b]	21.73[c]	(5x4)
8	10x4	.422[a]	.593[b]	18.74[c]	(5x3)
9	10x4	.19	.272	8.80	(6x3)

*Levels of congruence by levels of criterion.

a) $p < .0005$

b) $p < .01$

c) $p < .05$

In Task 2 ("boresight the M219") performance is simply not related to levels of congruence as we have defined it. Inspection of the underlying performance data reveals several instances in which the performance for a given level of congruence is high, intermediate and low, independent of the particular level of congruency considered. Finally, on Task 3 ("stow main gun rounds") only three levels of congruence were established, groups of loaders being pooled to form the three strata. Had the third category of loaders (comprised of personnel who presumably were "fills" for this crew position) been combined with those in the second congruence category, a stronger relationship might have been demonstrated. As it was, and contrary to our predictions, three of these troops (out of five for whom there were data) performed the task perfectly.

Expressions of loader confidence are strongly related to levels of congruity on all but one of the tasks. For both loaders and drivers, therefore, confidence clearly covaries with a relatively small set of variables reflecting one's on-the-job-experience.

<u>Gunners and tank commanders</u>. Analyses of the frequency tables developed for these two more senior crew positions reveal few relationships to be significant. In essence gunner performance is not related to judgments about congruence. None of the ten analyses is significant. With respect to the confidence criterion, relationships are found on three tasks but these are only marginally significant (Task 1, $D = .285$, $p < .05$; Task 2, $D = .266$, $p < .05$; and Task 4, $D = .289$, $p < .01$). With these exceptions, relationships between congruence and gunner criterion measures remain largely inscrutable.

The same situation is true for tank commander performance data. Congruence and performance are related on only one task and marginally so (Task 3, $D = .282$, $p < .025$). However, when the confidence data were examined, a number of relationships were uncovered. These are summarized in Table 18.

Table 18. Two-Way Frequency Table Analyses: Tank Commander Confidence

Task	Dimensions*	D	ρ	χ^2	Collapsed Dimension
1	5x4	$.356^b$	$.437^b$	11.76	(4x3)
2	5x3	.020	.030	2.79	(4x2)
3	3x4	$.472^b$	$.458^b$	8.38^c	(2x3)
4	3x4	$.314^c$	$.324^c$	4.17	(2x3)
5	6x4	$.406^b$	$.500^b$	16.29^c	(4x3)
6	2x4	**	**	**	**
7	6x4	$.248^c$	$.310^c$	6.47	(4x3)
8	5x4	$.302^c$	$.364^c$	10.99	(4x3)
9	5x4	$.358^b$	$.446^b$	11.47^b	(4x2)
10	5x4	$.376^b$	$.480^b$	15.06^c	(4x2)

*Levels of congruence by levels of confidence.

**Not computed; two congruence groups were formed in one of which n = 1.

a) $p < .0005$

b) $p < .01$

c) $p < .05$

Summary

One of the primary objectives of Phase II research was an intensive investigation of the "congruity hypothesis". Toward this end two outcomes were viewed as essential. First, it had to be determined whether any of the myriad independent (congruity) variables covaried with any of several criteria reflecting "effectiveness of performance on combat-related tasks". Second, it had to be determined whether such relationships or effects (if they existed) were consistent across different tasks, whether considered within or between crew positions; the latter distinction was necessitated by virtue of the heterogeneous sample of 11E troops available for study. To provide a fair evaluation of these issues substantial project resources were invested in cleaning up the underlying data base and then in analyzing it productively.

In both this and the preceding phase, investigation of the first issue focused on the demonstration that significant correlations (zero-order validities) existed between the independent and dependent variables. Such a demonstration has in fact occurred. Most if not all of the independent variables have been found to correlate significantly with a dependent measure in at least one instance. Several, particularly those drawn from the Army and job background and congruence categories, have repeatedly correlated with dependent variables. This fact has been amply documented in the zero-order validity results presented earlier in this section.

The rub, of course, is that while such a demonstration is necessary, it is not sufficient. It must also be demonstrated that the independent variables, or at least some subset of them, repeatedly and consistently relate to a given criterion measure. Alternatively, one must be able to account for and eventually predict those instances that prove to be exceptions. The required demonstration was not provided in Phase I, either in terms of the zero-order validities calculated for individual variables or the "best subsets" of variables identified during the regression of multiple (potential) predictors. In either case the result was the same: there were a few independent variables that repeatedly relate to a given criterion measure, did not do so with the degree of consistency one would like to have seen.

The issue was addressed in Phase II by attempting to distill a small subset of the Army and job background and congruence variables into a single index that could be used to portray congruence. The possibility of relating this index to various criterion measures was then investigated by means of the two-way frequency tables discussed above. With respect to the performance criterion measure, instances were found in which the congruence index related significantly to the proportion-of-steps-passed measure. The relationship varied, however, across criterion tasks and across crew positions. An examination was also conducted of the relationship between the congruence index and confidence estimates about one's ability to perform combat-related tasks. While the results were somewhat more consistent across tasks within a single position, they varied markedly between crew positions. In neither case, therefore, whether considering performance or confidence estimates, was the required level of consistency in results obtained. As a consequence, an alternative strategy was pursued. As discussed in the next section, attempts were made to construct a taxonomy of the combat-related tasks that might give rise to more consistent patterns of results, thereby permitting generalizations about the specific tasks comprising any position, job, or MOS.

AN ALTERNATIVE TASK TAXONOMY

The function-based task grouping adopted in the previous sections of this report is only one approach to analyzing the confidence and performance measures and their relationship to congruence. A problem with this approach, however, is that it is by design restricted to tank crew skills; any generalization of task groupings to other MOS designations is impossible. This section is concerned with the conceptualization, development, and partial evaluation of an alternative method for organizing tasks that could potentially be more valuable for unraveling the relationships of concern and extending them to other jobs.

One approach is to consider the congruency issues relative to a taxonomy that differentiates tasks on the basis of the cognitive and information-handling processes involved. Ideally, this approach would detect selective congruence effects that vary according to the cognitive processes involved in task performance. Such a taxonomy would have the added benefits of applicability to the other military occupations and compatibility with some of the new directions in selection and placement testing. This taxonomy could thus provide a good generalizable tool for analyzing the relationship between criterion-task performance and congruity.

Development of the Taxonomy

The hypothesized influence of job congruity on criterion-task performance stems from at least two factors: (1) the amount and kind of the practice a soldier receives on a criterion task on the job, and (2) the amount and nature of transfer from normal duty, on-the-job tasks to criterion tasks. In the first situation the soldier performs the criterion task as a part of his regular job and the issue is how often and under what conditions the task is performed. In the second situation the criterion task is not performed as part of the soldier's regular job and the issue is how the on-the-job tasks interfere with or facilitate performance of the criterion tasks. The latter issue is a very important one; however, it requires a task and job analysis that is much too complex and task-specific for present purposes. The practice issue fits much better into the goals of the preceding paragraph. Thus in developing the taxonomy described below, the overriding concern in selection of the dimensions was the availability of information about the nature of changes in performance on a potential dimension as a function of the recency and frequency of practice. A knowledge of the nature of these effects will allow us to relate the taxonomy to congruity through the practice issue.

The dimensions of the taxonomy were also made "independent and orthogonal" in the sense that the operational definitions of each dimension required distinct pieces of information. This provided maximum descriptive flexibility since the tasks could be described by either a set of values on dimensions or by a single dominant dimension. Using these criteria, five dimensions were derived for the present set of tasks. They are presented below.

Procedural Dimension. The procedural dimension involves those aspects of the task that require the operator to follow a fixed sequence of actions. For example, the dismounting and disassembly of the machinegun would be heavily weighted on the procedural dimension since the steps involved have to be performed in a pre-determined order; the soldier has little or no freedom in the order in which the steps are performed. Because the required sequence of responses is fixed, performance on this dimension should vary in much the same way as in the serial learning tasks extant in the experimental literature. Thus one would expect strong effects of practice, with task performance increasing with both the recency and the frequency of practice. Also, the longer or more complex the procedural aspects, the more influenced by the practice variables we might expect task performance to be.

Perceptual Dimension. The perceptual dimension involves the aspects of the task in which the operator is required to use perceptual processing in vigilance, inspection, detection, monitoring, identification or discrimination. For example, acquiring targets or checking the machinegun mainly involve perceptual processing for detection of targets and defects, respectively. These tasks would be heavily weighted on the perceptual dimension. Performance on this type of task has also been found to be strongly affected by practice. As Welford states,

> "On the perceptual side, practice seems to enable the skilled performer to select from among the mass of data impinging on his sense organs so that he neglects much of what is, to an unskilled person striking, and reacts strongly to data that a normal observer would fail to notice. Practice may also enable him to make absolute judgments with much greater precision than would otherwise be possible.[1]

It appears that for this dimension, frequency of practice might be more heavily weighted than the recency of practice.

[1] Welford, A. T. Fundamentals of Skill. London: Methuen & Co., Ltd., 1968.

Communicative Dimension. This dimension concerns those aspects of a task that require the operator to accurately receive or transmit information for successful completion of the task. Many of the firing procedures, for example, require communication among the crew for successful performance. The extent to which crew members are familiar with the standard procedures and language for communicating will determine the number of steps completed and the speed with which the steps can be completed. This communication, because it requires use of a standard, but unique language, is strongly affected by the practice variables. The ability to access the necessary codes and symbols in memory should increase with increasing frequency and recency of practice, and thus performance of tasks weighted heavily on this dimension should increase in a similar manner.

Feedback Dimension. This dimension concerns those attributes of the task that provide the operator with information about the correctness of his response which he can use to direct his behavior through completion of the task. Example behaviors would include aligning, tracking, and some adjusting activities. For the present situation, performance on this dimension should be relatively free of practice effects; since the task performance itself provides the information which directs behavior, the operator's ability to access information in memory should have little, if any, effect.

Motor Dimension. The motor dimension involves those aspects that require physical proficiency in a motor skill for successful task completion. An example would be the motor coordination required to drive the tanks. Such tasks should show few, if any, effects from the recency and frequency of practice, since the literature shows retention of motor skills to be very high.

In sum, the proposed taxonomy describes tasks in terms of their values on five dimensions: procedural, perceptual, communicative, feedback, and motor. Performance on the first three of these dimensions should decrease as the recency of practice or the amount of previous practice decreases, while performance on the last two dimensions should be relatively unaffected by the recency or amount of practice (given prior competence).

Evaluation

A two-phase approach was used to study the usefulness of the taxonomy as a tool for analysis. The first phase involved determining that performance on task groups derived from dimensions of the taxonomy were related to variables in the predictor set. The second phase involved demonstrating that the dimensions operated in application as predicted in theory. Here, another set of correlations relating specific experience and dimension performance was inspected for specific patterns predicted by the information available on practice effects for that dimension. The results of these two analyses were used as a basis for evaluating the taxonomy.

Prior to initiation of the first analysis, the criterion tasks were analyzed with respect to the taxonomic dimensions. This was accomplished by a consensual rating of the tasks by three judges. After a consensus was reached as to the dimensions present in each task, the tasks were then consensually rated as to the single dominant dimension in the task. In some cases it was impossible to assign a single dominant dimension; in these cases two or more dominant dimensions were assigned. In several other cases, a "minor dominant" dimension was also designated. These were cases where a secondary dimension was clearly judged to be potentially predictive of performance, but did not adequately describe the task if considered alone. The results of these assignments for each task are presented in Table 19 by crew position. The table shows that, in general, the tasks were judged to be predominantly procedural and that only the gunner and commander positions included tasks where communication was considered a dominant dimension.

Tasks were then combined according to the dominant dimensions. Two combination rules were used. The first combined those tasks that were rated major dominant and those rated minor dominant on a particular dimension. This grouping combined tasks on which the dimension(s) was considered as the primary influence on performance. In the second type of combination, only the tasks that were rated major dominant on a particular dimension were combined. This grouping thus eliminated the "minor dominant" dimensions for each task. These two types of grouping were used as the foundation for the subsequent analyses.

Descriptive data for performance (proportion of steps GO) are shown in Table 20. As can be seen, performance levels were quite consistent within groups, with the exception of the feedback tasks for loaders. These data parallel other performance data for the alternate task groupings, and unfortunately do not lead to the generation of any hypotheses of differential difficulty of task dimensions.

The first phase of evaluating the taxonomy involved looking at the patterns of correlations between performance on taxonomic task groupings and the background variables previously discussed. These correlations are similar to the zero-order validity coefficients computed for the alternate task grouping adopted in previous sections of this report. However, the emphasis,

Table 19. Consensual Rating of Drivers' Tasks by Taxonomic Dimension

TASKS[2]	TAXONOMIC DIMENSIONS[1] PROCEDURAL	PERCEPTUAL	FEEDBACK	MOTOR	COMMUNICATION
1. Remove M27 Periscope	YES**		YES		
2. Checks/Services On Periscopes		YES**			
3. Install M24 Periscope	YES**				
4. Place M24 Periscope Into Operation	YES**		YES*		
5. Checks/Services On Gas Particulate Unit	YES*	YES**			
6. Start Tank Engine	YES**				
7. Checks/Services On Tank Engine And Oil Transmission Levels	YES**				
8. Place Tank In Motion	YES**		YES	YES	YES
9. Check Track Tension			YES	YES**	YES
10. Prepare-to-Fire Procedures	YES**				YES
11. Operate Tank In Neutral Steer			YES	YES**	
12. Cross Vertical Obstacle			YES**	YES**	YES
13. Cross A Ditch			YES**	YES**	YES
14. Ascend A Steep Grade	YES	YES	YES**	YES**	
15. Descend A Steep Grade	YES*	YES	YES**	YES**	
16. Drive Over A Water Obstacle	YES*	YES*	YES**	YES**	
17. Evasive Maneuver—Following Commands			YES**	YES**	YES
18. Evasive Maneuver—Own Initiative		YES**	YES*	YES*	
19. Drive Into Defilade Position			YES*	YES*	YES**
20. Drive In Response to Fire Command		YES*	YES**	YES**	YES
21. Drive To A Halt		YES*	YES**	YES**	YES
22. Lock Brakes During Stationary Engagement	YES**				
23. Acquire Targets		YES*			YES

[1]YES means that dimension was judged to be present in the task; **=major dominant dimension for task; *=minor dominant dimension for task.

[2]Task titles are abbreviated for convenience; full titles are included in Table 1.

Table 19. Consensual Rating of Loaders' Tasks by Taxonomic Dimension (cont'd.)

TASKS[2]	TAXONOMIC DIMENSIONS[1] PROCEDURAL	PERCEPTUAL	FEEDBACK	MOTOR	COMMUNICATION
1. Before-Operations Check On Engine And Transmission Oil	YES	YES**			YES
2. Check Track Tension		YES**			YES
3. Adjust Track Tension	YES**		YES*		
4. Check Operation Of M3 Heater	YES**	YES			
5. Prepare Tank For Boresighting	YES**				
6. Check Boresight Alignment Of Main Gun		YES**			
7. Boresight M219 Machinegun	YES**		YES		
8. Stow Main Gun Rounds In Tank	YES			YES**	YES
9. Perform Main Gun Prepare-to-Fire	YES**	YES			YES
10A. Load Main Gun-- Sabot Loaded	YES**				YES
10B. Load Main Gun-- Main Gun Not Loaded	YES**			YES*	YES
10C. Load Main Gun-- Sabot Loaded But Other Ammu. Required	YES**			YES*	YES
11. Ready Coax In Response To Fire Command	YES**				YES
12. Rotate Round In Main Gun Misfire	YES			YES**	
13. Unload Misfired Main Gun Round	YES**				YES
14. Action To Reduce Stoppage Of M219 Machinegun	YES**				
15. Unload M219 Machinegun	YES**				YES
16. Remove M219 Machinegun From Tank	YES**		YES		
17. Disassemble M219 Machinegun	YES**				
18. Inspect Machinegun		YES**			
19. Assemble Machinegun	YES**				
20. Check Operation Of Machinegun	YES		YES**	YES**	
21. Mount Machinegun On Tank	YES**		YES		
22. Load Mahcinegun	YES**				
23. Disassemble Main Gun-- Breech Block Removal	YES**		YES	YES	
24. Disassemble Main Gun-- Breech Block Disassembly	YES**				
25. Assemble Main Gun Breech Block	YES**				
26. Installation Of Main Gun Breech Block	YES**		YES	YES	

[1]YES means that dimension was judged to be present in the task; **=major dominant dimension for task; *=minor dominant dimension for task.

[2]Task titles are abbreviated for convenience; full titles are included in Table 1.

Table 19. Consensual Rating of Gunners' Tasks by Taxonomic Dimension (cont'd.)

TASKS[2]	TAXONOMIC DIMENSIONS[1]				
	PROCEDURAL	PERCEPTUAL	FEEDBACK	MOTOR	COMMUNICATION
1. Check Operation Of M3 Heater	YES**	YES			
2. Charge Manual Elevation System			YES**	YES	
3. Place Turret In Power Operation	YES**	YES*			YES
4. Prepare-To-Fire Procedure	YES	YES**			
5. Prepare Tank For Boresight					YES**
6. Prepare Gunner's Telescope For Operation	YES*	YES**	YES		
7. Prepare Gunner's Telescope—Daylight	YES*	YES**	YES		
8. Prepare Azimuth Indicator	YES**	YES*	YES	YES	
9. Operate Elevation Quadrant	YES	YES**	YES		
10. Boresight Gunner's Telescope	YES**		YES	YES	
11. Boresight Daylight Sight Of Periscope	YES**	YES	YES	YES	
12. Boresight IR Sight Of Periscope	YES**	YES*	YES	YES	
13. Boresight Tank Search Light	YES**	YES*	YES	YES	
14. Boresight Machinegun		YES**	YES	YES	
15. Zero Tank Main Gun	YES**	YES	YES*	YES*	
16. Zero Machinegun	YES**	YES	YES*	YES*	
17. Main Gun Engagement, Stationary		YES**	YES*	YES*	YES
18. Main Gun Engagement, Moving		YES*	YES**	YES**	YES
19. Coax Engagement		YES*	YES**	YES**	YES
20. Main Gun Engagement, Moving To A Halt, Stationary Target		YES**	YES*	YES*	YES
21. Action On Main Gun, Failure To Fire	YES**				YES*
22. Unload Misfired Main Gun	YES**			YES	
23. Acquire Targets		YES**			YES

[1]YES means that dimension was judged to be present in the task; **=major dominant dimension for task; *=minor dominant dimension for task (see text).

[2]Task titles are abbreviated for convenience; full titles are included in Table 1.

Table 19. Consensual Rating of Tank Commanders' Tasks by Taxonomic Dimension (cont'd.)

TASKS[2]	TAXONOMIC DIMENSIONS[1] PROCEDURAL	PERCEPTUAL	FEEDBACK	MOTOR	COMMUNICATION
1. Check Gas Particulate Unit	YES**	YES			
2. Prepare To Fire	YES**	YES			YES
3. Load And Clear Machinegun	YES**	YES	YES	YES	
4. Dismounting Machinegun	YES**				
5. Disassemble Machinegun	YES**				
6. Maintain Machinegun	YES**	YES			
7. Prepare Machinegun	YES**				
8. Assemble Machinegun	YES**				
9. Mounting Machinegun	YES**				
10. Prepare Tank For Boresighting					YES**
11. Prepare Rangefinder For Operation	YES**	YES	YES		
12. Determined Range To Target		YES**	YES		
13. Boresight Rangefinder With Main Gun	YES**	YES*	YES	YES	
14. Boresight Tank Searchlight	YES**	YES*	YES		
15. Boresight M85 Machinegun	YES**	YES*	YES		
16. Zero Main Gun	YES**				
17. Zero M219 Machinegun		YES**	YES		
18. Zero M85 Machinegun	YES**	YES*	YES	YES	
19. Acquire Targets		YES**			YES
20. Main Gun Engagement--T.C. Fires, Moving, Stationary		YES**	YES*	YES*	YES
21. Main Gun Engagement--Gunner Fires, Stationary, Stationary		YES			YES**
22. Main Gun Engagement--Gunner Fires, Moving, Moving		YES			YES**
23. Coaxial Machinegun Engagement--T.C. Fires, Moving, Moving		YES*	YES**	YES**	YES**
24. Coaxial Machinegun Engagement--Gunner Fires, Moving, Stationary		YES			YES**
25. Caliber .50 Engagement--T.C. Fires, Moving, Moving		YES*	YES**	YES**	YES
26. Main Gun Engagement--Gunner Fires, Moving, Stationary		YES			YES**

[1]YES means that dimension was judged to be present in the task; **=major dominant dimension for task; *=minor dominant dimension for task.

[2]Task titles are abreviated for convenience; full titles are included in Table 1.

Table 20. Descriptive Data: Performance for Taxonomic Groups

	Driver N ≈ 45		Loader N ≈ 40		Gunner N ≈ 49		Tank Commander N = 45	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
Procedural-Major	.82	.22	--	--	.84	.17	--	--
Procedural-All[1]	.81	.19	.76	.30	.84	.17	.94	.09
Perceptual-Major	.85	.20	--	--	.88	.13	.97	.07
Perceptual-All	.86	.16	.78	.32	.86	.14	.96	.07
Feedback-Major	.91	.10	.58	.50	.93	.12	.96	.07
Feedback-All	.89	.12	.63	.44	.92	.09	.97	.06
Motor-Major	.89	.12	.70	.34	.96	.08	.96	.07
Motor-All	.90	.12	.75	.33	.93	.08	.97	.06
Communication-Major	--	--	--	--	.94	.11	.96	.06
Communication-All	--	--	--	--	.98	.14	--	--

[1] "All" are all tasks with the dimension judged as major or minor dominant.

here is slightly different in that we were hoping to find different patterns of correlations for different taxonomic dimensions. This would demonstrate that the dimensions were in fact differentially predictable, and would add evidence for the construct validity of the taxonomy.

The correlations of performance on the task groupings and the background variables are shown by position in Table 21. Inspection of the table shows that grouping the tasks by dominant taxonomic dimension did indeed produce different patterns of correlations for different dimensions. The pattern of significant correlations differed from position to position, as one would expect since the specific tasks involved were different. Nevertheless, the background variables involved in the significant correlations were surprisingly similar, even across positions. With the exception of the gunners, variables involving task experience or practice (e.g., job time, duty time, Table VIII experience in position) tended to be involved in the significant correlations. Thus, the results of this first analysis show that the taxonomic dimensions operated as different dimensions in this application and that these dimensions tended to relate strongly to the experience and practice variables included in the predictor set.

The next analysis was designed to investigate the extent to which specific predictions about the effect of experience on performance in the task groupings were supported. For this phase, we returned to the congruence groupings of soldiers discussed previously (see Table 13). As has been stated, these divisions between groups of subjects were based on the amount of past and present experience at the job. In essence the groups represent different levels of potential for practice. The second phase of the evaluation took advantage of this variation in the soldier grouping to examine predictions about the taxonomic dimensions. The groups in each crew position were assigned ranks in the order in which they occur in the table. That is, the most experienced soldiers were given the highest rank and the least experienced were given the lowest rank. Correlations were computed between this variable and performance scores (proportion of steps GO) for each taxonomic category. These correlations are shown in Table 22. However, since correlations by themselves represent only a summary index of relationship, an additional step was undertaken: scatterplots of taxonomic category performance as a function of each congruence group were constructed. This produced a situation in which performance on each dimension could be visually evaluated over varying levels of practice. In this manner, the predictions about practice effects discussed earlier could be evaluated for each dimension.

Table 21: Correlations[1] for Taxonomic Grouping and Background Variables for Drivers

TASK GROUPINGS

		PROCEDURAL: MAJOR & MINOR	PROCEDURAL: MAJOR ONLY	PERCEPTUAL: MAJOR & MINOR	PERCEPTUAL: MAJOR ONLY	FEEDBACK: MAJOR & MINOR	FEEDBACK: MAJOR ONLY	MOTOR: MAJOR & MINOR	MOTOR: MAJOR ONLY
Personal Background	Weight	—	—	—	—	—	-.29*	—	—
	Height	—	—	—	—	—	-.27*	-.28*	-.30*
	Age	—	—	—	—	—	—	—	—
	Education	—	—	—	—	—	—	—	—
	Like Job	—	—	—	—	—	—	—	—
	Reenlist	.35*	.28*	.29*	.34*	.29*	—	—	—
Army & Job Background	Months Served	—	—	—	—	—	—	—	—
	Rank	—	—	—	—	—	—	—	—
	Months in Primary MOS	—	—	—	—	—	—	—	—
	Skill Level-Primary MOS	—	—	—	—	.32*	.26*	—	—
	Years in Crew	.30*	.30*	—	—	.40**	.31*	—	—
	Number Positions Held in Training	.37**	.42**	—	—	.33*	—	—	—
	Number Table VIIIs	.44***	.44***	—	.27*	.29*	—	—	—
Congruence	Months in Duty MOS	.33*	.27*	—	—	.29*	.32*	.27*	.33*
	Skill Level-Duty MOS	—	.29*	—	—	—	—	—	—
	Test Position-Job Title Match	.27*	.26*	—	—	—	—	—	.25*
	Position Training	.44***	.42**	.28*	.31*	.32*	—	—	—
	Position Table VIIIs	.42	.40	.27*	.28*	.35**	—	—	.27*
	Months in Position Past Year	.39**	.38**	.25*	—	.36**	.30*	.32*	.35**
	Hours Per Day in Position	.30*	—	—	—	—	—	—	—
	Platoon Sergeant Ratings	—	—	—	—	—	—	—	—

1. Significance levels are indicated as follows:

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Pearson product-moment correlations, rounded to the nearest hundredth.

Table 21. Correlations[1] for Taxonomic Grouping and Background Variables for Loaders (cont'd.)

		TASK GROUPINGS							
		PROCEDURAL: MAJOR & MINOR	PROCEDURAL: MAJOR ONLY	PERCEPTUAL: MAJOR & MINOR	PERCEPTUAL: MAJOR ONLY	FEEDBACK: MAJOR & MINOR	FEEDBACK: MAJOR ONLY	MOTOR: MAJOR & MINOR	MOTOR: MAJOR ONLY
Personal Background	Weight	SAME AS PROCEDURAL: MAJOR ONLY	—	SAME AS PERCEPTUAL: MAJOR ONLY	—	—	—	—	—
	Height		—		—	—	—	—	—
	Age		—		—	—	—	—	—
	Education		—		—	-42*	-38*	-30*	-35*
	Like Job		—		—	—	—	—	—
	Reenlist		—		—	—	—	—	—
Army & Job Background	Months Served		—		—	—	—	—	—
	Rank		—		—	—	—	—	—
	Months in Primary MOS		—		—	—	—	—	—
	Skill Level-Primary MOS		59***		54***	—	—	50***	49***
	Years in Crew		44**		43**	—	—	—	—
	Number Positions Held in Training		54***		53***	—	—	48***	49***
	Number Table VIIIs		39**		36**	—	—	—	—
Congruence	Months in Duty MOS		—		—	—	—	—	—
	Skill Level-Duty MOS		53***		48**	—	—	42**	42**
	Test Position-Job Title Match		33*		32*	—	—	38**	38**
	Position Training		46***		51***	28*	28*	38**	39**
	Position Table VIIIs		34**		35**	—	—	—	—
	Months in Position Past Year		44**		42**	—	—	29*	—
	Hours Per Day in Position		42**		33*	49**	55***	45**	50***
	Platoon Sergeant Ratings		—		—	—	—	—	—

1. Significance levels are indicated as follows:

*p < .050
**p < .010
***p < .001

Pearson product-moment correlations, rounded to the nearest hundredth.

Table 21. Correlations[1] for Taxonomic Grouping and Background Variables for Gunners

		TASK GROUPINGS									
		Procedural: Major & Minor	Procedural: Major Only	Perceptual: Major & Minor	Perceptual: Major Only	Feedback: Major & Minor	Feedback: Major Only	Motor: Major & Minor	Motor: Major Only	Communication: Major & Minor	Communication: Major Only
Personal Background	Weight	--	--	-30*	-36**	--	--	--	--	--	--
	Height	--	--	--	-29*	--	--	--	33*	--	--
	Age	--	--	--	--	--	--	--	--	--	--
	Education	--	--	--	--	--	--	--	--	--	--
	Like Job	--	--	--	--	--	--	--	--	--	--
	Reenlist	--	--	--	--	--	--	--	--	--	--
Army & Job Background	Months Served	26*	25*	26*	--	--	--	--	--	--	--
	Rank	--	--	--	--	--	--	--	-30*	--	--
	Months in Primary MOS	--	--	--	--	--	--	--	--	--	--
	Skill Level-Primary MOS	--	--	--	--	--	--	--	--	--	--
	Years in Crew	--	--	--	--	--	--	--	--	--	--
	Number of Positions Held in Training	--	--	--	--	--	--	--	--	--	--
	Number Table VIIIs	--	--	--	--	--	--	--	--	--	--
Congruence	Months in Duty MOS	--	--	--	25*	--	--	--	--	--	--
	Skill Level-Duty MOS	--	--	--	--	--	--	--	--	--	--
	Test Position-Job Title Match	--	--	--	--	--	--	--	--	--	--
	Position Training	--	--	--	--	--	--	--	--	--	--
	Position Table VIIIs	--	--	--	--	--	--	--	--	--	--
	Months in Position Past Year	--	--	24*	28*	--	--	--	--	--	--
	Hours Per Day in Position	--	--	--	--	--	--	--	-26*	--	--
	Platoon Sergeant Ratings	--	--	30*	31*	29*	29*	27*	29*	--	--

1. Significance levels are indicated as follows:

* $p \leq .05$.

** $p < .010$.

Table 21. Correlations[1] for Taxonomic Grouping and Background Variables for Commanders (Excluding Officers) (cont'd.)

TASK GROUPINGS

		Procedural: Major & Minor	Procedural: Major Only	Perceptual: Major & Minor	Perceptual: Major Only	Feedback: Major & Minor	Feedback: Major Only	Motor: Major & Minor	Motor: Major Only	Communication: Major & Minor	Communication: Major Only
Personal Background	Weight	SAME AS PROCEDURAL: MAJOR ONLY	--	--	--	--	--	--	--	SAME AS COMMUNICATION: MAJOR ONLY	--
	Height		--	--	--	--	--	--	--		--
	Age		--	--	--	--	--	--	--		-.33*
	Education		--	--	--	--	--	--	--		--
	Like Job		.31*	--	--	--	--	--	--		--
	Reenlist		--	--	--	-.37*	-.29*	-.37*	-.29*		-.39**
Army & Job Background	Months Served		--	--	-.25*	--	--	--	--		--
	Rank		--	--	-.29*	--	--	--	--		--
	Months in Primary MOS		.29*	--	--	--	--	--	--		.26*
	Skill Level-Primary MOS		--	--	-.28*	--	--	--	--		--
	Years in Crew		--	--	--	--	.29*	--	.29		.35**
	Number of Positions Held in Training		.27*	--	--	.35*	.38**	.35*	.38**		.30*
	Number Table VIIIs		--	--	--	--	--	--	--		.25*
Congruence	Months in Duty MOS		.34*	--	--	.23*	--	.27*	--		--
	Skill Level-Duty MOS		--	--	-.25*	--	--	--	--		--
	Test Position-Job Title Match		--	--	--	--	--	--	--		--
	Position Training		.29*	--	--	.32*	.29*	.32*	.29*		--
	Position Table VIIIs		.26*	--	--	.26*	--	.26*	--		--
	Months in Position Past Year		.34*	--	--	.27*	--	.27*	--		--
	Hours Per Day in Position		.43**	.44**	--	.48***	.43**	.48***	.43**		.47***
	Platoon Sergeant Ratings		--	--	--	--	--	--	--		--

1. Significance levels are indicated as follows:

* $p \leq .050$.

** $p < .010$.

*** $p < .001$.

Table 22. Correlations[1] Between Taxonomic Group Performance and Congruence Categories

	Driver n = 45	Loader n ≈ 33	Gunner n = 49	Tank Commander n ≈ 43
Procedural-Major	**46**	--	-05	--
Procedural-All	**45**	29	-02	10
Perceptual-Major	24	--	**34**	13
Perceptual-All	22	10	17	08
Feedback-Major	32	**45**	00	27
Feedback-All	**45**	**44**	05	24
Motor-Major	30	**39**	21	27
Motor-All	29	**40**	14	23
Communication-Major	--	--	-10	16
Communication-All	--	--	08	--

[1] Pearson product-moment correlations, rounded to nearest hundredth; decimals omitted. Correlations with significant levels $\leq .01$ are indicated in bold face.

Considering only the correlations, one finds that only the driver and loader data produced meaningful results. The tank commander performance data did not have enough variance to allow meaningful comparisons; performance at all taxonomic groupings over all levels of experience was clustered around perfect performance, thereby precluding significant correlations. For gunners, performance again did not appear to vary with experience. Only in the driver and loader data were the correlations significant, and thus consideration of the predicted practice effects is limited to these groups.

The plot for the drivers' procedural tasks matched predictions in that mean performance decreased and within-level variances increased with decreasing experience. The perceptual tasks showed a similar increase in within-level variance with decreasing experience but here mean performance was a U-shaped function over experience. Closer investigation showed that the U-shaped function stemmed from three groups of mechanics who were difficult to evaluate with respect to experience. If these groups were dropped performance would be as predicted. Plots for the drivers' feedback tasks showed that performance decreased slightly with decreasing experience and that within-level variance increased slightly with decreasing experience. Finally, the motor tasks showed almost constant mean performance and within-level variance across experience. This is exactly as predicted. In sum, the driver data showed clear support for the predicted practice effects.

The loader data were not quite as clear in support of the predictions. Mean performance on the procedural tasks was fairly constant across decreasing experience, with a relatively large within-level variance for some ranks and zero variance (one score) for other ranks. The perceptual tasks were much the same as the procedural tasks. The feedback tasks tended to produce "all-or-none" performance. In the high experience groups almost all subjects performed perfectly and in the low experience groups almost all subjects failed completely. The motor tasks showed large variances and tended to have fairly constant or slightly decreasing mean performance across ranks. In sum, the loader data clearly support the predictions only in the case of the motor tasks. For the taxonomic groupings, correspondence between predicted and observed performance is questionable at best.

Unfortunately, in neither the loader nor driver data is the analysis as strong as one would like. Both sets of data contain groups that were difficult to evaluate with respect to experience and groups that often contained only one or two persons. It is impossible to tell whether the failure to find predicted results is due to problems with the taxonomy or with the set of ranks.

Considering now the results of the two-phase evaluations of the taxonomy, we can conclude (1) that the dimensions of the taxonomy appear to be separate entities that differ from one another both in theory and in application, and (2) that with respect to predicted practice effects, there is evidence to suggest that the dimensions might operate as predicted but the tests are not conclusive. We are left with the conclusion that the taxonomy appears promising; there were enough interesting relationships to perhaps warrant further exploration of this approach. However, in this specific situation, the taxonomy did not provide the degree of organization (i.e. consistency) hoped for. As discussed below, this outcome might not be solely due to deficiencies in the taxonomy, but also to the problems inherent in the subject population and performance tests used in this study.

DISCUSSION AND CONCLUSIONS

Conceptually, the fundamental question addressed by this research effort is a simple one: are there demonstrable differences in combat effectiveness between soldiers whose normal, day-to-day activities are similar to their projected combat roles and those whose daily activities are not? However, a realistic restatement of this question makes it apparent that it is not as simple as it first appears: combat effectiveness depends upon many job-related factors. What are these factors, how are they related to each other, and how are they related to combat effectiveness?

A significant proportion of project activities was devoted to the translation of this proposition into a set of meaningful empirical hypotheses. In the course of this translation, numerous subsidiary issues arose that had to be resolved before more central questions could be addressed. Some of these subsidiary issues were resolved by fiat; others, due to limitations of time and resources, were resolved by "expert judgment"; and others were resolved empirically. A brief review of project activities will help to put these issues into perspective; it will also serve to summarize the major findings. We will then return to the central questions, presenting our "answers" and the implications of these answers for Army decision-makers.

Phase I. To reiterate what was discussed in the synopsis, the activities of this phase consisted basically of selecting the 11E MOS for examination and identifying a subject group, collecting a variety of information concerning their experience, confidence, and performance, and conducting several analyses designed to examine the relationships in the data. The key aspects of this phase relating to the central questions of the project were the delineation of the set of potential performance predictors, the decisions regarding the measurement of criterion performance, and the decisions regarding the analytical methodology used to determine the relationships between predictor and criterion variables. It was apparent that there was no simple way to characterize and organize all the potential influences on a particular soldier's combat readiness at any particular point in time, nor could one easily delineate those aspects of a soldier's previous experience and day-to-day activities most important for prediction. Thus, rather than restricting ourselves to a narrow and possibly infertile set of predictors, it was decided to collect a broad range of information concerning a soldier's personal background, his Army and job history, and his particular job experiences. Later analyses would be conducted to consolidate these variables into a parsimonious set. The same approach was taken for criterion performance variables: A large initial set of data was collected, with the plan being to later reduce the set to a meaningful (and manageable) performance index. The analytical and statistical techniques employed, while perhaps unnecessarily complex, enabled us to provisionally answer some of the central questions. Thus, it was found that level of performance was predictable for most soldiers, and that the hypothesized "congruence" variables were important predictors. However, the results also showed that different variables predicted performance for different crew positions and for different tasks within a given position. The attempt to unravel the complexities of these findings was the major focus of the second phase of the project.

Phase II. In essence, the activities of this phase consisted of consolidating the predictor and criterion variable sets, and conducting several types of analyses that attempted to clarify the relationships found between congruence and performance. As discussed in various sections of this report, the first of these activities produced several "byproducts":

1. The analyses of the interrelationships among the various criterion performance measures;
2. The analysis and ultimate incorporation of confidence ratings into the assessment of performance;
3. The development of two distinct organizational taxonomies for the performance tasks; and
4. The development of what is essentially a "congruence scale"--a method by which soldiers could be located and ordered along a single multidimensional scale.

The predominant "central" results are summarized at the end of the Congruence Effects section; hence, the following discussion will be of a more general nature, reflecting our current "position" on the congruence-performance issue.

The conclusion we are forced to draw from this study is essentially negative: we were unable to discover any systematic, generalizable connection between congruence and performance. We suspect that this is a function of two conditions that severely limited even the possibility of uncovering significant relationships. First, there were simply not enough "incongruous" soldiers in our sample. Especially for gunners and tank commanders, crew positions for the performance tests were overwhelmingly filled by "appropriate" soldiers. It was not the case, as was anticipated in the early stages of this project, that soldiers were engaged in widely disparate day-to-day activities. At least for the 11E MOS at Fort Carson during the time frame of this study, most soldiers were performing combat-related activities as part of their normal routine jobs.

Unfortunately, we do not know whether this phenomenon is typical of other MOS's or unique to this sample. Second, there was simply not sufficient variance in performance to permit reliable correlations with the predictor variables. Again, we do not know if such high performance levels are idiosyncratic for the particular group of soldiers in this study, nor can we determine whether it was due to the new performance test used. We would caution against assuming that the first alternative is true--that performance levels could be expected to be uniformly high throughout the Army. Nor are we willing to assume that the performance tests used here were inadequate samples of combat effectiveness. For future research concerned with this issue, it is possible that the implementation of the EPMS system, including wide scale utilization of SQT's, Soldier's Manuals, etc., could ameliorate the difficulties experienced in this effort.

On the other hand, a "positive" conclusion to be drawn from this study is that it seems possible to predict some soldier's combat effectiveness on a task-by-task basis, given a relatively small amount of information about the soldier's specific experience on the task. Furthermore, the soldier's confidence in his own ability might be an adequate surrogate for performance, and does not require the substantial investment of resources necessary to physically test his performance.