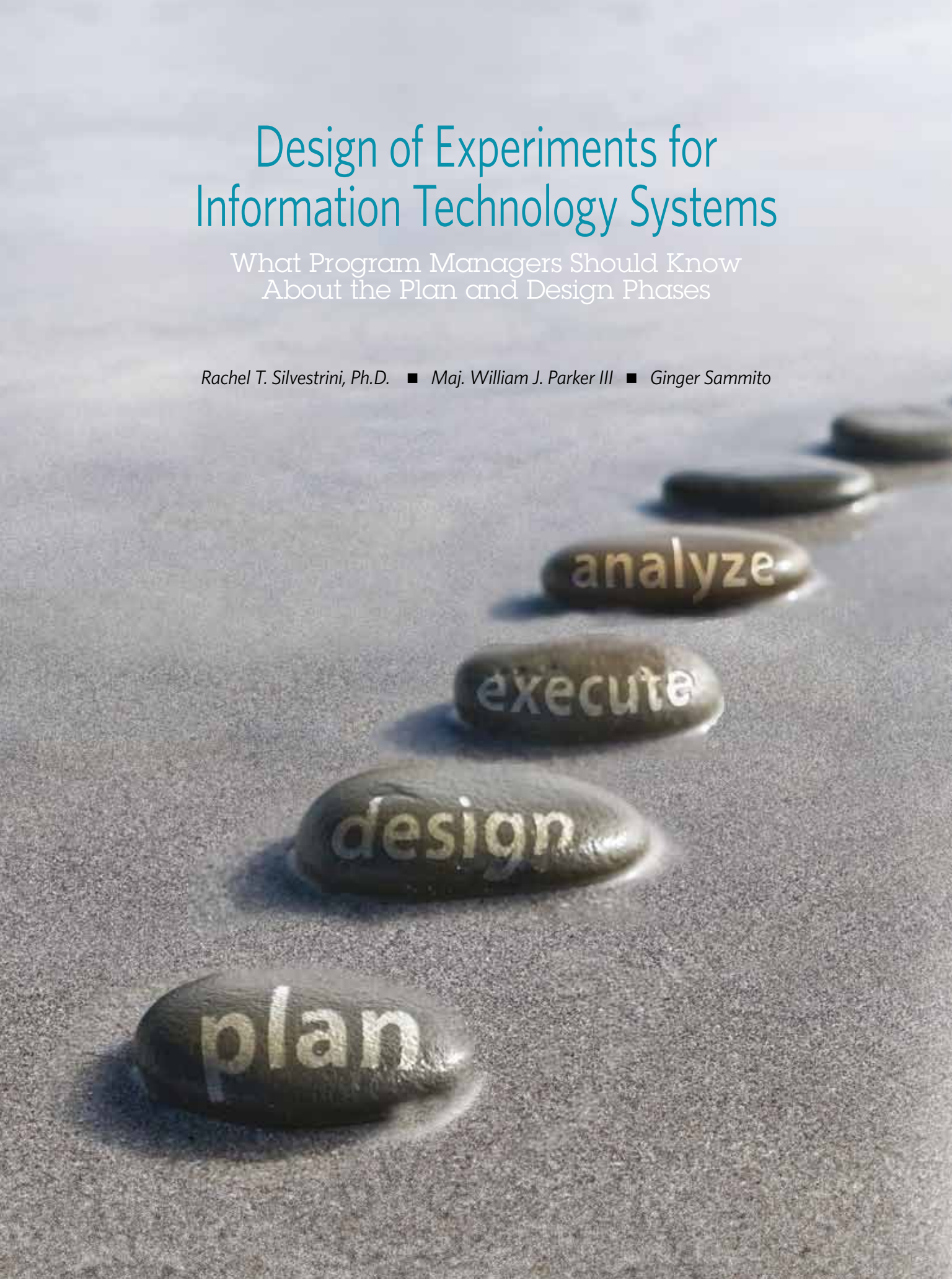
Design of Experiments for Information Technology Systems

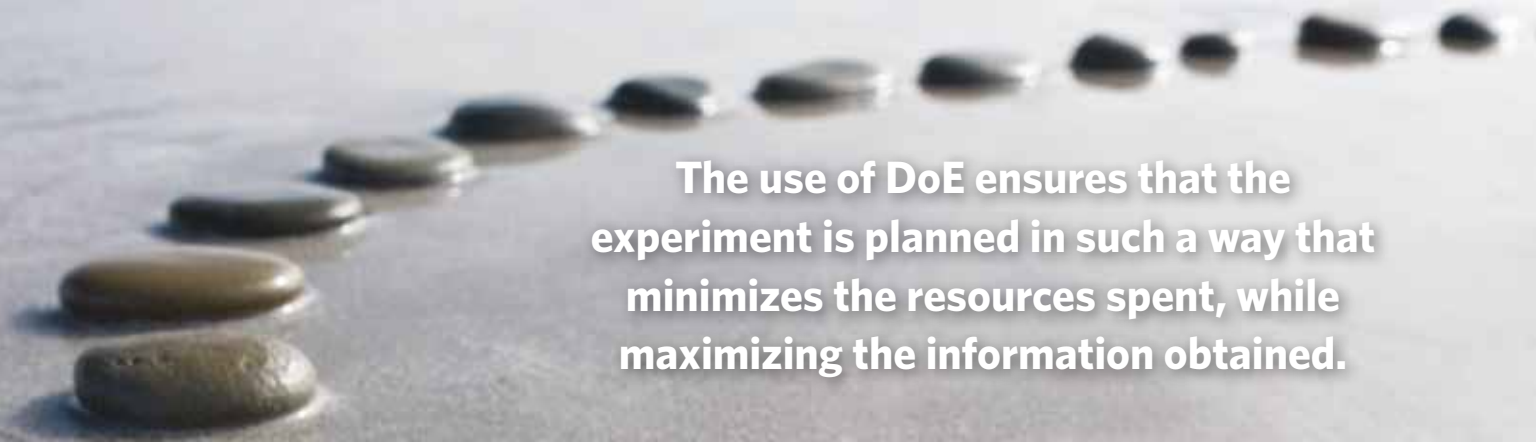
What Program Managers Should Know About the Plan and Design Phases

Rachel T. Silvestrini, Ph.D. ■ *Maj. William J. Parker III* ■ *Ginger Sammito*



Recent mandates require that rigorous statistical and mathematical approaches be applied to all tests that fall under developmental and operational test and evaluation (T&E). On October 19, 2010, J. Michael Gilmore, director of Operational Test and Evaluation, released a memorandum to the T&E community within the DoD that describes an initiative designed to increase the use of scientific

Silvestrini *is an assistant professor in the Operations Research Department at the Naval Postgraduate School.* **Parker** *is a C4ISR systems operational test director and operations research system analyst for the Homeland Security/Information Assurance Portfolio at the Joint Interoperability Test Command.* **Sammito** *is a principal operations research system analyst for the Force Application/Force Protection Portfolio at the Joint Interoperability Test Command.*



and statistical methods to develop rigorous methods for test and data analysis. Dr. Gilmore's memo specifies the need for using rigorous statistical based testing methods in order to ensure that proper and sufficient data is collected to answer the question of interest. In addition, Edward R. Greer, the director of Developmental Test and Evaluation, has championed the skillsets of design of experiments (DoE), statistics, and test design principles in the rejuvenation and development of the T&E workforce as one of his top initiatives to the practice of T&E.

The framework that encompasses the statistical and mathematical approaches for T&E is called scientific based test design (SBTD). SBTD can be applied to all fields and application areas within the T&E realm. There is no set of T&E experiments in which SBTD does not apply. For example, consider the program manager (PM) who is involved with IT systems and feels that SBTD cannot be applied to his/her respective system because the variable measures of interest in the experiment results in a binary outcome. In other words, did the system work (yes or no)? Although this is a formidable challenge that must be considered prior to running the experiment, it is not a showstopper.

SBTD is a framework that includes statistical based methods for T&E such as DoE and regression analysis. DoE is a formal approach for the development of a set of tests to be carried out in an experiment. An experiment is a large number of individual tests (also called trials or runs) where variables are manipulated and data is collected.

There are abundant sources of literature on DoE that describe the mathematical and statistical based tactics for designing and analyzing the results of an experiment that can meet the needs of any experimental goals. These methods ensure that valid, objective, and scientific conclusions are reached. Additionally, the use of DoE ensures that the experiment is planned in such a way that minimizes the resources spent, while maximizing the information obtained. Figure 1 highlights the four phases of the DoE approach: Plan, Design, Execute, and Analyze.

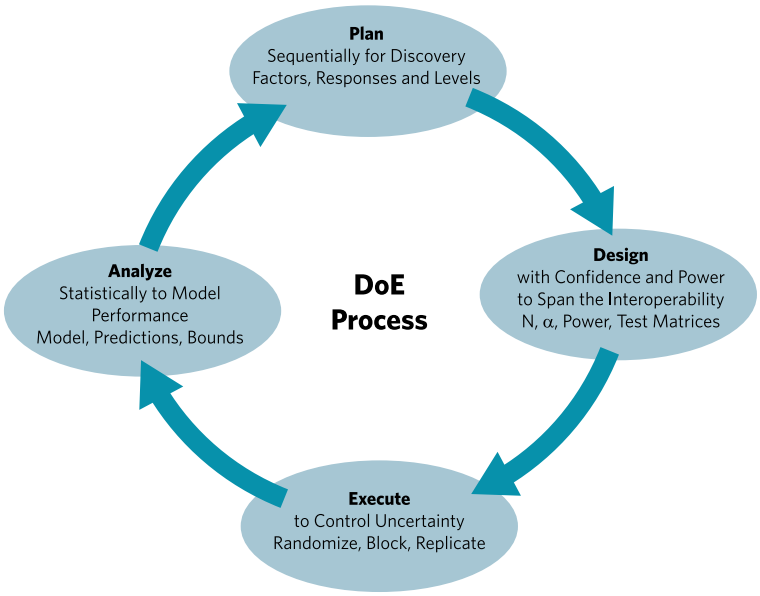
Unlike the T&E of traditional weapons systems such as aircraft, tanks, artillery, maritime vessels, etc., the PM involved with IT systems testing may experience slightly different challenges associated with the T&E processes. However, the phases of DoE process do not change for anyone. While this article is primarily aimed at the PM within T&E of IT systems, it is intended to be beneficial reading for any PM involved with T&E in the DoD. The remainder of this article will briefly cover how to apply the first two phases of DoE through an example application to an IT system. When appropriate, specific challenges one might encounter will be highlighted.

Applying Science Based Testing Designs

The DoE approach to the experiments conducted during the T&E process is displayed in Figure 1. The first two phases of this process (Plan and Design) will be discussed through an example application to an IT system.

Suppose that a PM is in charge of oversight for a new software application being developed as a test tool. The experiment used to test the software is called Bravo Test. During Bravo Test different message types for multiple platforms with an Identification Friend or Foe (IFF) system are both transmitted and received. A DoD architecture framework is illustrated in Figure 2. Bravo Test will take place at the systems level (middle view).

Figure 1. Design of Experiments (DoE) Process



Compliments: 46th Test Wing OA

Phase 1: Plan

The first phase in the DoE process is Plan. This phase includes statement of the goal of the experiment as well as the development of a list of variables involved in the experiment. There are three types of variables important to list:

- variables that will be manipulated or controlled during the experiment
- variables that cannot be controlled, but may change during the experiment
- variables used to measure the system (outcomes)

The goal of Bravo Test is to test the accuracy and timeliness of messages transmitted and received. The first objective of Bravo Test is to determine whether or not each of four different platforms transmits or receives messages with accuracy rate above 99 percent. The second objective is to model the expected time to transmit and receive a message as a function of the different platforms, identification systems, and type of message. The PM should be aware that the recognition of the goal and objectives in a test often aid in identifying the variables present in the experiment.

Table 1 illustrates the three different controllable variables that will be manipulated (changed) over the course of Bravo Test. Remember; variables that can be controlled as well as those that cannot be controlled should be identified. For example, during Bravo Test the average system load during the transmission of a message may be measurable, but it may not be a variable that is directly controllable. The PM should be eager to identify all uncontrollable variables possible and additionally keep in mind that it is possible that a few variables may not be known initially, but will emerge later. This should not be a stumbling point, but an opportunity for the PM to refine the test during the next cycle with more information. This involves going back to the planning phase and proceeding from there.

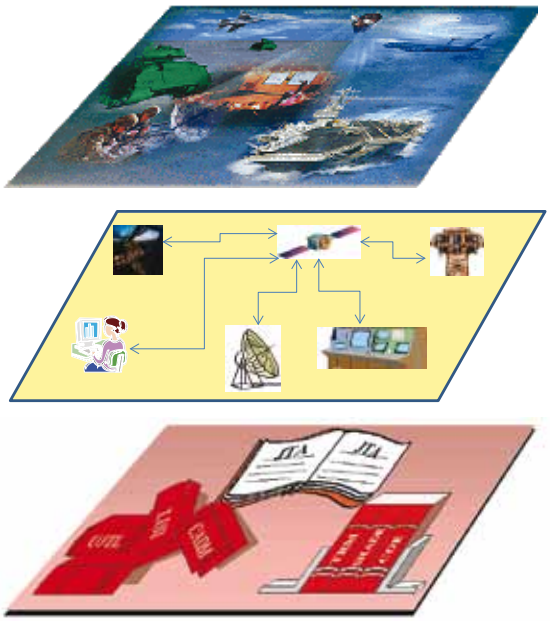
Example Factors to be varied during Bravo Test

Controllable Variables	Settings During Test
IFF (Identification, Friend, or Foe)	Range 0 - 5
Message types	UTF-8, UTF-16, UTF-32 (UTE = Unicode Transformation Format)
Producing or Consuming Platforms	A, B, C, D

In Bravo Test, there are two outcome variables: (1) accuracy of message and (2) time to transmit/receive message. Accuracy is a binary variable: if the message is 100 percent correct, the data point will be considered 1 (accurate); otherwise 0 (not accurate). In IT systems testing, a binary response is a common metric of interest. Also, many outcome variables may be collected for a single test within the experiment; this is important to note and is used when assessing the quantity of tests required for the experiment.

Without proper care in the Plan phase of the experiment, the direction of the experiment may become unclear. This leads to the collection of erroneous or incomplete information, which will prevent the experimental goals from being met. Often, determining the variables of interest in an experiment can be a difficult task that should be undertaken with caution. Fishbone diagrams as well as other brainstorming techniques often work well during subject matter expert meetings to discuss variable selection.

Figure 2. DoD Architecture Framework with Systems View in Center



The **Operational View** describes and interrelates the operational elements, tasks and activities, and information flows required to accomplish mission operations.

The **Systems View** describes and interrelates the existing or postulated technologies, systems, and other resources intended to support the operational requirements.

The **Technical View** describes the profile of rules, standards, and conventions governing systems implementation and forecasts their future direction.

DoD Architectural Framework (DoDAF)

Phase 2: Design

The Design Phase involves mapping out the sets of tests that will be conducted during the experiment. Specifically, this phase involves the selection of the design type and the determination of the number of tests to be conducted in the experiment (also known as sample size). Each test involves the control and manipulation of variables identified in the Plan Phase. There are a number of different experimental design techniques found in various textbooks, journal articles, technical reports, and case studies.

Examples of design selections include factorial design, fractional factorial design, central composite design, covering array, and optimal design. While

Figure 3. JMP—User Interface for the Development of Full Factorial Design

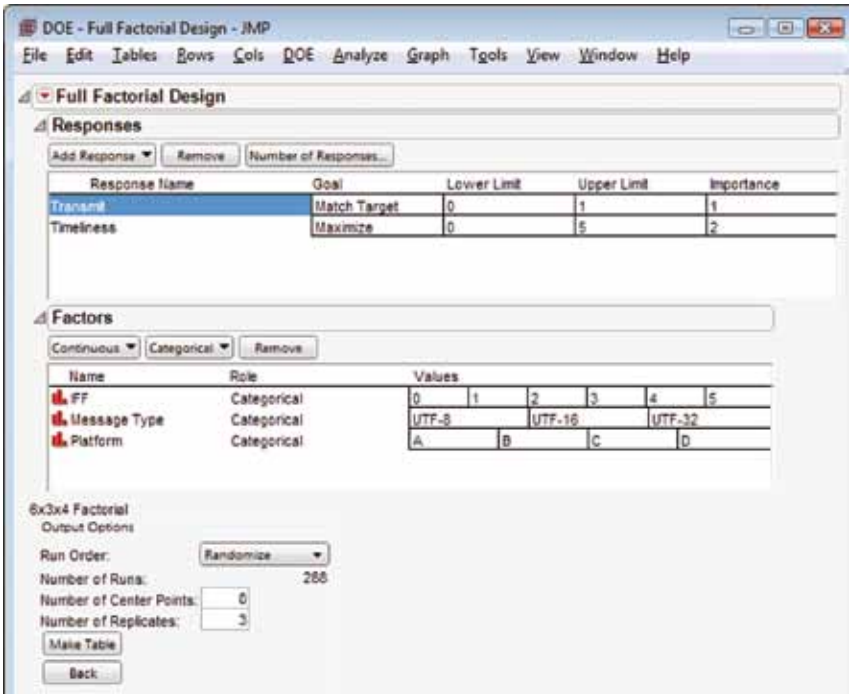
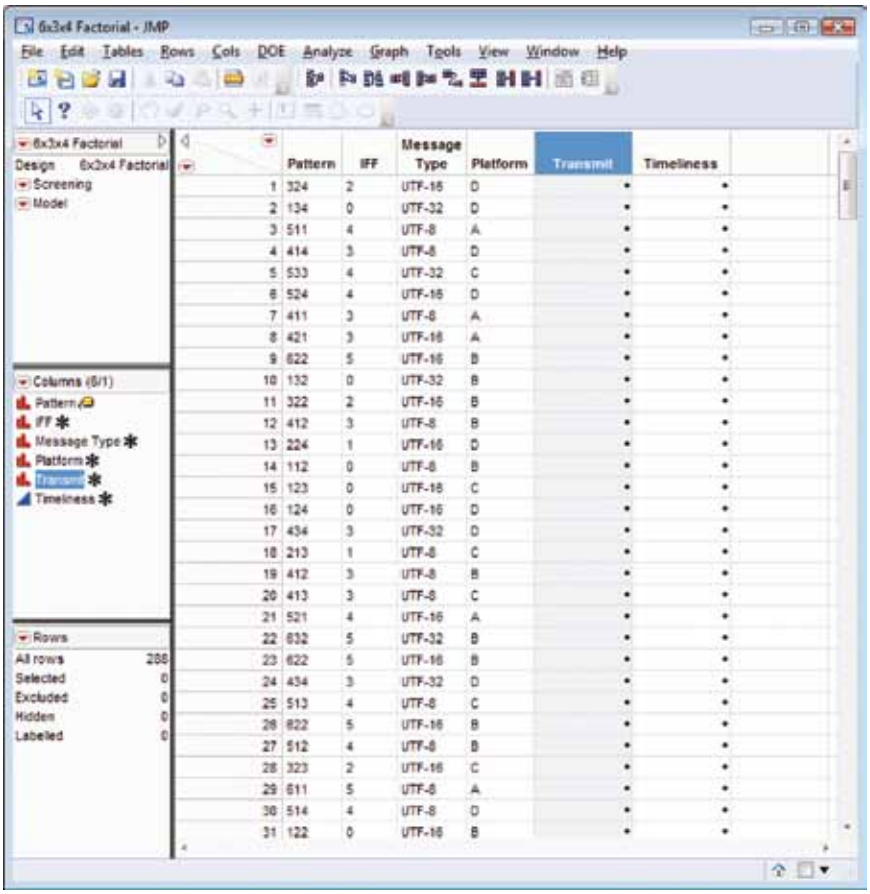


Figure 4. JMP—Full Factorial Table Design



a PM does not necessarily need to know each different design, they should recognize that different designs are appropriate for different experimental goals. For example, a fractional factorial design is an appropriate design choice when the experimental goal involves finding the subset of factors that influence the outcome variable of interest. This is a goal typically encountered in the early phase of testing. For situations involving multiple responses with overlapping or conflicting goals, a hybrid design approach, in which different design choices are combined, can be used to satisfy all objectives of the experiment.

In addition to design choice, the number of tests to run (or the sample size) of an experiment must be determined during this phase. Given the opportunity, a PM might prefer to choose an unlimited sample size. However, cost, time, and resource constraints often drive sample size choices.

For Bravo Test, a full factorial design with four replicates is selected to support the goals of testing the accuracy and timeliness of messages transmitted and received. A statistical software package, such as JMP (illustrated), can be used to create the design. Snapshots of the design creation are shown in Figure 3 and Figure 4. Figure 3 illustrates the user interface that guides the inputs to the development of the design. Figure 4 contains the design. The design dictates the running of every experimental test. For example, the first experimental test will be conducted with IFF = 2, Message Type = UTF-16, and Platform = D.

A full factorial design is appropriate for the needs of Bravo Test. In Bravo Test, simple relationships between IFF, Message Type and Platform will be investigated. In other situations, different designs may be more apt. The factorial design dictates a baseline number of runs in the experiment. That number can be altered by repetition of the experiment (as seen in one of the selection tabs in Figure 3). It is important for the PM to realize that within a resource-constrained environment, a single experiment cannot provide unlimited answers. Both design choice and sample size restrictions translate to restrictions on what information can be obtained. Statistical and math-

ematical analysis can greatly help overcome sample size dilemma by focusing on answering the following:

- Given a fixed sample size, what information can be measured and modeled?
- Given measurement or modeling requirements, what sample size is required?

Approach (1) involves identifying risks in the constrained environment and approach (2) involves determining requirements of sample size based on the risks the experimenter is willing to accept. Risks can be discussed in terms of confidence level and/or power of mathematical estimation. These are two terms related to statistical analysis that PMs should be or become familiar with.

During the Design Phase, the PM should encourage documentation of the methodology that includes rationale for selecting a design, sample size, and lessons learned from the process. Clear documentation will help the PM face the challenges of the iterative DoE process and development stages as the software moves towards maturity.

Conclusion

SBTD methods, specifically DoE, can and should be applied to T&E of IT systems. There are many case studies that document the success of the DoE approach for both IT and non-IT systems. This article covered the Plan and Design phases in the DoE approach. It is believed that the Plan and Design phases are of utmost importance because an inadequately designed experiment will result in poor results and possibly incorrect conclusions, thus making the Execute and Analyze phases meaningless.

The Execute Phase refers to the running of each test in the experiment. For Bravo Test, the experiment to be run is illustrated in Figure 4. During this phase, it is imperative that each test is run to specification. This involves ensuring that proper blocking, randomization, and replication are carried out as specified by the design. The Analyze Phase encompasses a mathematical study of the resulting data to obtain valid and objective conclusions.

Sometimes the challenges and decisions in the creation of an experimental design approach appear endless for the PM, especially as requirements shift from traditional testing to rigorous SBTD for IT systems. The PM must ensure compliance with applicable policies. The PM is also responsible for the quality and consistency to those standards while developing test reports based on a sound, scientific rigor that have not formally been a part of any IT system/program. The PM needs to look beyond the present in facing these SBTD challenges in IT systems and focus on the valid, objective, and measureable approach that ultimately saves time and money over the development cycle of the IT system.

The authors can be reached at **rtsilves@nps.edu**, **william.j.parker60.mil@mail.mil**, *and* **ginger.j.sammito.civ@mail.mil**.

DAU Alumni Association

JOIN THE SUCCESS NETWORK

The DAU Alumni Association opens the door to a worldwide network of Defense Acquisition University graduates, faculty, staff members, and defense industry representatives—all ready to share their expertise with you and benefit from yours.

Be part of a two-way exchange of information with other acquisition professionals.

- Stay connected to DAU and link to other professional organizations.
- Keep up to date on evolving defense acquisition policies and developments through DAUAA newsletters and symposium papers.
- Attend the DAUAA Annual Acquisition Community Conference/ Symposium and earn Continuous Learning Points (CLPs) toward DoD continuing education requirements.

Membership is open to all DAU graduates, faculty, staff, and defense industry members. It's easy to join, right from the DAUAA website at *www.dauaa.org*.

For more information,
***call* 703-960-6802** or **800-755-8805**, or e-mail *dauaa2(at)aol.com*.