Design of Experiments – Example for Software-Intensive System

(The following section would appear in the body of the TEMP for a Command and Control System at MS C. Appendix material begins on page 4.)

3.2 Test and Evaluation Framework

The Operational Test Activity (OTA) will accomplish the following during integrated testing:

- Determine if thresholds in the approved capabilities documents and COIs have been satisfied
- Determine Operational Effectiveness, Survivability, and Suitability of the system under realistic operational conditions
- Assess the contribution of the system to combat operations
- Provide additional information on the system’s operational capabilities and limitations.

The OTA’s evaluation plan creates a framework and methodology for evaluating the entirety of program data, obtained from late developmental testing, an operational assessment and IOT&E. The evaluation plan is intended provide a transparent, repeatable, and defensible approach to evaluation. The evaluation framework is captured in Table 3-1. The test team developed the test strategy by employing Design of Experiments (DOE) to ensure that a rigorous methodology supports the development and analysis of test results. DOE is used to design the tests to evaluate the data fusion KPP and the three COIs outlined in Table 3-1. A designed experiment is used to determine the effect of a factor or several factors (also called independent variables) on one or more measured responses (also called dependent variables). All COI DOEs are designed with mission-oriented response variables. Each design will include an estimation of the power of the test, which is included in the DOE Appendix. When gaps in the design are identified, these gaps will be listed as limitations, and a risk assessment will be provided in the appropriate Detailed Test Plan. In addition, the team will work with all appropriate parties to determine the most appropriate way to mitigate and/or manage the risks.

The OTA intends to exercise the command and control system during multiple training exercise (for a list of resources, see section 4.0) and dedicated test events. Real operators will be using the system for all tests where the data is considered in the evaluation of the COIs and data fusion KPP.

The Integrated test team has identified the response variables, factors and levels that will be exercised during each event in Table 3-2 to 3-5. The exact test size, experimental design, including expected trial replications, and confidence and power levels are outlined in the DOE Appendix. The identified confidence level and power are the maximums expected in a completely randomized event, due to restrictions in randomization. The major risk of not completely randomizing the design is that some factors may become confounded with uncontrollable variables. The OTA will work to avoid any obvious confounding of variables. Data collected in training exercise will be supplemented by dedicated test events to mitigate any risks of data loss due to exercise objectives.

Table 3-2. Overview of DOE Strategy to Assess the Data Fusion KPP

		Test Phase		
		DT	**OA**	**IOT**
Critical Responses →		Track Accuracy, Timeliness, and Completeness	Track Accuracy, Timeliness, and Completeness	Track Accuracy, Timeliness, and Completeness
Factors	**Factor Levels**			
Connection	Categorical Factor with 5 levels: JREAP A/B/C, Link-16, CTN	SV*	SV	Record*
Number of Tracks	Low, Threshold, Objective	SV	SV	SV (simulated tracks in addition to live tracks)
Type of Track	Real time, Near real time, non-real time	SV	SV	Record

*Factors labeled systematically vary (SV) will be included in the DOE for data fusion. The data fusion DOE will be primarily executed in DT and the OA, IOT data will be used to confirm the results from DT and OT. If major configuration updates are made to the system between the OA and IOT, the factor management strategy for OT may need to be updated.

Tables 3-3 and 3-4 follow a similar format to Table 3-2 but are specific to each agency's respective mission.

Finally, a minimum of 3,000 hours of operation, equally spread across all three of the agencies employing the system are required to evaluate RAM and Ao requirements. These operation hours will be collect across late DT testing, the operational assessment, and the IOT&E. In order for the hours to count in the operational suitability assessment the system must be in a near final configuration and operated by operationally representative users.

Table 3-3. Overview of DOE Strategy to assess COI 1: System's ability to support mission of agency 1.

		Test Phase

Design of Experiments – Example for Software-Intensive System

		DT	OA	IOT
Critical Responses →		1.Response time for critical information download/upload. 2.Number of missions successfully controlled.	1.Response time for critical information download/upload. 2. Rating of ability to control aircraft. 3.Number of missions successfully controlled.	1.Response time for critical information download/upload. 2.Rating of ability to control aircraft. 3.Number of missions successfully controlled.
Factors	**Factor Levels**			
Mission Load	Standard, High	SV	SV	SV
Track density	Standard, High	SV	SV	SV (simulated tracks in addition to live tracks)
Mission Duration	Short (4 hours), 24 hour operations	SV	SV	SV
Configuration	Small, Medium, Large	HC (Small)	HC (Medium)	HC (Large)
Environment	Desert, Hot & Humid, Cold	HC (Desert)	HC (Hot & Humid)	HC (Desert)

Sample DOE Appendix – Design of Experiment for COIs and Data Fusion KPP

Data Fusion KPP

Response variables

The data fusion KPP will be evaluated using the following critical measures, which have threshold requirements:

- Track Accuracy
- Track Completeness
- Track Timeliness

Factors

The following factors were considered for the data fusion KPP:

- Connection Method (JREAP A/B/C, Link-16, CTN)
  - Connection methods will be used both independently and simultaneously to assess an interoperability issues that may result
- Number of tracks (Low, Threshold, Objective)
- Type of Tracks (Real time, Near real time, Non-real time)

Table D-1 below provides the experimental design along with replications for achieving high power at the 95% confidence level to detect significant differences in factor levels. The power for detecting differences in the outcome based on the connection method is 91%, the power for detecting differences in the outcome based on the number and type of track is 99%. This design will be executed between both the developmental testing and the operational assessment. Half of each of the four runs will be conducted in DT, the other half will be conducted in the operational assessment. If for any reason this testing is not completed in DT and the OA it will be completed in the OT.

Table D-1. Experimental Design for Data Fusion KPP

		Connection Method					
Number Tracks	**Track Type**	JREAP A	JREAP B	JREAP C	Link-16	CTN	All Links
Low	Real time	4	4	4	4	4	4

	Near-real	4	4	4	4	4	4
	Non-real	4	4	4	4	4	4
Threshold	Real time	4	4	4	4	4	4
	Near-real	4	4	4	4	4	4
	Non-real	4	4	4	4	4	4
Objective	Real time	4	4	4	4	4	4
	Near-real	4	4	4	4	4	4
	Non-real	4	4	4	4	4	4

Figure D-1 shows power as a function of the number of replicates for each condition. Four replicates provide adequate power at the 95% confidence level to assess the data fusion KPP across all test conditions.

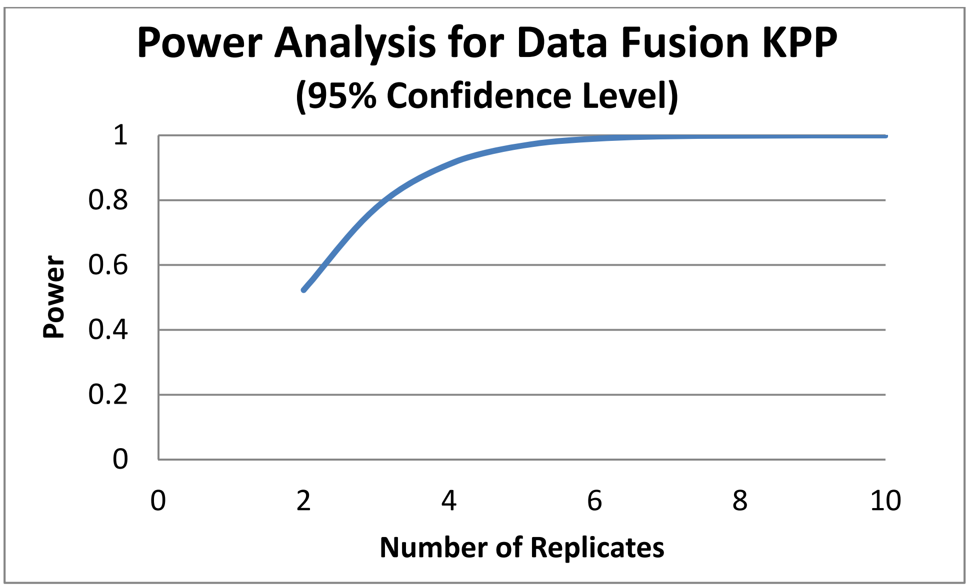

Figure D-1. Power Analysis for Data Fusion KPP

A similar discussion should follow for each of the additional COIs including the responses, factors, a proposed experimental design, and rational for the number of test points.