

Planning for LVC Simulation Experiments

Casey L. Haase, Raymond R. Hill, Douglas Hodson

Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, USA
Email: rayrhill@gmail.com

Received 12 May 2014; revised 15 June 2014; accepted 23 June 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The use of Live, Virtual and Constructive (LVC) simulations are increasingly being examined for potential analytical use particularly in test and evaluation. In addition to system-focused tests, LVC simulations provide a mechanism for conducting joint mission testing and system of systems testing when fiscal and resource limitations prevent the accumulation of the necessary density and diversity of assets required for these complex and comprehensive tests. LVC simulations consist of a set of entities that interact with each other within a situated environment (*i.e.*, world) each of which is represented by a mixture of computer-based models, real people and real physical assets. The physical assets often consist of geographically dispersed test assets which are interconnected by persistent networks and augmented by virtual and constructive entities to create the joint test environment under evaluation. LVC experiments are generally not statistically designed, but really should be. Experimental design methods are discussed followed by additional design considerations when planning experiments for LVC. Some useful experimental designs are proposed and a case study is presented to illustrate the benefits of using statistical experimental design methods for LVC experiments. The case study only covers the planning portion of experimental design. The results will be presented in a subsequent paper.

Keywords

DOE, Live-Virtual-Constructive (LVC), Simulation, Statistical Experimental Design

1. Introduction

Live, virtual, and constructive (LVC) simulation is a test capability being pursued by the Department of Defense (DoD) to test systems and system of systems in realistic joint mission environments. The DoD was made acutely aware of the need for designing and testing systems in a joint environment during the first joint operations conducted in Operation Desert Storm. Operation Desert Storm highlighted a host of interoperability issues, namely that systems across services were incompatible with one another [1]. The Secretary of Defense (SECDEF) re-

sponded by mandating a new capability-based approach to identify gaps in services' ability to carry out joint missions and fill those gaps with systems designed with joint missions in mind [2]. Additionally, the SECDEF mandated that all joint systems be tested in a joint mission environment so that systems can be exercised in their intended end-use environment. This implies that future testing of systems be capability focused [3].

In response to the SECDEF's mandate, the Director of Operational Test and Evaluation (DOT & E) set up the Joint Test Evaluation Methodology (JTEM) project. The purpose of JTEM was to investigate, evaluate, and make recommendations to improve test capability across the acquisition life cycle in realistic joint environments. One result of JTEM's efforts was the development of the capability test methodology (CTM). CTM is a set of "best practices" that provide a consistent approach to describing, building, and using an appropriate representation of a joint mission environment across the acquisition life cycle. The CTM enables testers to effectively evaluate system contributions to system-of-systems performance, joint task performance, and joint mission effectiveness [3].

CTM is unique in that it focuses not only on the materiel aspects of the system but also on aspects of doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF). The inclusion of these joint capability test requirements adds significant complexity to the T&E process. Because of this increase in complexity, the CTM Analyst Handbook states that future tests will require innovative experimental design practices as well as the use of a distributed LVC test environment to focus limited test resources [3].

LVC is a central component of CTM due to its ability to connect geographically dispersed test facilities over a persistent network and potentially reduce test costs. Figure 1, from the CTM Handbook [3], illustrates the centrality of LVC to CTM. LVC simulations can scale to different levels of fidelity thus making it well suited to experiments across the acquisition life cycle. Simple joint mission environments can be developed using mostly constructive entities in the early stages of system development with live and virtual entities added as the system matures. Cost is yet another reason that LVC is being pursued as a core test capability. While the cost of LVC experiments can be significant, it is often less expensive when compared to joint mission experiments using only live assets. Furthermore, LVC simulation can build joint mission scenarios of greater complexity than can be assembled at any single DoD test facility.

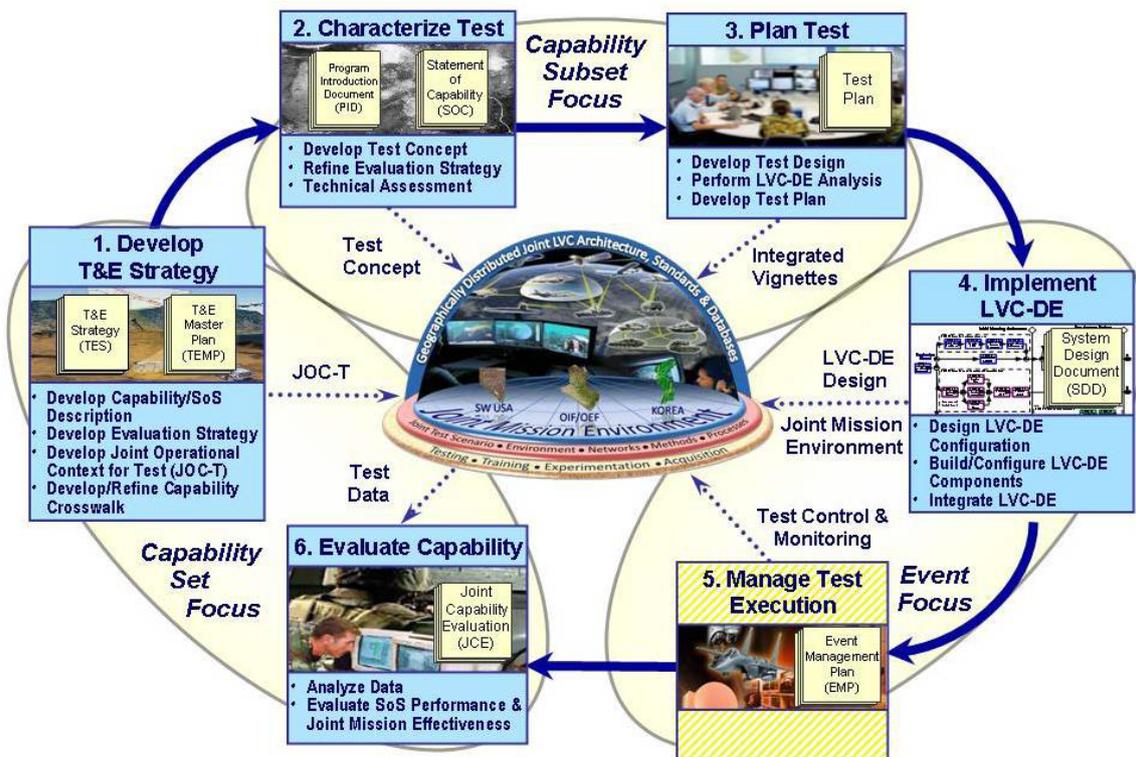


Figure 1. Capability test methodology [3].

1.1. Live-Virtual-Constructive Simulation

LVC simulations consist of a set of entities that interact with each other within a situated environment (*i.e.*, world) each of which are represented by a mixture of computer-based models, real people and real physical assets [4]. The physical assets often consist of geographically dispersed test assets which are interconnected by persistent networks and augmented by virtual and constructive entities to create the joint test environment for evaluation. Because they include live, virtual and constructive entities, LVC simulation technology is able to create the necessary variety and density of assets representative of a joint environment.

When viewed as a software system, LVC simulations create an environment where multiple, geographically dispersed users interact with each other in real-time via a persistent network architecture [5]. LVC simulations consist of a set of entities from three DoD defined classes of simulations: live, virtual, and constructive. In a live simulation, real people operate real systems. A pilot operating a real aircraft for the purpose of training under simulated operating conditions is a live simulation. In a virtual simulation, real people operate simulated systems or simulated people operate real systems. A pilot in a mock-up cockpit operating a flight simulator is a well-known example of virtual simulation. In constructive simulations, simulated people operate simulated systems.

LVC simulations have the potential to provide experimenters with several benefits not found in purely live system tests. First, systems can be tested in robust joint environments at a fraction of the cost of using only live assets. Test ranges, threats, emitters, and conceptual next-generation capabilities can be included in the simulation without purchasing the live asset. These assets are expensive and their specific inclusion could significantly increase the cost of a test program using just a live system test. The reduced cost of LVC experiments can sometimes allow for more runs and consideration of more design factors when cost is the limiting resource. More runs using LVC can result in more information than could be obtained in a similar test only utilizing live assets.

The virtual and constructive elements of LVC give experimenters increased flexibility in designing the experiment. Statistical experiments are founded on completely randomizing the order of the experiments. Split-plot designs provide approaches when complete randomization is restricted. In some situations completely randomized designs can be used in the LVC instead of the more complex split-plot designs often found in live test because the virtual and constructive elements can be easily reconfigured before each run. An important caveat is to use caution when changing virtual and constructive elements if humans are active in the experiment; changing test conditions too often can lead to operator confusion and introduce bias in the results.

Another benefit of LVC is that it allows the user to exercise greater control over the test environment. Increased control improves the repeatability of the experiment potentially increasing the precision of the estimate of the experimental error used when making statistical statements regarding the results. Reduced experimental error also means more precise effect estimates for the active factors in the experiment. With the exception of live assets, all entities in the simulation experiment can be controlled with greater precision which allows the analyst to scale the fidelity of the model as needed to suit the experimental objective.

The LVC environment is also fairly easy to instrument. This provides an improved capability to gather data to support decisions pertaining to the test objectives. The design team does, however, need to spend time evaluating potential measures and implementing only those needed.

1.2. Change the LVC Paradigm

The LVC concept was introduced to the DoD by the Joint National Training Center, which was established in January 2003 to provide war fighters across all services training opportunities in a realistic joint mission environment [6]. In a training environment, large, complex, noisy environments are preferred because it appropriately prepares soldiers for the “fog of war”. Further, training outcomes do not always require quantitative-based, objective results. For analytical purposes such as test, “fog” is a detriment because it obscures the underlying factors that are driving system performance and effectiveness. In test, we want to abstract away certain aspects of the representative environment so that we can identify the factors that affect the system in its end-use environment. If LVC is going to be successfully implemented as a core test capability LVC practice will require a fundamental shift from the way LVC users currently employ the technology and towards a paradigm in which the LVC generates quantitative-based, analytically defensible results.

If LVC simulation is properly utilized it offers significant test capability to T & E practitioners. Care must be taken to ensure that users understand the limitations of LVC or risk collecting meaningless data. Statistical experimental design techniques greatly increase the likelihood of collecting useful data and doing so in an efficient

manner. Statistical experimental design is a methodical design process that plans, structures, conducts, and analyzes experiments to support objective conclusions in complex test environments. Statistical experimental design gives experimenters a firm foundation for conducting LVC experiments but its use represents a fundamental shift in how LVC is used currently. In Section 2 we give an overview of the experimental design process and a summary of designs useful for LVC. In Section 2.2 we discuss additional considerations for conducting experiments with LVC. Lastly, a case study is presented to illustrate the benefits of experimental design for LVC experiments in Section 4.

2. The Statistical Experiment Design Process

Experimental design is a strategy of experimentation to collect and analyze appropriate data using statistical methods resulting in statistically valid conclusions. Statistical designs are quite often necessary if meaningful conclusions are to be drawn from the experiment. If the system response is subject to experimental errors then statistical methods provide an objective and rigorous approach to analysis. Often in test, the system response is measured as a point estimate (such as the mean response) when the individual responses are actually subject to a random component. Oversimplifying the system response can often lead to erroneous conclusions because the random component of the response is unaccounted for.

The three basic principles of statistical experimental design are randomization, replication, and blocking [7]. Randomization is the cornerstone of statistical methods. Statistical methods require that the run-to-run experimental observations be independent. Randomization typically ensures that this assumption is valid. Randomization also spreads the experimental error as evenly as possible over the entire set of runs so that none of the effect estimates are biased by experimental error. A replication is an independent repeat of each factor combination and provides two important benefits to experimenters. Replication provides an unbiased estimate the pure error in an experiment. This error estimate is the basic unit of measurement for determining whether observed differences in the data are statistically different. More precise effect estimates is another benefit of replication. In general, the more times an experiment is replicated the more precise the estimates of error will be and any inferences pertaining to factor effects will be more informed.

Blocking is a design technique that improves the precision of estimates when comparing factors. Blocking controls the variability of nuisance factors; factors that influence the outcome of the experiment but are not of interest in the experiment. To illustrate blocking, consider a machining experiment where two different operators are used in the experiment. The operators themselves are not of interest to the experiment but experimenters are concerned that any differences between the operators may confound the results and lead to erroneous conclusions. To overcome this, the operators are assigned to two separate blocks of test runs. By assigning the operators to blocks any variability between operators can be estimated and those effects removed from the experimental error estimates, thus increasing overall experiment precision.

A statistical experiment design process for LVC must not only consider the three basic principles of statistical experiment design, but also include considerations such as:

- Models, simulations and assets used in the experiment;
- Scenarios considered during the experiment;
- Factors that change each run and how to control those that do not change;
- The fidelity of models and simulations used; and
- How human operators might influence results.

The above complications truly call for an LVC experimental design process.

2.1. An Experimental Design Process

To apply statistical methods to the design and analysis of experiments, the entire test team must have a clear understanding of the objectives of the experiment, how the data is to be collected, and a preliminary data analysis plan prior to conducting the experiment. [8] proposes guidelines to aide in planning, conducting, and analyzing experiments. An overview of their guidelines follow, keep in mind these guidelines pertain only to the development of the experimental plan, not the myriad of other factors that arise when planning and coordinating the resources for actual experiments. These guidelines are useful for defining an LVC-experiment design process.

- 1) **Recognition and statement of the problem.** Every good experimental design begins with a clear statement

of what is to be accomplished by the experiment. While it may seem obvious, in practice this is one of the most difficult aspects of designing experiments. It is no simple task to develop a clear, concise statement of the problem that everyone agrees on. It is usually necessary to solicit input from all interested parties: engineers, program managers, manufacturer, and operators. At a minimum a list of potential questions and problems to be answered by the experiment should be prepared and discussed among the team. It is helpful if not necessary to keep the objective of the experiment in mind. Some common experiment objectives are given in **Table 1**.

At this stage it is important to formulate large problems into a series of smaller experiments each answering a different question about the system. A single comprehensive experiment often requires the experimenter to know the answers to many of the questions about the system in advance. This kind of system knowledge is sometimes unlikely and the experiment often results in disappointment. If the experimenters make incorrect assumptions about the system, the results could be inconclusive and the experiment wasted. A sequential approach using a series of smaller experiments, each with a specific objective, is a superior test strategy.

- 2) **Selection of the response variable.** The response variable measures system response as a function of changes in input variable settings. The best response variables directly measure the problem being studied; they provide useful information about the system under study as it relates to the objectives of the experiment. Test planners need to determine how to be measure response variables before conducting the experiment. The best response variables directly measure the problem being studied. Sometimes a direct response is unobtainable and a surrogate measure must be used instead. When surrogate measures are used test planners must ensure that the surrogate adequately measures how well the system performs related to the objectives and the system is properly instrumented to capture the surrogate measure information.
- 3) **Choice of factors, levels, and range.** Factors are identified by the design team as potential influences on the system response variable. Two categories of factors frequently emerge: design and nuisance factors. Design factors can be controlled by either the design of the system or the operator during use. Nuisance factors affect the response of the system but are not of particular interest to experimenters. Often nuisance factors are environmental factors. Blocking is a design technique that can be used to control the effect of nuisance factors on an experiment. For more details on techniques that deal with nuisance factors see [7].

After choosing the factors it is necessary to choose the number of levels set for each factor in the experiment. Quantitative factors with a continuous range are usually well represented by two levels but more levels often arise in the more complex, comprehensive designs. When factors are qualitative the number of levels is generally fixed to the number of qualitative categories. Unlike continuous factors, there is no way to reduce the number of factor levels for categorical factors without losing the ability to make inferences on that level's effect on system response. The range of factors levels must also be carefully considered in the design process. Factor levels that are too narrowly spaced can miss important active effects while factor levels that are too wide can allow insignificant effects to drive the system response. A subject matter expert working in conjunction with the statistical experimental design expert is invaluable when choosing the range of factors levels to ensure sufficient detectability of effects.

- 4) **Choice of experimental design.** Choosing an experimental design can be relative easy if the previous three steps have been done correctly. Choosing a design involves considering the sample size, randomizing the run order, and deciding whether blocking is necessary. Software packages are available to help generate alternative designs given the number of factors, levels, and number of runs available for the experiment. More unique designs like orthogonal arrays and nearly orthogonal arrays can be created with available computer algorithms. Some good resources for creating unique designs are given in Section 3.

Table 1. Common objectives for experiments.

Objective	Type of System	Rationale for Usage
System Characterization	New System	Little understanding how control variables affect system response
Optimization	Mature System	Seek control settings for best system response performance
Robustness	Mature System	Seek control settings to reduce system response variation from noise

- 5) **Performing the experiment.** In this step it is vital to ensure that the experiment is being conducted according to plan. Conducting a few trial runs prior to the experiment can be helpful in identifying mistakes in planning thus preventing a full experiment from being wasted. While tempting, changing system layouts or changing factors during the course of an experiment, without considering the impact of those changes, can doom an experiment.
- 6) **Statistical analysis of the data.** If the experiment was designed and executed correctly the statistical analysis need not be elaborate. Often the software packages used to generate the design help to seamlessly analyze the experiment. Hypothesis testing and confidence interval estimation procedures are very useful in analyzing data from designed experiments. Common analysis techniques include analysis of variance (ANOVA), regression, and multiple comparison techniques. A common statistical philosophy is that the best statistical analysis cannot overcome poor experimental planning. The important aspect of statistical analysis is to involve the professional statistician for the analysis.
- 7) **Conclusions and recommendations.** A well designed experiment is meant to answer a specific question or set of questions. Hence, the experimenter should draw practical conclusions about the results of the experiment and recommend an appropriate course of action. The beauty of a well designed and executed experiment is that once the data have been analyzed the interpretation of the data should be fairly straightforward, objective and defensible.

Coleman and Montgomery [8] give details on the steps of experimental design. Additionally, most texts on experimental design, including [7], provide some experimental design methodology.

2.2. Additional Design Considerations for LVC

The Coleman and Montgomery [8] guidelines offer comprehensive general guidelines for industrial experiments. However, LVC experiments are non-industrial representing a more dynamic process. There are several experimental design issues that need to be addressed before the benefits of LVC can be fully realized.

- 1) **Scoping the Experiment.** Scoping LVC experiments require more careful treatment than most traditional experiments. LVC is flush with capability; users and experimenters can build very large, complex, joint mission environments. Experimenters are often enticed to create environments that are more complex than required to actually satisfy the experiment's objective. When these LVC environments are used for analytical purposes, such as the case in T&E, more discipline must be exercised to ensure the test environment is not overbuilt but remains constructed to align with the analytical objectives. LVC has enormous data generation capability making the number of possible problems that can be researched significantly larger than that of live asset tests. An LVC builder can instrument just about any process included in the environment. Experimenters are faced with vast alternatives to choose from when designing the experiment. This means planners have to avoid the temptation to grow the size of the experiment in an attempt to answer all questions, and stay focused on investigating those that are most important.

Over-scoping the experiment not only affects the quality of data garnered from the experiment but also leads to delays in experiment execution. LVC simulation developers work off of the requirements supplied by the test team; if too many requirements are demanded then developers can become task saturated and unable to deliver the LVC environment in time for the test event. Breaking the experiment up into a series of smaller experiments that build on each other can improve the experiment data quality and increase the likelihood of meeting test deadlines. When used for training or assessments, increased complexity in the LVC environment has become accepted. When used for analytical insight, this same increased complexity can ruin any meaningful results.

- 2) **Qualitative Objectives.** Objectives in LVC experiments are often qualitative in nature. LVC is used primarily for joint mission tests to evaluate system-of-systems performance, joint task performance, and joint mission effectiveness. Nebulous qualities such as task performance and mission effectiveness are often difficult to define and measure. More often than not there are no direct metrics to quantify system performance and mission effectiveness. Questionnaires and opinions are often used. Consequently choosing an appropriate response variable is not straightforward. Surrogate measures need to be circumspectly examined to make certain that the experiment objectives are actually measured. This may actually require some innovative thinking on the part of the design team to build instrumentation into the LVC environment to gather the data necessary to support otherwise qualitative assessments of system performance in a system-of-systems context.

- 3) **Mixed Factor Levels and Limited Resources.** Joint mission environments are complex often containing many mixed-level, qualitative factors with scant resources available. Mixed-level factors refer to multiple factors where at least one factor contains a differing number of levels than the other factors. Often mixed-level designs require a large sample size making them inappropriate for tests that demand a small sample size due to resource constraints. Mixed-level designs can be fractioned into smaller designs but doing so can be tedious and independent estimates are not guaranteed for all fractioned designs. For the LVC experiment planners, early consideration of these mixed factor problems can lead to changes in experiment focus, objectives, or even design to accommodate the problem.
- 4) **Interaction Effects.** Unlike most traditional experiments, large simulation experiments can have a significant number of higher order interaction effects (*i.e.*, 3-way or higher factor interactions). When using small designs these higher order effects may be aliased with the main effects meaning that the source of the effect is difficult, if not impossible to isolate and estimate (the main effect and interaction effect are intermingled). Active higher order interactions can wreck the outcome of the experiment unless they are considered and appropriately accounted for in the choice of experimental design. The multi-disciplinary experiment design team can anticipate these interactions and choose designs for the LVC experiment that avoid the aliasing problem.
- 5) **Noisy Test Environments.** The joint mission environment contains copious sources of noise that must be prudently considered. Noise in the test environment can be harmful to an experiment if appropriate measures are not taken to control it or measure it. Effects that are thought to be important may not appear to be so because of over-estimated experimental error. To overcome this problem appropriate statistically-based noise control techniques are used in the LVC experiment planning process. Often human operators are the largest contributors of noise in the experiment and thus should only be used as necessary in LVC experiments. The benefits or necessity of including human subjects in the experiment must outweigh the risk that is assumed by including them. This judicious use of the human component in the LVC experiment is likely one of the larger paradigm shifts when moving LVC from a training environment to an analytical environment. Increasing system complexity by integrating additional (possibly unnecessary) assets can also increase noise in test.
- 6) **Human System Integration.** HSI principles should be applied to LVC experiments since LVC is a software system that requires extensive human interaction. Madni [9] states that HSI practices propose that human factors be considered an important priority in system design and acquisition to reduce life-cycle costs. Furthermore, he states that each of the seven HSI considerations is necessary to satisfy operational stakeholders needs. We would add that HSI principles should be applied across all T&E activities where humans interact with software systems and offer some HSI considerations for T&E when human-software system interaction is central to the experiment, as is often the case with LVC. HSI considerations for LVC-based T&E activities ensure that:
 - a) The right tradeoffs have been made between the number of humans included in the experiment and the quality of data required.
 - b) Including joint human-machine systems in the experiment supports the objectives with human-machine systems only included when the experiment's analytical requirements can still be satisfied.
 - c) The design of the experiment circumvents the likelihood of excessive experimental error caused by human-machine systems by using appropriate experimental noise control techniques.
 - d) Data planning and analysis takes into account the additional variability introduced when humans adapt to new conditions or respond to contingencies (e.g., consider and avoid human learning invalidating the experimental results).

Human System Integration is native to the systems engineering process from a design point of view but foreign to T&E activities. For LVC experimentation to be effective, HSI considerations must be included across all test planning activities; such HSI considerations for LVC experimental planning are left for future research.
- 7) **Improved Test Discipline.** An LVC environment is extremely flexible. Assets can be added, deleted or modified, in some cases, quite easily. Given its strong history in training and demonstration events, LVC experimenters often "tweak" the LVC based on early results. Changing the LVC system mid-way through a randomized experimental design changes the fundamental assumptions of subsequent experiments from those already completed. In other words, the experimental design is compromised and no amount of statis-

tical analysis can save poor designs.

- 8) **Experimental Design Size.** Unfortunately, there may be the belief that large, complex LVC experiments can answer any questions pertaining to the system (or systems) of interest. While the LVC may seem to address such questions, answering quantitatively those questions would require far too many experimental runs; LVC experiments have run budgets like any other experimental event. Fortunately there are a range of reduced sample size experimental designs quite applicable to LVC experimentation. Some are fundamental, usually covered in basic training guides. Others are more advanced but powerful in their ability to obtain meaningful results. The experimental design goal is how to achieve compromise in the final design choice to achieve sufficient coverage of test objectives while achieving desired levels of statistical power in the design as a function of some agreed to level of significance.

3. Some Useful Experimental Designs for LVC Applications

The LVC environment offers many unique capabilities to T&E. However, to use LVC results in the analytically rigorous manner required by T&E necessitates that experimental designs be scrutinized to ensure they satisfy the objectives of the LVC-based joint mission tests. Several advanced designs seem well suited to the LVC test environment: orthogonal arrays, nearly orthogonal arrays, optimal designs, and split-plot designs. The first three designs can be used in experiments that allow full randomization while the split-plot designs are useful when there are restrictions on randomization.

An array is considered orthogonal if every pair of columns in the array is independent. This is accomplished by making each level combination in each column occur equally often [10]. Orthogonality improves our ability to estimate factor effects. To illustrate the usefulness of OAs, consider an experiment with a three-level factor and four two-level factors where testing resources only allow for 12 runs. A full factorial (all combinations of all factor levels) design requires 48 runs (3×24) and fractioning the design into a smaller, useful design would be very complicated. An orthogonal array can be constructed with 12 runs and will generate independent estimates of each of the 5 main effects. Table 12.7 in [11] contains many mixed-level orthogonal arrays for the interested reader.

At times orthogonal arrays cannot sufficiently reduce the run size while accommodating the necessary number of factors. A design team can relax the orthogonality requirement and reduce the experiment run size through the use of a nearly orthogonal array. A drawback to nearly orthogonal arrays is that the estimates of the effects are somewhat correlated (*i.e.*, loss of independence when orthogonality was relaxed) making the data analysis somewhat more difficult [10]. Several researchers such as [12]-[14] have constructed nearly orthogonal arrays using algorithmic approaches with nice results.

Optimal designs are another excellent way to construct mixed-level designs. Optimal designs are nearly orthogonal designs optimized to some design criterion. Statistical software packages help create optimal designs making them a convenient choice for experimenters faced with mixed-level factors and limited resources. The D-optimal criterion (arguably the most widely used) measures the overall degree of orthogonality of the design matrix. The G-optimal criterion measures the extent that the maximum prediction variance for regression parameters is minimized. The G-optimal criterion is useful if a regression model is built from the experimental data to be used to make predictions about the system response. There are other optimal designs but not as pertinent to LVC experimentation in our view (see [15] for a cursory introduction to these other designs).

Split-plot designs are used when there are restrictions on experiment run randomization that prevent the use of a completely randomized design. Randomization restrictions make a completely randomized design inappropriate and can lead the experimenter to erroneous conclusions if the responses are analyzed in a manner inconsistent with the design and execution of the experiment [16]. In split-plot designs, hard-to-change factors are assigned to a larger experimental unit called the whole plot while all other factors are assigned to the subplot. Each of the whole plot and subplot carry an error component that must be estimated. Split-Plot designs are thus more difficult to analyze than completely randomized designs because of this more complicated error structure. See [16] for more details on split-plot designs.

There are of course many other classes of designs that may be applicable to LVC experimentation for T&E. The three classes discussed above provide, in our opinion, a broad range of options the LVC experimental design team should consider. Final design choices must be appropriate to the specifics of the LVC experiment planned. Use of orthogonal and nearly-orthogonal array designs is discussed in the subsequent case study.

4. Conducting a Data Link Experiment with LVC¹

Currently there are aircraft that can only receive Link-16 communications from Command and Control (C2) assets in denied access environments. The Multifunctional Advanced Data Link (MADL) is a technology that would allow aircraft to transmit to other friendly forces in a denied access environment. The Air Force Simulation and Analysis Facility (SIMAF) was tasked with assessing the suitability of the MADL data link for aerospace operations in a denied access environment using a distributed LVC environment [17]. The experiment will connect two geographically separated virtual aircraft simulators and augment them with constructive entities to make up the complete joint mission environment. Two separate test events are funded with enough resources to conduct two weeks of testing for each event. The experiment is characterized as a factor screening experiment aimed at gaining insight into the usefulness of the MADL network. Additionally, we want to ascertain which factors affect MADL usability in a denied access environment. Aircrew are in short supply with only two aircrew available per week per test phase. This case study focuses on the planning process for this LVC experiment. The experiment execution, data analysis, and conclusions will be discussed in a subsequent paper.

4.1. MADL Data Link

MADL allows aircrews to use voice communication in denied access environments and introduces two other capabilities: text chat, and machine-to-machine communication as shown in **Table 2**. To effectively transmit communications in a denied access environment the data link must not greatly increase the vulnerability of the aircraft to enemy air defenses. To prevent detection during communication, MADL transmits a narrow beam of data between aircraft. With MADL, each aircraft in the network is assigned as a node in the communication chain. To communicate with specific aircraft the subsequent traffic may go direct to that aircraft or be delivered to the aircraft through other aircraft nodes. This network structure can create latency, even failure, in message delivery. Suppose aircraft A, B, and C are linked via MADL and aircraft A wants to communicate with aircraft C. If aircraft B transmits at the same time as aircraft A then aircraft B “steps on” A’s transmission and the message never reaches aircraft C. In other instances, if an aircraft in the network is in an unfavorable geometry at the time of transmission, the MADL chain is broken and the message could be lost. These two issues are of particular interest in the study and can be studied in a controlled manner using the LVC environment.

A simple scenario with an aircraft operating in a denied access environment includes: command and control aircraft operating, friendly fighter forces performing combat air patrol, and targets inside the denied airspace. **Figure 2** depicts a notional MADL operation sufficient to support our discussion. The potential exists for the aircraft or other fighter aircraft to encounter enemy aggressors at any point in the denied airspace. Current operational procedures have the aircraft following pre-planned routes that minimize the probability of detection by enemy integrated air defense (IADS). An experiment objective includes determining if communicating in the denied access environment is useful enough to justify acquiring such capability. This represents an ideal example of using computing power to ascertain the operational effectiveness of proposed upgrades without investing in changes to the weapon systems.

4.2. Defining Experiment Objectives

The first task in an experimental design process is to clearly define the problem to be studied. Defining a clear, agreed upon problem statement for the LVC experiment was the most difficult task in the design process. Four to five months were spent defining the problem statement because influential members of the planning team were focused on defining the requirements for the LVC test environment instead of the data link problem being

Table 2. MADL capabilities.

Level	Available Communication Capability
1	Voice Only
2	Voice and Text
3	Voice, Text, and Machine-to-Machine

¹This case study is an actual event with specific weapons systems unnamed.

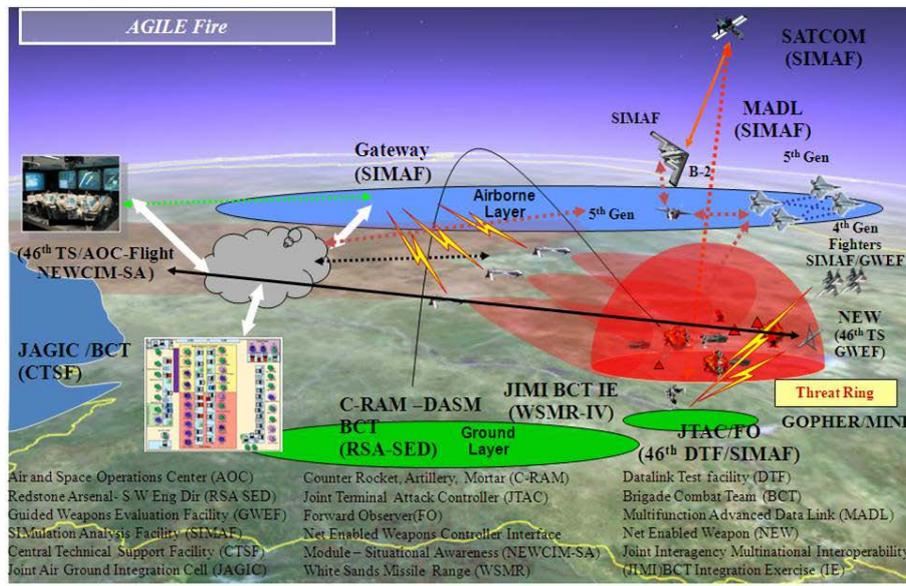


Figure 2. Notional LVC representation of a joint operation network in a denied access environment [18].

investigated; the test should drive what LVC provides. This distraction slowed the progress of the planning phase appreciably, but is really attributable to the paradigm shift associated with using LVC for new purposes. After much deliberation, two related objectives were chosen, one for each phase of the test program.

- 1) **Phase I:** Assess the usefulness of data messages passed on the MADL network assuming a perfect network configuration and performance.
- 2) **Phase II:** Assess the usefulness of the MADL network given a realistic level of degraded network performance.

Phase I assesses whether the message content and message delivery capabilities of MADL are useful to aircrews in prosecuting targets in denied access airspace. Sub-objectives include determining which factors affect the usability of MADL for aircrews and find out which message delivery capabilities are preferred. Following phase I, the set of MADL messages and capabilities will be evaluated with useful messages and capabilities carried forward to phase II. The messages and capabilities deemed not useful will be dropped from the test set. The objective of phase II is to evaluate the usability of MADL messages and capabilities in a realistic environment when network degradation is present (as will likely occur in actual operations).

Breaking the test into two phases is important because it ensures that factor effects are easily identifiable in the data analysis. Consider what would happen if only phase II of the experiment were conducted and the degraded network makes the system so cumbersome that aircrew give it an unfavorable rating. This test method makes it more difficult to tell whether the MADL messages and delivery capabilities are problematic or whether poor network service is the problem. Experimental design helps to focus and clarify the objectives and the data required to achieve the objective.

4.3. Choosing Factors of Interest and Factor Levels

The factors of interest came primarily out of the requirements for the LVC test environment. Initially MADL and the vignettes (operational environment scenarios for the test) were the only two factors proposed for the study. This created an overly simplistic model for study especially when you consider that several other test conditions were to be varied across runs. Such a simplistic yet changing model of the experiment would have yielded results with factor effects confounded with hidden effects. Analytically, no defensible insights could come from such an experiment. Accidental factor confounding is not uncommon if statistical experimental design issues are ignored. Unfortunately, subsequent analyses may proceed without knowledge of the confounding.

Statistical experimental design was re-emphasized at this point in the planning process. Brainstorming re-

sulted in an initial set of 10 (**Table 3**) factors with further consideration reducing the set to 4 factors for phase I and 6 factors for phase II, given in **Table 4** and **Table 5**, respectively. Additionally, one of the MADL factor levels was dropped from the test requirements. Besides MADL as the factor of interest, the operational context (vignettes), ingress route, target location, and aircrew were included as factors in phase I of the experiment. The three latter factors were not of primary interest but were chosen to prevent learning effects in the aircrew during the experiment and its biasing of the outcome. The routes and target locations vary systematically while the aircrew factor is considered a blocking effect. These statistical techniques help guard the experiment against excessive noise introduced by human operators influencing the final results.

In phase II, two additional factors, node position and quality of network service, are added to the phase I design. The additional factors allow a measure of the variation caused by the degraded network. The rule of thumb for choosing factors of interest is to consider adding any setting or test condition changed from run to run as a factor of interest in the experiment.

Table 3. Proposed factors of interest.

Factors	Levels
MADL	4
MADL Node Position	2
Quality of Service	2
Vignettes	4
Route	2
Target Location	2
Aircrew	2
Size of Enemy Air	2
Position of Enemy Air	2
Size of Friendly Air	2
Position of Friendly Air	2

Table 4. Final set of factors of interest for phase I.

Factor	Level
MADL	3
Vignettes	4
Route	2
Target Location	2
Aircrew	2

Table 5. Final set of factors for phase II.

Factor	Level
MADL	3
Vignettes	4
Route	2
Target Location	2
Aircrew	2
Node Position	2
Quality of Service	2

4.4. Selecting the Response Variable

Selecting an appropriate response variable is never easy and can be particularly troublesome in an LVC experiment where many test problem statements are qualitative in nature. Quite often LVC tests employ user surveys to assess qualitative aspects and thus aircrew surveys were proposed for the current test. However, an LVC can collect system state data quite easily. Such state data, if properly defined provides potential insight into the potential benefits of improved system capabilities. In other words, state data can be correlated to qualitative measures, such as aircrew surveys, to develop quantitative measures on qualitative aspects. The approach agreed upon was to use the aircrew survey as a primary response variable with the system state data collected to cross-check and verify aircrew responses and perceptions of the system capabilities.

4.5. Choice of Experimental Design

LVC test requirements can be dynamic; the current case was no exception. Since an LVC offers a tremendous flexibility to expand the test event, unlike comparable live test events, the temptation is to continue to expand the LVC. Due to the ever-changing nature of the test requirements, several experimental designs were considered at various stages in the design process. As requirements were refined, more information about the size and scope of the experiment, the number of virtual and constructive simulation entities, environmental constraints, and aircrew availability came to light. A few of the designs that were contemplated are discussed below along with the rationale for considering that design.

A 16-run 4×4 factorial design was initially considered. The design was discounted as overly simplistic because it ignored potentially important environmental factors. A split-plot design was then considered since the experiment involved a restricted run order. The experimental design team was concerned that completely randomizing MADL capabilities would confuse operators due to large changes in available capability from one level to another. To avoid potential operator confusion the team considered a restricted run order where the run order is chosen by fixing MADL at a particular level then randomizing the run order for the remaining factors. Once all runs have been completed for a given level of MADL, a new MADL level is chosen and the process is repeated until all test runs have been completed for all MADL levels. Such randomization restriction makes the use of split-plot analysis an imperative. [16] shows that analyzing restricted run order experiments as completely randomized designs can lead to incorrect conclusions, a conclusion echoed in [19].

Future use of LVC for test is quite likely to examine impacts of new methods or technology and such examinations affect the design. In the current setting, the MADL-voice-only option was removed as a factor, run separately, and used as a baseline for performance measurement. The rest of the design, now smaller given the removal of a factor, was completely randomized. A replicated, 12-run orthogonal array, shown in **Table 6**, was chosen for phase I. Four additional, replicated runs are completed using voice only to provide a baseline capability for comparison. The orthogonal array is a good option for factor screening experiments since it provides estimates of each of the main effects and select interactions of interest.

Phase II will add two more factors to the experiment making an orthogonal array unusable for a sample size of 12. This led to choosing a nearly orthogonal array (NOA) with replicates. The NOA used for phase II is shown in **Table 7**. If phase I reveals that some factors are inactive then those factors may be dropped from phase II and orthogonality in the design could potentially be restored since Phase II will involve fewer factors.

5. Conclusions

LVC offers the T&E community a viable means for testing systems and system-of-systems in a joint environment. However, the added capability is not without cost and a shift in the paradigm of LVC use. Planning joint mission tests using LVC is a challenging endeavor and requires careful upfront planning. The nature of LVC experiments requires experimenters to decide what should be studied in the experiment when defining the objectives. There is a strong lure toward unnecessary complexity in LVC that entices experimenters to tackle excessively large tests with a misplaced hope that many questions about the system can be addressed simultaneously in that one large experiment. Experimenters need to be aware of this lure and exercise good test discipline by structuring LVC experiments to gain system knowledge incrementally thereby ensuring sound test results. This experimental design method is easily manageable for planning, executing, and analyzing data and builds system knowledge piece by piece.

Table 6. Run matrix for phase I test in standard order.

Run	MADL	Vignette	Route	Target Location
1	1	1	2	2
2	1	2	1	2
3	1	3	1	1
4	1	1	2	1
5	2	1	1	1
6	2	2	2	2
7	2	3	1	2
8	2	2	2	1
9	3	1	1	2
10	3	2	1	1
11	3	2	2	1
12	3	3	2	2

Table 7. Run matrix for phase II test in standard order.

Run	MADL	Vignette	Route	Target Location	Node Position	Quality of Service
1	1	1	2	2	2	2
2	1	2	1	2	1	1
3	1	3	1	1	1	2
4	1	1	2	1	2	1
5	2	1	1	1	1	2
6	2	2	2	2	1	1
7	2	3	1	2	2	1
8	2	2	2	1	2	2
9	3	1	1	2	1	1
10	3	2	1	1	2	2
11	3	2	2	1	1	2
12	3	3	2	2	2	1

LVC test environments have many sources of random error. Considering and exploiting statistical experimental design techniques allow for objective conclusions when the system response is affected by random error. The system response variable should be chosen based on how well that measure relates to the experiment objectives. The response variable should measure this relation as directly as possible. Direct measurements are unobtainable in many LVC experiments so surrogate measures should be devised and examined for suitability. The factors of interest should be chosen from the set of environmental and design parameters that are thought to have an effect on the system response. A good rule of thumb when choosing factors is to consider including any test parameter that will be varied across the runs. Additional design considerations for LVC experiments were proposed to deal with the nuances of LVC. The additional design considerations are by no means exhaustive and should be updated as new challenges are encountered in LVC.

The reported data link experiment demonstrates how experimental design techniques can be used to ultimate-

ly better characterize the performance and effectiveness of a new system in a joint environment generated by LVC. The application of experimental design principles uncovered substantial mistakes in test planning and improved the overall test strategy by using an incremental test approach. Important factors that were initially missed were added to the system as a result of using statistical experimental design. Noise control techniques were used to improve the quality of the data collected. These techniques added necessary complexity to the experiment but improve data quality. The experiments also showed how innovative experimental designs, such as orthogonal and nearly orthogonal arrays, effectively accommodate the large, irregular factor space with limited test resources that are typical of most LVC experiments.

The experimental designs proposed were accepted by the study team. Events prevented the actual conduct of the experiment but the experience of following a systematic experimental design process yielding the specific experimental designs was invaluable to the entire team. The improved quantitative focus on the response variable coupled with the executed experiment would have helped meet the study objectives.

Future LVC experiments can benefit greatly from using such statistical experimental design techniques. This paper did not address the myriad technical issues involved in realizing an LVC environment. Much of the work (and finding) in LVC focuses on solving these technical issues. Our focus in this paper is the design of the experiment that uses the LVC to generate results used in analytical settings. We understand technical issues can affect system responses and we understand that experimental design choices can affect LVC system technical aspects. We leave this discussion to future work for now.

Disclaimer

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense or the US Government.

References

- [1] Bjorkman, E.A. and Gray, F. B. (2009) Testing in a Joint Environment 2004-2008: Findings, Conclusions, and Recommendations from the Joint Test and Evaluation Methodology Project. *ITEA Journal*, **30**, 39-44.
- [2] Department of Defense (2003) Joint Capabilities Integration and Development System. Chairman of the Joint Chiefs of Staff Instruction 3170.01C.
- [3] Department of Defense (2007) Capability Test Methodology v3.0. https://www.jte.osd.mil/jtemctm/handbooks/PM_Online/PM_Handbook_17Apr09.htm
- [4] Hodson, D.D. and Hill, R.R. (2014) The Art and Science of Live, Virtual, and Constructive Simulation for Test and Analysis. *Journal of Defense Modeling and Simulation*, **11**, 77-89. <http://dx.doi.org/10.1177/1548512913506620>
- [5] Hodson, D.D. (2009) Performance Analysis of Live-Virtual-Constructive and Distributed Virtual Simulations: Defining Requirements in Terms of Temporal Consistency. Ph.D. Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- [6] Thorp, H.W. and Knapp, G.F. (2003) The Joint National Training Capability the Centerpiece of Training Transformation. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC)*, Orlando, FL.
- [7] Montgomery, D.C. (2009) Design and Analysis of Experiments. 7th Edition, John Wiley & Sons, New York.
- [8] Coleman, D.E. and Montgomery, D.C. (1993) A Systematic Approach to Planning for a Designed Industrial Experiment. *Technometrics*, **35**, 1-12. <http://dx.doi.org/10.1080/00401706.1993.10484984>
- [9] Madni, A.M. (2010) Integrating Humans with Software and Systems: Technical Challenges and a Research Agenda. *Systems Engineering*, **13**, 232-243.
- [10] Wang, J. and Wu, C. (1992) Nearly Orthogonal Arrays with Mixed Levels and Small Runs. *Technometrics*, **34**, 409-422. <http://dx.doi.org/10.1080/00401706.1992.10484952>
- [11] Heydat, A.S., Sloane, N. and Stufken, J. (1999) Orthogonal Arrays: Theory and Applications. Springer-Verlag, New York. <http://dx.doi.org/10.1007/978-1-4612-1478-6>
- [12] Nguyen, N.-K. (1996) A Note on the Construction of Near-Orthogonal Arrays with Mixed Levels and Economic Run Size. *Technometrics*, **38**, 279-283. <http://dx.doi.org/10.1080/00401706.1996.10484508>
- [13] Xu, H. (2002) An Algorithm for Constructing Orthogonal and Nearly-Orthogonal Arrays with Mixed Levels and Small Runs. *Technometrics*, **44**, 356-368. <http://dx.doi.org/10.1198/004017002188618554>
- [14] Lu, X., Li, W. and Xie, M. (2006) A Class of Nearly Orthogonal Arrays. *Journal of Quality Technology*, **38**, 148-161.

- [15] Myers, R.H., Montgomery, D.C. and Anderson-Cook, C.M. (2009) *Response Surface Methodology*. 3rd Edition, John Wiley & Sons, New York.
- [16] Jones, B. and Nachtshiem, C.J. (2009) Split-Plot Designs: What, Why, and How. *Journal of Quality Technology*, **41**, 340-361.
- [17] Horton, M.J.D. (2012) *Conflict: Operational Realism versus Analytical Rigor in Defense Modeling and Simulation*. Graduate Research Project, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- [18] Bjorkman, E.A. (2010) *USAF War Fighting Integration: Powered by Simulation*. In *WinterSim'10: Proceedings of the 2010 Winter Simulation Multiconference*, San Diego, CA, Society for Computer Simulation International.
- [19] Cohen, A.N. (2009) *Examining Split-Plot Designs for Developmental and Operational Testing*. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

