F1

The Role of Field Experimentation in Streamlining the Redesign of the U.S. Army in the 21st Century

Dr. Henry C. Dubin

ABSTRACT

The post Cold War era has brought four significant changes to the U.S. Army (1) transition from regionally based forces to predominantly home country based forces (2) considerable diversification in missions, operational environments and threats, (3) downsizing of the force, and (4) reduction in modernization budgets. Under these changes and constraints, the Army is challenged to develop versatile, lethal, easily deployable forces, by exploiting information technology and redesign concepts. To overcome the constraints and achieve the best possible combination of technological, procedural and organizational concepts, the U.S. Army is conducting a series of field experiments with operational units. This progression of warfighting experiments provides a venue for exploring warfighting concepts, learning strengths and weaknesses in implementing the concepts, and validating the operational effectiveness of new force designs. The paper describes the "rolling baseline" assessment strategy which updates the baseline with validated capabilities that serve as points of reference for the next cycle of proposed improvements. Tracking improvements in technologies, procedures and organizations over time enables the force designers to observe which capabilities are contributing the most to force effectiveness. The benefits of this approach are: (1) it provides a means of finding a balance between what is useful, what is achievable and the requirements for modernization, and (2) based on what soldiers can accomplish with the new concepts, leaders can make informed decisions on the design of the U.S. Army for the 21st century. Additionally, challenges in conducting such experiments and analyzing the results are addressed.

The Role of Field Experimentation in Streamlining the Redesign of the U. S. Army in the 21st Century

Dr. Henry C. Dubin

Background

The post-Cold War era has brought four significant changes to the U. S. Army. First, the Army is transitioning from regionally based forces to predominantly global projection forces. Second, the ground forces now face considerable variability and uncertainty in missions, operational environments and threats. Third there is a considerable downsizing of the force. Fourth, all these changes must be accommodated with a significant reduction in modernization budgets. In spite of these changes and constraints, the Army is challenged to develop a force which is:

- more lethal,
- more survivable,
- able to control the conditions and tempo of battle,
- easily deployable, and
- supportable in remote locations.

A key opportunity for achieving these improvements is to consider force redesign concepts which capitalize on exploiting information technology.

As a point of departure for introducing the Army's approach to meeting the above challenges, consider the traditional approach to force modernization. Battlefield deficiencies were first identified from combat modeling, analysis of threat capabilities, and combat or field experience. For those deficiencies that required a materiel solution, analyses were performed to determine the best of several alternative materiel concepts. With a concept in hand, proponents sought to develop and sell a modernization program to those empowered to grant resources. Competition for DoD funding tempted proponents to promise optimistic performance improvements in a short time frame and at a low cost.. Consequently, the approved programs were generally marginally resourced, had optimistic schedules, and pushed beyond the leading edge of technology. Cost and schedule overruns were common. Nevertheless, this flawed development process produced many fine warfighting and support systems.

Experimentation Concept

In today's environment of severely constrained modernization resources, we cannot afford cost overruns or programs that terminate without satisfying the materiel need. Therefore, the U. S. Army has adopted an approach to build confidence in new warfighting capabilities are technologically achievable and useful in the hands of soldiers before committing to acquisition programs. In this new framework, the Battle Laboratories in the Training and Doctrine Command are the repositories for new technologies. Five Battle Labs (Depth and Simultaneous

Attack, Early Entry, Mounted Battlespace, Dismounted Battlespace, Battle Command and Combat Service Support) are each responsible to improve capabilities in each of their respective battlefield functional areas. The Battle Labs use analysis and military judgment to screen new concepts and select those which are most likely to have a high payoff and can participate in field experiments in operational units. These experiments are called Advanced Warfighting Experiments or AWEs.

With the help of the technology base, acquisition, and Army test and evaluation organizations, the Battle Labs will conduct field experiments to explore strengths and weaknesses of the new technologies as well as investigate tactics, techniques and procedures for their use. Lessons learned from experiments at the small unit level (platoon or company) will be used to improve the prototype systems to facilitate their insertion in battalion or brigade-level experiments. If some concepts do not prove to be practical in the field they will be discarded or set aside for future consideration. Over the next five years, the Army intends to progressively advance the experiments through each level of command up to division and corps levels.

Whenever possible the AWEs will be conducted with units in the context of their normal training cycles. Occasionally experiments will be added to planned operational tests or if necessary will be scheduled as a dedicated exercise for experimentation purposes. The AWEs at battalion level or higher will be conducted in conjuction with planned training because the costs to deploy large forces only for experimentation are not affordable. Since the experiments are focused on warfighting capabilities, they are typically done in conjunction with force-on force training or testing exercises. An additional benefit to the trainers in these experiments is that the Army's Real Time Casualty Assessment instrumentation and data collection capabilities will enhance realistic combat conditions and provide valuable information for after action reviews.

The main advantages of field experiments for exploring new warfighting concepts are:

- trade-offs among technologies, organizational structures, and procedures that really work, can be identified;
- concepts or technologies that are not yet ready or practical for soldiers can be deferred for further maturation or discarded;
- soldier ingenuity can be applied to find better ways to use new technologies;
- concept exploration is based on demonstrated capabilities rather than assumptions used in analyses; and
- there is confidence that resulting systems acquisitions are likely to be achieved and are useful.

Implementation Challenges

In theory, the experimentation concept has great potential for reducing risks normally associated with modernization programs. A new program based on demonstrated capabilities in field experiments should have a reasonable chance of success. There are, however, significant pitfalls and challenges in conducting the experiments and in using the results.

Challenges in the Conduct of Experiments

Three issues must be addressed for the successful conduct of AWEs. The first is ***accommodating the differences between training and experimentation objectives.*** In general field training is conducted to assess or maintain the unit's collective warfighting skills. Consequently, training drills are repeated only when proficiency needs to be improved. Furthermore, when task or mission drills need to be repeated, the terrain and opposing force players that happen to be available are used. However, for valid experimental designs, conditions such as terrain and opposing forces should be held constant while other independent variables (such as alternative procedures or various prototype designs) are varied during repetitions of the same task or mission drills. Furthermore, the experimenters normally want to repeat the training drills with participating units that already have a high level of proficiency. Therefore, for training to be a viable venue for experimentation, agreements must be reached on scheduling and resourcing extra trials. If not, experimentation objectives will be sacrificed in favor of training objectives. The experiment proponents should be prepared to augment the training budget, and the trainers must be prepared to conduct extra trials or extend the normal training period. The operational units must accommodate new systems and non-standard procedures and tasks. Finally, the units must accommodate the time and inconvenience of instrumentation of their systems, training on instrumentation, instrumentation checkouts and calibrations prior to each trial. This is analogous to pre-combat inspections.

The second issue is the need to ***focus on a limited number of objectives.*** There is a great desire to introduce as many new gadgets and operational concepts as possible into an experiment. Unfortunately it is counterproductive to do so. Too many new things in an experiment puts a significant new equipment training burden on the experimentation units. It is difficult to achieve proficiency with many new capabilities in a relatively short time. Also, too many independent variables make it difficult to isolate the contributions of each new system in the conduct of combat missions. In addition to a limited number of trials, the outcomes of combat training missions typically have a large variance which exacerbates the problem for the analyst.

The final issue concerning the conduct of field experiments addresses ***early delivery of the new systems and the procedures for using them.*** Most unit training schedules are fixed well in advance, leading to a large scale event such as a joint exercise or a rotation to one of the Army's Combat Training Centers. Therefore it is extremely difficult to slip the date of the culminating event. New systems must be delivered early enough to be introduced to the units for new equipment training before the sequence of collective training exercises begin. This progression starts from a low level, such as platoon, to the high level, typically battalion or brigade for combat units. Ample train-up time is essential for the success of new operation concepts if the new systems and procedures are not mature enough for the soldiers to use reasonably well, both the training and experimentation objectives will be in jeopardy.

There are two steps than can mitigate this problem. The first is to plan for and adhere to a series of equipment checks. These checks should start with typical user validation of reasonable human interfaces with the equipment in early demonstrations or participation in the development of the interfaces. Then in a laboratory environment verify the system functionality and required interoperability with other systems. The final equipment check should occur in a field environment with the equipment installed on the platforms to be used in the experiment. The second step is to develop draft procedures for using the new systems. This needs to be done in time for the procedures to be taught as part of the new equipment training. Initial tactics, techniques and procedures are essential for getting the units started in their training program. Although the users will undoubtedly find new and better ways to employ the new equipment, it is not reasonable to simply give units new equipment early in their training cycle and expect wonderful improvements in combat effectiveness to occur.

A key to successful operational experiments is to exercise steadfast discipline in adhering to the cut-off dates for delivery of equipment and procedures to the experimentation units. The experiment proponents must stay focused on achieving new operational capabilities, not simply trying to get new gadgets in the hands of soldiers.

Challenges in Using the Results

In addition to the problems inherent in conducting operational experiments, there are related problems in using the results. In the broadest context, senior Army leaders want to use the AWEs to support decisions on the design of the land forces for the 21st Century. The Force XXI concept is to use the findings of several experiments to make decisions on new systems acquisition, force structures and concepts of operation. The experiments will ideally be progressive and will not repeat demonstrations of proven capabilities. A considerable challenge is found in **making decisions based on a disparate set of experiments conducted at different echelons, in different conditions, and with different objectives, over an extended period of time.**

To overcome this challenge, the Army established a framework for organizing the various findings from the AWEs. This framework is called the Rolling Baseline Assessment Strategy. Its four primary features are:

- a "rolling" baseline,
- common measures of force effectiveness,
- integration of constructive (combat models), virtual (interactive manned), and live

(operational field exercise) simulations, and

- a central database organized by simulation type, unit type and echelon.

The rolling baseline is central to this assessment concept. As the Army advances toward the design of Force XXI, this concept supports an assessment, at any time, of current force capabilities. It also documents trends of improved system performance and force effectiveness in support of decisions on new concepts, systems, and force structure.

The rolling baseline avoids the single event-oriented "success or failure" philosophy. It uses data from relevant preceding experiments as the baseline for the next cycle of analyses and experiments. It enables analysts to track changes in force effectiveness over time as new system technologies, doctrine, procedures, and force structures are introduced. In addition, as various experiments focus on different employment concepts, the idea of tracking force effectiveness improvements provides a means to assess which concepts have the greatest potential. If trade-offs among some of the force capabilities (e.g., less ease of deployability for greater survivability) occur, they can be identified and presented to the senior decision makers.

The common threads tying many of these disparate activities together are an overarching set of common force effectiveness measures. These measures are generic and can be applied to all echelons. Their purpose is to measure the degree of improvement in force effectiveness in the force capability areas of: lethality, survivability, controlling tempo (command and control), deployability, supportability, joint and combined interoperability, and versatility. These are the measures that are used to track progress in reaching the objective force design. These high level measures of effectiveness (MOE) are supported by lower level measures of performance (MOP).

There are two classes of MOP, operational MOP and system MOP. The operational MOP focus on battlefield functions directly dependent on procedural, organizational or materiel changes in the experimentation force. System MOP focus on capabilities of the new system technologies as they are used by the soldiers. The logical relationships between these measures are: the system MOP should indicate how the new systems enable improvements in the operational MOP, and the operational MOP should indicate how the new operational concepts enable improvements in the force level MOE.

For each cycle of improvements, the rolling baseline assessment strategy employs the "model-experiment-model-validate" method. The first phase of a cycle is normally constructive simulation to screen new candidate concepts. Based on the modeling analysis, goals for the cycle are established. These goals coupled with promising ideas from the Battle Labs are the basis for the experimentation objectives to be addressed in the next phase, experiment. In phase two of the cycle small scale experiments would be conducted in order to find out how well soldiers can implement the concepts with the prototype systems and procedures. Typically virtual simulations are used to explore initial procedural candidates, and live simulations are used to see how well operational concepts can be implemented by soldiers in the field environment.

The results of the experiments are then fed back into the constructive models in the third phase of the cycle. This modeling phase serves three purposes. First, the constructive, virtual and live simulation results can be compared, and adjustments can be made to any of the simulations as appropriate. Second, the constructive modeling provides feedback on whether or not the achieved capabilities have sufficient potential to yield acceptable improvements in force effectiveness. Phases two and three can be repeated to weed out the less useful concepts and refine the more promising ones. The third purpose of the phase three modeling is to establish the analytical baseline for the fourth phase, validation. In phase two the emphasis is on operational and system MOP. Phase three provides the transition to the MOE which are the focus of the validation phase.

The final phase of the cycle, validation, is normally accomplished with a large scale AWE. Here the objective is to see how well the new operational concepts and systems can be integrated to achieve improvements in force effectiveness. The efficacy of the force changes is assessed from three perspectives. First and foremost is the chain of command's assessment of the warfighting readiness of the experimentation force. Second is the Battle Labs' assessment of the doctrine, training, leader development, organization and soldier implications of the changes. Third is the positive trends in the MOE and the supporting correlations among the MOP and MOE. The substantiated improvements in the MOP and MOE are documented and update the rolling baseline for the next cycle of improvements. Supporting the model-experiment-model-validate method is the rolling baseline database. This database is structured to store the conditions, observations, findings, and MOP and MOE for each experiment and analytical study. This structure enables the analyst to explore correlations among measures, force effectiveness trends and trends of correlations over time by simulation type, unit type, echelon and mission.

The Rolling Baseline Assessment Strategy described above is intended to provide a framework for helping to support making decisions which guide an extremely complex force redesign process. It is instructive to compare the characteristics of this process with those of the classic scientific method to illustrate some of the significant analytical challenges posed. A simple description of the scientific method is outlined below.

1. Make observations about a phenomenon.
2. Formulate a theory.
3. Develop a method to test (i. e. challenge) the theory. This is normally done by manipulating one or more independent variables and observing whether or not the predicted state of a dependent variable is obtained.
4. Conduct the test to gather the appropriate data.
5. Accept or reject (or refine) the theory.

If the theory is deterministic, the decision on accepting or rejecting is straightforward. If the theory is stochastic, the decision has an element of risk associated with the statistical test used to accept or reject. If the theory is accepted we usually infer a causal relationship between the independent and dependent variables. By comparison the rolling baseline for AWEs is replete with problematic characteristics. Examples are, **very many variables, inability to precisely**

control any of the variables, many unobserved variables, and AWE objectives change for each repetition. Given this partial list of problems, there is little chance of having a situation which poses a decision based on known statistical risk and virtually no chance for a straightforward decision.

Even though we are conducting "field experiments," analytically we are relegated to the domain called "nonexperimental methods." We are stuck with predominantly passive observations and dependent on "natural" variations. What analytical objective would be appropriate given these circumstances? As a minimum we would hope to show positive correlation between the force design changes and the changes in force capabilities. Correlation methods are often suitable for forecasting, but they are extremely risky for inferring causality. To illustrate the point, suppose we find a strong positive correlation between attendance at advanced military schools and reaching the general officer ranks. One could infer that people selected for the advanced schools are more likely to become generals than those who are not selected for the schools. It would be a mistake, however, to infer that the advanced schooling caused people to be good general officers. It would be a bigger mistake to decide to send all soldiers to these schools to give them the capabilities of generals. It is likely that there are personal trait variables that result in both selection to the schools and the higher chance of becoming a general.

Therefore, if in fact we can find positive correlations in a series of AWEs, we must carefully guard against the mistake of inferring causality when none can be substantiated. We should not invest our limited resources in systems simply because they happened to be present when force effectiveness was observed to be better than when those systems were not present. It may be that some other factors resulted in the improved effectiveness. This is more likely to be the case when we deal with many simultaneous changes in systems, procedures, and organizational structures in an AWE. Worse yet would be a case where one would invest in a concept which actually had a negative influence in the context of an overall positive correlation. That is, the force would have done even better without the selected concept. These mistakes usually stem from unique variances in measured variables or the influence of unobserved variables.

The social sciences have long been confronted with such challenges. Although there are no foolproof methods for avoiding errors of improper inferences, there have been several techniques developed for the purpose of reducing the occurrence of such errors. Techniques such as path analysis and cross-lagged correlations should be considered when analyzing the correlations system MOP, operational MOP and the force effectiveness MOE. Other approaches, such as the Delphei Method or interrater correlations among subject matter experts, should also be considered for getting military judgment consensus among decision makers.

Wrap Up

In spite of the AWE conduct and analytical challenges, I believe that the advantages of learning from soldiers in operational exercises is well worth the investment in focused field experiments. If one wants to reduce the risk of having requirements for new systems which are not readily achievable or may not be very effective, focused field experiments are a viable tool for generating practical requirements. I assert that field experiments with operational units, when adequately planned and executed with the discipline to keep the process event driven, are superior to paper or computer only analyses for conducting concept exploration. If the experiments are focused on achieving new battlefield capabilities and not just on getting new technologies to work in the field, they can be used successfully to find the balance between requirements and what is reasonably achievable and useful. Lastly, I believe that along with organized findings from AWEs, it is essential to incorporate military judgment in the form of structured consensus building into the process for making decisions on the design of Force XXI.