## ANNEX A – ACRONYMS AND ABBREVIATIONS

| Acronym/Abbreviation | Meaning |
|---|---|
| AB | Analytical Baseline |
| ABM | Agent Based Model |
| ABMS | Agent Based Modeling and Simulation |
| ABS | Agent-Based Simulation |
| ACMC | Assistant Commandant of the Marine Corps |
| ACV | Airspace Control Volume |
| AFATDS | Advanced Field Artillery Tactical Data System |
| AFB | Air Force Base |
| AFOTEC | Air Force Operational Test & Evaluation Center |
| AFPD | Air Force Policy Directive |
| AI | Area of Interest |
| Analyst's Handbook | Analyst's Handbook for Testing in a Joint Environment |
| ANOVA | Analysis of Variance |
| AO | Action Officer; Area of Interest |
| AO's Handbook | Action Officer's Handbook for Testing in a Joint Environment |
| AoA | Analysis of Alternatives |
| AOI | Area of Influence |
| AP | Analysis Plan |
| API | Application Programming Interface |
| ARSM | Advanced Response-Surface Methodology |
| ASD(NII) | Assistant Secretary of Defense for Networks and Information Integration |
| AT&L | Acquisition, Technology and Logistics |
| ATCCS | Army Tactical Command and Control System |
| AUTL | Army Universal Task List |
| AV | All View |
| AW | Air Warfare |
| AWACS | Airborne Warning and Control System |
| BCT | Brigade Combat Team |
| BOV | Blue Operational View |
| BSV | Blue Systems and Services View |
| C2 | Command and Control |
| C4I | Command, Control, Communications, Computers, and Intelligence |
| C4ISR | Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance |
| CAB | Combat Action Brigade |
| CADM | Core DoD Architecture Data Model |
| CAP | Capability Analysis Plan |
| CART | Classification and Regression Tree |
| CAS | Close Air Support |
| CBA | Capabilities-Based Assessment |

| Acronym/Abbreviation | Meaning |
|---|---|
| CBP | Capability-Based Planning |
| CCA | Close Combat Attack |
| CCD | Central Composite Design |
| CCI | Critical Capability Issue |
| CCJO | Capstone Concept for Joint Operations |
| CDD | Capability Development Document |
| CDR | Critical Design Review |
| CEM | Capability Evaluation Metamodel |
| CFF | Call for Fire |
| CGF | Computer Generated Forces |
| CIO | Chief Information Officer |
| CJCS | Chairman of the Joint Chiefs of Staff |
| CJCSI | Chairman of the Joint Chiefs of Staff Instruction |
| CJCSM | Chairman of the Joint Chiefs of Staff Manual |
| CLE | Nomenclature designation for a Continuous Learning Module |
| CM | Capability Manager |
| CMU | Connectivity Matrix Utility |
| COCOM | Combatant Command |
| COI | Critical Operational Issue |
| COIN | Counterinsurgency |
| COMBAT$^{XXI}$ | Combined Arms Analysis Tool for the XXIst Century |
| CONOPS | Concept of Operations |
| COTS | Commercial Off-the-Shelf |
| CPD | Capability Production Document |
| CPM | Capability Portfolio Manager |
| CRN | Common Random Number |
| CSB | Controlled Sequential Bifurcation |
| CTA | Capstone Threat Assessment |
| CTM | Capability Test Methodology |
| CTM 1 | CTM Step 1 |
| CTM 2 | CTM Step 2 |
| CTM 3 | CTM Step 3 |
| CTM 4 | CTM Step 4 |
| CTM 5 | CTM Step 5 |
| CTM 6 | CTM Step 6 |
| CTO | Combine Test Organization |
| CTP | Critical Technical Parameter |
| d.f. | Degree of Freedom |
| DAB | Defense Acquisition Board |
| DAP | Data Analysis Plan |
| DARS | DoD Architecture Registry System |
| DAU | Defense Acquisition University |
| DAUVS | Digital Army USMTF VMF Stimulator |
| DCARS | Digital Collection, Analysis, and Review System |

| Acronym/Abbreviation | Meaning |
|---|---|
| DCIT | Distributed Capabilities Integration Toolbox |
| DCM | Data Collection Matrix |
| DCP | Data Collection Plan |
| DCR | DOTMLPF Change Recommendation |
| DD | Deputy Director |
| DD,AW | Deputy Director, Air Warfare |
| DE | Distributed Environment |
| DecSecDef | Deputy Secretary of Defense |
| DELT | Data Elements List Table |
| DIA | Defense Intelligence Agency |
| DIACAP | DoD Information Assurance Certification and Accreditation Process |
| DIS | Distributed Interactive Simulation |
| DMAP | Data Management and Analysis Plan |
| DMP | Data Management Plan |
| DoD | Department of Defense |
| DoDAF | DoD Architecture Framework |
| DoDI | DoD Instruction |
| DOE | Design of Experiment |
| DOT&E | Director, Operational Test and Evaluation |
| DOTMLPF | Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities |
| DP | Decision Point; Design Point |
| DPG | Defense Planning Guidance |
| DPS | Defense Planning Scenario |
| DRCM | Distributed Range Capabilities Matrix |
| DRCT | Distributed Range Coordination Team |
| DSB | Defense Science Board |
| DSIG | Domain Special Interest Group |
| DT | Development Test; Developmental Test |
| DT&E | Developmental Test and Evaluation |
| DWS | Data Warehouse System |
| EA | Enterprise Architecture |
| EM | Event Manager |
| EMP | Event Management Plan |
| EPG | Electronic Proving Ground |
| ES | Executive Summary |
| EV | Evaluation View |
| EW | Electronic Warfare |
| FAA | Functional Area Analysis |
| FARA | Federal Acquisition Reform Act |
| FBCT | Force Brigade Combat Team |
| FCS | Future Combat System |
| FDD | Federation Object Model Document Data |
| FED | Federation Execution Data |

| Acronym/Abbreviation | Meaning |
|---|---|
| FFBD | Functional Flow Block Diagram |
| FID | Foreign Internal Defense |
| FNA | Functional Needs Analysis |
| FO | Forward Observer |
| FOM | Federation Object Model |
| FoS | Family of Systems |
| FOT&E | Follow-on Test and Evaluation |
| FPC | Final Planning Conference |
| FSA | Functional Solutions Analysis |
| FSE | Fire Support Element |
| FY | Fiscal Year |
| FYAB | Fiscal Year Analytical Baseline |
| GCCS-M | Global Command and Control System - Maritime |
| GIG | Global Information Grid |
| GIS | Geographic Information System |
| GL | Glossary |
| GOSC | General Officer(s) Steering Committee |
| GPS | Global Positioning System |
| GRSM | Generalized Response-Surface Methodology |
| HIMARS | High Mobility Artillery Rocket System |
| HITL | Human-in-the-Loop |
| HLA | High Level Architecture |
| HPCC | High Performance Computing Cluster |
| HSD | Honestly Significant Difference |
| HSLT | High Speed LAN TAP |
| ICD | Initial Capabilities Document |
| ICP | Integrated Capability Portfolio |
| IDEF | Integrated Definition for Data Modeling |
| IDFW | International Data Farming Workshop |
| IDRL | Integrated Data Requirements List |
| IEEE | Institute of Electrical and Electronics Engineers, Inc. |
| IER | Information Exchange Requirement |
| IF | Indirect Fires |
| INC | Interface Network Controller |
| IOT&E | Initial Operational Test and Evaluation |
| IPC | Initial Planning Conference |
| IPL | Integrated Priority List |
| IPT | Integrated Product Team |
| ISR | Intelligence, Surveillance, and Reconnaissance |
| IT | Information Technology |
| ITWA | Initial Threat Warning Assessment |
| JAGS | Joint Air-to-Ground System |
| JBD2 | Joint Battlespace Dynamic Deconfliction |
| JCA | Joint Capability Area |

| Acronym/Abbreviation | Meaning |
|---|---|
| JCAS | Joint Close Air Support |
| JCD | Joint Capabilities Document |
| JCE | Joint Capability Evaluation |
| JCER | Joint Capability Evaluation Report |
| JCIDS | Joint Capabilities Integration and Development System |
| JCS | Joint Chiefs of Staff |
| JDS | Joint Data Support |
| JE | Joint Environment |
| JFC | Joint Functional Concept |
| JFEO | Joint Forcible Entry Operation |
| JFIRES | Joint Fires |
| JFM | Joint Mission Environment (JME) Foundation Model |
| JIC | Joint Integrating Concept |
| JM | Joint Mission |
| JMe | Joint Mission Effectiveness |
| JME | Joint Mission Environment |
| JMET | Joint Mission Essential Task |
| JMETC | Joint Mission Environment Test Capability |
| JMETL | Joint Mission Essential Task List |
| JOA | Joint Operational Area |
| JOC | Joint Operating Concept |
| JOC-T | Joint Operational Context for Test |
| JOpsC | Joint Operations Concepts |
| JP | Joint Publication |
| JPD | Joint Potential Designator |
| JPME | Joint Professional Military Education |
| JROC | Joint Requirements Oversight Council |
| JS | Joint Staff |
| JSTARS | Joint Surveillance Target Attack Radar System |
| JT&E | Joint Test and Evaluation |
| JTAC | Joint Terminal Attack Controller |
| JTEM | Joint Test and Evaluation Methodology |
| JVMF | Joint Variable Message Format |
| KPP | Key Performance Parameter |
| KSA | Key System Attribute |
| LCIM | Level of Conceptual Interoperability Model |
| LDM | Logical Design Model |
| LF | Live Fire |
| LFT&E | Live Fire Test and Evaluation |
| LH | Latin Hypercube |
| LRIP | Low Rate Initial Production |
| LS | Launch System |
| LVC | Live, Virtual, Constructive |
| LVC-DE | Live, Virtual, Constructive Distributed Environment |

| Acronym/Abbreviation | Meaning |
|---|---|
| M&P | Methods and Processes |
| M&S | Modeling and Simulation |
| MADM | Multi-Attribute Decision Making |
| MANA | Map Aware Non-Uniform Automata |
| MARS | Multivariate Adaptive Regression Spline |
| MCO | Major Combat Operation |
| MDA | Milestone Decision Authority |
| METOC | Meteorological and Oceanographic |
| MFMEA | Matrix Failure Modes and Effects Analysis |
| MHS | Message Handling System |
| MIL-STD | Military Standard |
| Mission MOE | Mission Measure of Effectiveness |
| MMOE | Mission Measure of Effectiveness |
| MODAF | Ministry of Defence Architecture Framework (UK) |
| MOE | Measure of Effectiveness |
| MOP | Measure of Performance |
| MOS | Measure of Suitability |
| MOSA | Measure of System/SoS Attribute |
| MPC | Mid-Planning Conference |
| MRSM | Modified Response-Surface Methodology |
| MSCO | Modeling and Simulation Coordination Office |
| MSEL | Master Scenario Events List |
| MSFD | Multi-Service Force Deployment |
| MSRR | Modeling and Simulation Resource Repository |
| MTBF | Mean Time Between Failure |
| MTM | Metamodel-Test-Metamodel |
| MTTR | Mean Time to Repair |
| MVR | Maneuver |
| NASA | National Aeronautics and Space Administration |
| NAVSTAR | Navigation Satellite Timing and Ranging |
| NBC | Nuclear, Biological, Chemical |
| NCE | Net-Centric Environment |
| NEW | Network Enabled Weapon |
| NLOS | Non-Line of Sight |
| NLOS-LS | Non-Line of Sight Launch System |
| NMS | National Military Strategy |
| NOLH | Nearly Orthogonal Latin Hypercube |
| NPS | Naval Postgraduate School |
| NSS | Naval Simulation System |
| ODUSD(A&T)SSE/DTE | Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering, Development Test and Evaluation |
| OMB | Office of Management and Budget |
| OneSAF | One Semi-Automated Forces |

| Acronym/Abbreviation | Meaning |
|---|---|
| OOS | OneSAF Objective System |
| OPLAN | Operational Plan |
| OPORD | Operation Order |
| ORSA | Operation Research Systems Analyst |
| OSD | Office of the Secretary of Defense |
| OSD (PA&E) | Office of the Secretary of Defense Program Analysis and Evaluation |
| OT | Operational Test |
| OT&E | Operational Test and Evaluation |
| OTA | Operational Test Agency |
| OTG | Over-the-Horizon Gold |
| OUSD(AT&L) | Office of the Under Secretary of Defense for Acquisition, Technology and Logistics |
| OUSD(P&R) | Office of the Under Secretary of Defense for Personnel and Readiness |
| OUSD(P) | Office of the Under Secretary of Defense for Policy |
| OV | Operational View |
| PA&E | Program Analysis and Evaluation |
| PAID | Process, Application, Infrastructure, Data |
| PDM | Physical Design Model |
| PDR | Preliminary Design Review |
| PI | Program Introduction |
| PID | Program Introduction Document |
| PM | Program Manager |
| PM's Handbook | Program Manager's Handbook for Testing in a Joint Environment |
| PMJ | Professional Military Judgment |
| PPBE | Planning, Programming, Budgeting, and Execution |
| PPBS | Planning, Programming, and Budgeting System |
| PTP | Program Test Plan |
| RCS | Restricted Cubic Spline |
| RCT | Regimental Combat Team |
| RGS | Requirements Generation System |
| ROE | Rules of Engagement |
| ROMO | Range of Military Operations |
| ROZ | Restricted Operating Zone |
| RPG | Recommended Practices Guide |
| RSM | Response-Surface Methodology |
| RTI | Runtime Infrastructure |
| RTO | Responsible Test Organization |
| SC | Statement of Capability |
| SCS | Simulation Collection System |
| SDD | System Design Document |
| SEED | Simulation Experiments and Efficient Design |
| SIPRNET | Secret Internet Protocol Router Network |

| Acronym/Abbreviation | Meaning |
|---|---|
| SME | Subject Matter Expert |
| SOA | Service Oriented Architecture |
| SOC | Statement of Capability |
| SoS | System of Systems; Systems of Systems |
| SPG | Strategic Planning Guidance |
| SQL | Standard Query Language |
| SSAA | System Security Authorization Agreement |
| STAR | System Threat Assessment Report |
| SUT | System under Test |
| SV | Systems and Services View |
| T&E | Test and Evaluation |
| TADIL | Tactical Digital Information Link |
| Task MOP | Task Measure of Performance |
| TBMCS | Theater Battle Management Core Systems |
| TCR | Test Concept Review |
| TD | Test Director |
| TEMP | Test and Evaluation Master Plan |
| TENA | Test and Training Enabling Architecture |
| TES | Test and Evaluation Strategy |
| TGOA | Test Goal, Objectives, and Approach |
| TIJE Roadmap | Testing in a Joint Environment Roadmap |
| TMOP | Task Measure of Performance |
| TOE | Table of Organization and Equipment |
| TOEL | Time Ordered Event List |
| TOV | Threat Operational View |
| TP | Test Plan |
| TPR | Test Plan Review |
| TRAC-MTRY | US Army Training and Doctrine Command Analysis Center in Monterey |
| TRR | Test Readiness Review |
| TSSG | Testing in a Joint Environment Roadmap Senior Steering Group |
| TTP | Tactics, Techniques, and Procedures |
| TV | Technical Standards View |
| UAS | Unmanned Aircraft System |
| UCP | Unified Command Plan |
| UJTL | Universal Joint Task List |
| UML | Unified Modeling Language |
| US | United States |
| USA | United States Army |
| USAF | United States Air Force |
| USC | United States Code |
| USD(AT&L) | Under Secretary of Defense for Acquisition, Technology and Logistics |
| USD(P) | Under Secretary of Defense for Policy |

| Acronym/Abbreviation | Meaning |
|---|---|
| USJFCOM | US Joint Forces Command |
| USMC | United States Marine Corps |
| USMTF | US Message Text Format |
| USN | United States Navy |
| V&V | Verification and Validation |
| VCJCS | Vice Chairman of the Joint Chiefs of Staff |
| VCNO | Vice Chief of Naval Operations |
| VCSA | Vice Chiefs of Staff of the US Army |
| VCSAF | Vice Chiefs of Staff of the US Air Force |
| VMF | Variable Message Format |
| VRT | Variance-Reduction Technique |
| VV&A | Verification, Validation, and Accreditation |
| WAN | Wide Area Network |
| WIPT | Working-level Integrated Product Team |
| XML | eXtensible Markup Language |
| XSD | XML Schema Definition |

*This page intentionally left blank.*

# ANNEX B - DEVELOP EVALUATION STRATEGY

## B.1 Purpose

This annex describes in detail the methodology for conducting Capability Test Methodology ***(CTM) 1.2, Develop Evaluation Strategy***, to include the development of critical capability issues (CCI) and critical operational issues (COI), development and refinement of the evaluation strategy, and identification and mitigation of risks. The development of evaluation strategy includes the processes needed to identify and define the factor test space and the measures framework that will be used to evaluate a capability under test.

## B.2 Overview

This process, shown in figure B-1, creates the capability evaluation strategy elements that are included in the Test and Evaluation Master Plan (TEMP). The evaluation strategy is developed based upon identified warfighter needs and includes the capability gaps, solutions, and implementations from the appropriate source documents (including Defense Planning Scenarios [DPS], Analytical Agenda, Initial Capabilities Document [ICD], and Capabilities Development Document [CDD]). The initial evaluation strategy document will be a Test and Evaluation Strategy (TES). As stated in the Defense Acquisition Guidebook, "TES is an early T&E planning document that describes the T&E activities starting with Technology Development and continuing through Engineering and Manufacturing Development into Production and Deployment. Over time, the scope of this document will expand; the TES will evolve into the TEMP due at Milestone B."

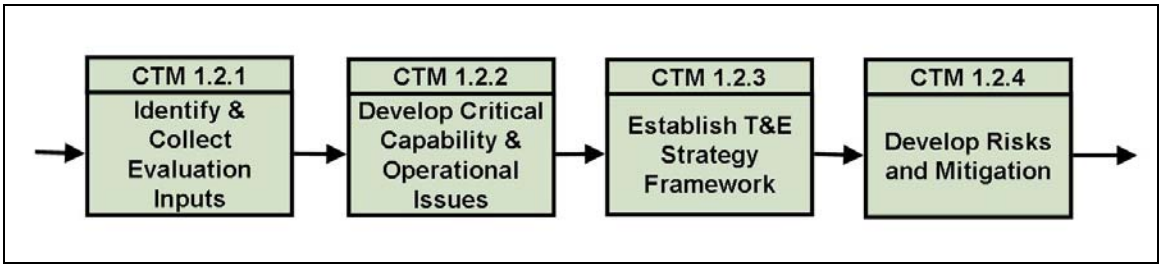



**Figure B-1. CTM 1.2, Develop Evaluation Strategy**

The first step in this process, ***CTM 1.2.1, Identify and Collect Evaluation Inputs***, involves the identification and collection of appropriate evaluation inputs. It is discussed in detail in section B.3.1 of this annex. Once all inputs have been collected, the analyst will derive CCIs and COIs from the warfighter needs in ***CTM 1.2.2, Develop Critical Capability and Operational Issues***. CCIs are used to assess SoS performance pertaining to capabilities that support joint missions and their impact on joint mission effectiveness. COIs are used to assess systems under test attributes and ability to perform tasks. CCI and COI development is discussed in detail in section B.3.2. The analyst will map these CCIs and COIs to an evaluation framework in ***CTM 1.2.3, Establish Test and Evaluation Strategy Framework***, that consists of initial test factors, measures, and relationships in order to demonstrate that warfighter needs are adequately addressed. Further discussions of developing the T&E framework can be found in section B.3.3 of this annex. Once the T&E strategy framework has been developed, the analyst will complete an initial assessment of risks and possible mitigations in ***CTM 1.2.4, Develop Risks and Mitigations***. In particular, the focus of the discussions in this handbook is on those risks

associated with the framework development and modeling limitations and assumptions identified during the previous steps. Further discussion of this process can be found in section B.3.4.

## B.3 Methods & Process for CTM 1.2

### B.3.1 Identify and Collect Evaluation Inputs – CTM 1.2.1

Because the overall evaluation process begins here, it is important to identify the collection of inputs explicitly as a process to ensure that the analysis team makes an explicit effort to identify and gather all of the necessary documentation as the first activity. The main steps in this process, representing the primary categories of input documents and stakeholders, are shown in figure B-2.

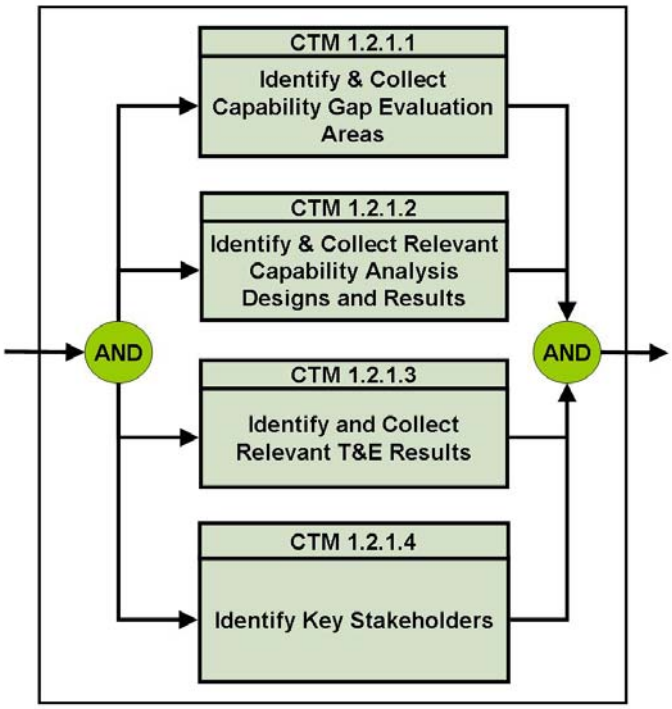



**Figure B-2. CTM 1.2.1, Identify and Collect Evaluation Inputs**

The overall evaluation for the system or system of systems (SoS) is developed based on identified warfighter needs including the capability gaps, solutions, and implementations from the appropriate source documents.

#### B.3.1.1 Capability Gap Evaluation Areas – CTM 1.2.1.1

Documentation from the JCIDS and related processes describe the capability gap in terms of the joint capability area (JCA), joint capability key characteristics, functional area(s), mission, desired effects, tasks, constraints, limitations, and key attributes across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) with appropriate measures. Potential JCIDS documents include:

- **DOTMLPF Change Request (DCR):** "The joint DCR is defined as a recommendation for changes to existing joint resources when such changes are not

associated with a new defense acquisition program." (CJCSI 3170.01G, 1 March 2009)

- **Initial Capabilities Document (ICD):** The ICD "[s]ummarizes a capabilities-based assessment and justifies the requirement for a materiel or non-materiel approach, or an approach that is a combination of materiel and non-materiel, to satisfy specific capability gap(s). It defines the capability gap(s) in terms of the functional area, the relevant range of military operations, desired effects, time and doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) and policy implications and constraints. The ICD summarizes the results of the DOTMLPF and policy analysis and the DOTMLPF approaches (materiel and non-materiel) that may deliver the required capability. The outcome of an ICD could be one or more joint DCRs or recommendations to pursue material solutions." (CJCSI 3170.01G, 1 March 2009)
- **Capability Development Document (CDD):** "A document that captures the information necessary to develop a proposed program(s), normally using an evolutionary acquisition strategy. The CDD outlines an affordable increment of militarily useful, logistically supportable, and technically mature capability. The CDD may define multiple increments if there is sufficient definition of the performance attributes (key performance parameters, key system attributes, and other attributes) to allow approval of multiple increments." (CJCSI 3170.01G, 1 March 2009)
- **Capability Production Document (CPD):** "A document that addresses the production elements specific to a single increment of an acquisition program. The CPD defines an increment of militarily useful, logistically supportable, and technically mature capability that is ready for a production decision. The CPD defines a single increment of the performance attributes (key performance parameters, key system attributes, and other attributes) to support a MS c decision" (CJCSI 3170.01G, 1 March 2009)
- **Analysis of Alternatives (AoA):** An AoA is an analytical comparison of the operational effectiveness, suitability, and lifecycle cost of alternatives that satisfy established capability needs. It is a vehicle used by senior leadership to debate and assess a program's desirability and affordability. Initially, the AoA process typically explores numerous conceptual solutions with the goal of identifying the most promising options, thereby guiding the concept refinement phase. An AoA is important in determining whether or not a system should be procured. It must make a case for having identified the most cost effective alternative and it must also make a compelling statement about the military utility of acquiring it.
- **Joint Capabilities Evaluation (JCE):** The JCE documents the analysis of one or more relevant capability test events used to support Milestone A, B, or C acquisition decisions.
- **Capability Gap Evaluation Areas:** The capability gap evaluation areas are developed in terms of the JCA, joint capability key characteristics, functional area(s), mission, desired effects, tasks, constraints, limitations, and key attributes across the spectrum of DOTMLPF with appropriate measures.

- **Attributes of Capabilities:** The capability gap discussion should also describe the attributes of the required capabilities in terms of desired outcomes. Where multiple characteristics are identified, they should be prioritized based on their value to delivering the capability within the context of the CONOPS described earlier.

The ICD will provide the concept of operations (CONOPS) and/or unified command plan (UCP)-assigned mission to which this capability contributes; what operational outcomes it provides; what effects it must produce to achieve those outcomes; how it complements the integrated joint warfighting force; and what enabling capabilities are required to achieve its desired operational outcomes. Section 4 of the related ICD for the system/SoS will provide the capability gaps in operational terms. The missions, tasks, and functions that cannot be performed or are unacceptably limited will be identified, as well as DOTMLPF impacts and constraints. Capability gaps will also be identified as being due to lack of proficiency in existing capabilities or due to lack of sufficient capabilities. This will provide the linkage between required capabilities and the key characteristics identified in the family of joint future concepts and/or CONOPS. Traceability between different aspects of capability gaps (such as mission effect to task to capability characteristic to KPP) is derived from authoritative sources. The focus is on providing information relative to mission and system-level operational effectiveness. These capability gaps will provide the basis to develop the CCIs and COIs.

#### B.3.1.2 Capability Analysis Designs and Results – CTM 1.2.1.2

These collected designs and results are compared to the joint operating concepts (JOC) to establish reference points for the initial design activity.

#### B.3.1.3 Relevant Test and Evaluation Results – CTM 1.2.1.3

Relevant test and evaluation (T&E) results can help scope factors and/or issues for the test design.

#### B.3.1.4 Key Stakeholders – CTM 1.2.1.4

There are a number of individuals and organizations that should provide input into the evaluation thread throughout the process. Such key individuals and organizations could include program managers (PM), capability managers (CM), Joint Requirements Oversight Councils (JROC), responsible test agencies, etc. It is critical to ensure that the right stakeholders are considered throughout the process to ensure buy-in to the evaluation framework.

### B.3.2 Develop Critical Capability & Operational Issues – CTM 1.2.2

#### B.3.2.1 <u>Inputs</u>

**Capability/System of Systems (SoS) Descriptions:** The initial capability description and SoS description describe the key features and subsystems, and hardware and software, for each increment's configuration. The SoS description will decompose the system into testable systems that can be integrated into the joint mission environment (JME). It includes a mapping of a capability configuration, its sub-systems, and relevant

systems under each sub-system. SoS attribute areas may also be identified from capability gaps and key characteristics. Information on the capability and SoS required for CCI development is covered in this document.

**Joint Operational Context for Test (JOC-T):** The JOC-T defines the operational organizations, resources, missions, and threats that interact with the SoS. The JOC-T annex includes several Department of Defense Architecture Framework (DoDAF) operational and system views. Task, conditions, and mission desired effects required for CCI development will come from the JOC-T products.

**Key Stakeholders:** There are a number of individuals and organizations that should have input into the identification of CCI. Such individuals and organizations might include PMs, CMs, JROCs, responsible test agencies, etc. It is critical to ensure that the right stakeholders are considered here, as the CCIs developed in this step provide the context for all subsequent experimentation and evaluation of the joint capability.

### B.3.2.2 Discussion

Testing must address the critical issues that can be used to assess performance as it pertains to capabilities supporting joint missions. COIs address SUT performance and capability contribution to SoS operational effectiveness; and CCIs address joint capability contributions to achieving mission desired effects.

CCIs and COIs should be carefully structured to address the key capability attributes described in joint capabilities documentation (ICD, CDD, etc.). It is important to state how the test issue contributes to achieving the desired mission end state in terms of mission desired effects. CCIs should address the SoS ability to accomplish joint operational tasks and/or the SoS, system, or service attribute performance; on the other hand, COIs focus on SUT abilities to perform tactical tasks.

The majority of the discussion in the subsequent paragraphs will focus on CCI development, as they are unique to joint testing. COI development is a normal part of Service-level testing, and, therefore, will be discussed only briefly in this handbook.

#### B.3.2.2.1 Critical Capability Issues (CCI)

CCIs are used to assess performance pertaining to capabilities that support joint missions. The essential elements of a CCI are represented in the Capability Evaluation Metamodel (CEM) as shown in figure B-3. They include a capability's essential tasks, mission desired effects, blue SoS (across DOTMLPF), and conditions involving threat and environmental factors. These essential elements are contained in the capabilities crosswalk, which is described in Annex E. There will probably be more than one CCI developed.

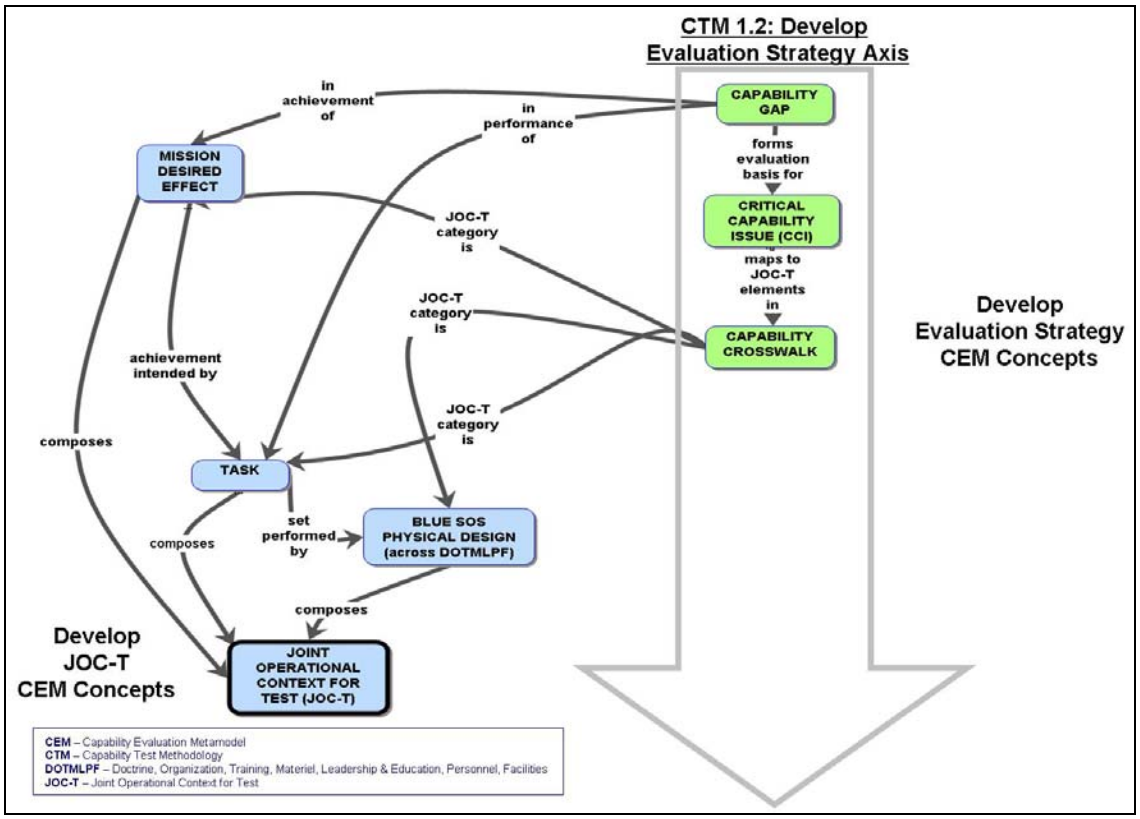



**Figure B-3. CEM Representation of CTM 1.2.2, Develop Critical Capability Issues**

It is important to state how the test issue contributes to achieving the mission end state outcomes in terms of mission desired effects. A CCI should address the SoS capability to perform joint operational tasks and/or the SoS, system, or Service attribute performance. CCIs are of primary importance to the appropriate authority when deciding whether to allow the SoS to advance to the next phase of development.

An example set of questions related to the development of a CCI is described below:

- Is the CCI used to assess performance as it pertains to capabilities supporting joint missions?
- Is the CCI carefully structured to address joint capability areas described in joint capability documentation?
- Does the CCI phrasing include SoS contribution to achieving the desired mission end state outcomes in terms of mission desired effects?
- Is the CCI addressing the SoS' ability to perform joint operational tasks and/or the SoS, system, or Service attribute performance?

An example CCI format that captures the essential elements would be:
***Assess the ability to perform Task X by SoS Configuration Y under Condition Set A to achieve Mission Desired Effect Z.***

Referring to the definition of a capability, "the ability to achieve a desired effect under specified standards and conditions through combinations of means and ways to perform a set of tasks," there is a need to tie together the desired effects, the means and ways (joint capability or SoS), and the joint tasks in a clear and consistent manner. The methodology draws on the previously discussed mapping of the evaluation framework to create CCI. Similar to operational test and evaluation (OT&E) critical operational issues, CCI form the basis to determine the SoS ability to perform its mission. A mapping of SoS components to UJTL to desired effects will give us the combinations to develop CCI for the joint capability.

Whether the CCI format is "how well, or can the, or assess the" is not as important as ensuring that the key elements and relationships in the CCI are captured. These key elements include the task, conditions, SoS configuration, and mission desired effects. The format of the CCI is dependent on the analysis conducted; as long as the key components are present, the format of a question or a statement is tailorable and dependent on the needs of the analyst.

#### B.3.2.2.2 Critical Operational Issues (COI)

COIs are the operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined in operational test and evaluation (OT&E) to evaluate and/or assess the system's capability to perform its mission. These test issues/questions should also identify those tasks (operational and tactical) identified by the ICD and/or CDD associated with the operational tasks that will be performed to accomplish the mission and ensure the intent of test is met. COIs include a description of the operational task threads and task interactions.

A COI is typically phrased as a question that must be answered in order to properly evaluate operational effectiveness (e.g., "Will the system detect the threat in a combat environment at adequate range to allow successful engagement?") and operational suitability (e.g., "Will the system be safe to operate in a combat environment?").[1]

Example COI:
***How capable is the X System of making in-flight trajectory corrections?***

#### B.3.2.3 <u>Output</u>

The primary outputs of this process are the CCIs and the COIs. Details of these outputs are discussed in the preceding sections.

### B.3.3 Establish Test and Evaluation Strategy Framework – CTM 1.2.3

The TES framework is refined through a metamodel-test-metamodel approach shown in Figure B-4. Once the "possible" framework space (likely to be characterized by a large number of test factors and interrelationships) has been defined, the test space initially described by all these factors must be refined and scoped down to a "probable" test space that

[1] The reader may also consult the following service-level evaluation and analyst guides for more information about COI: AFOTEC (2007), pp 2-8 through 2-10; OPTEVFOR (200 ), p 1-2

retains those test factors and/or factor interrelationships that are of significant impact to CCI-derived measures. These refined factors and interrelationships will provide the basis for the development of models for the test. This will allow for simulation of the test utilizing the test factors and test factor interrelationships in order to provide statistical output to:

- Validate the significance of test factors and interrelationships;
- Allow for analysis that will indicate where models need to be changed; and
- Indicate where other factors or interrelationships may need to be modeled.

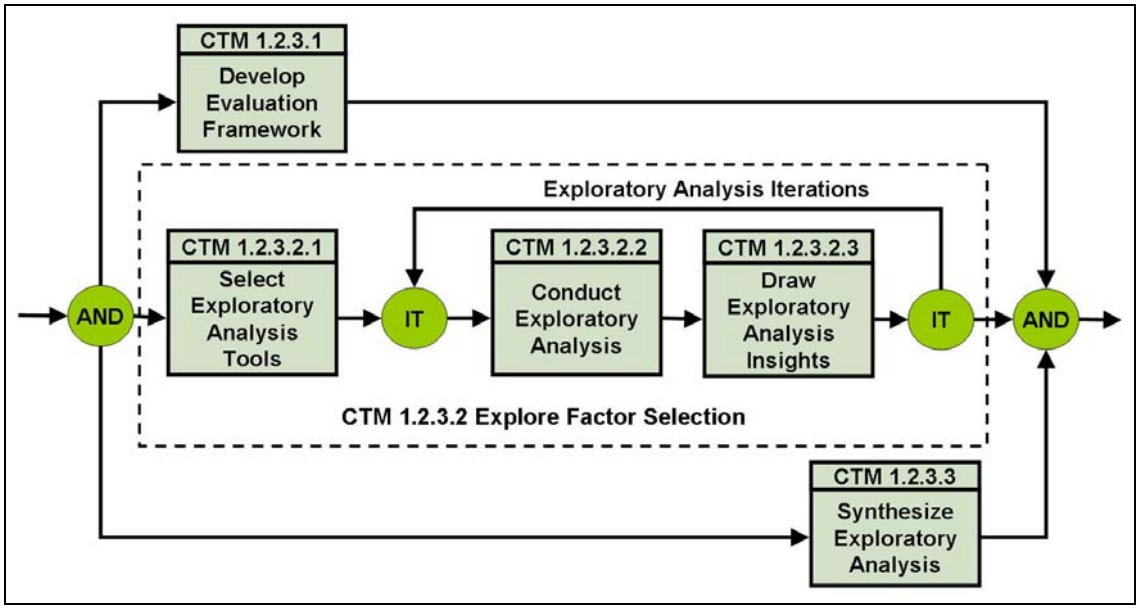



**Figure B-4. CTM 1.2.3, Establish Test and Evaluation Strategy Framework**

The initial Capability Test Design is developed in order to describe, populate, and refine the joint mission effectiveness (JMe) test space based on factors from a number of dimensions or axes. The essential CEM elements of the Capability Evaluation Strategy are shown in Figure B-5.

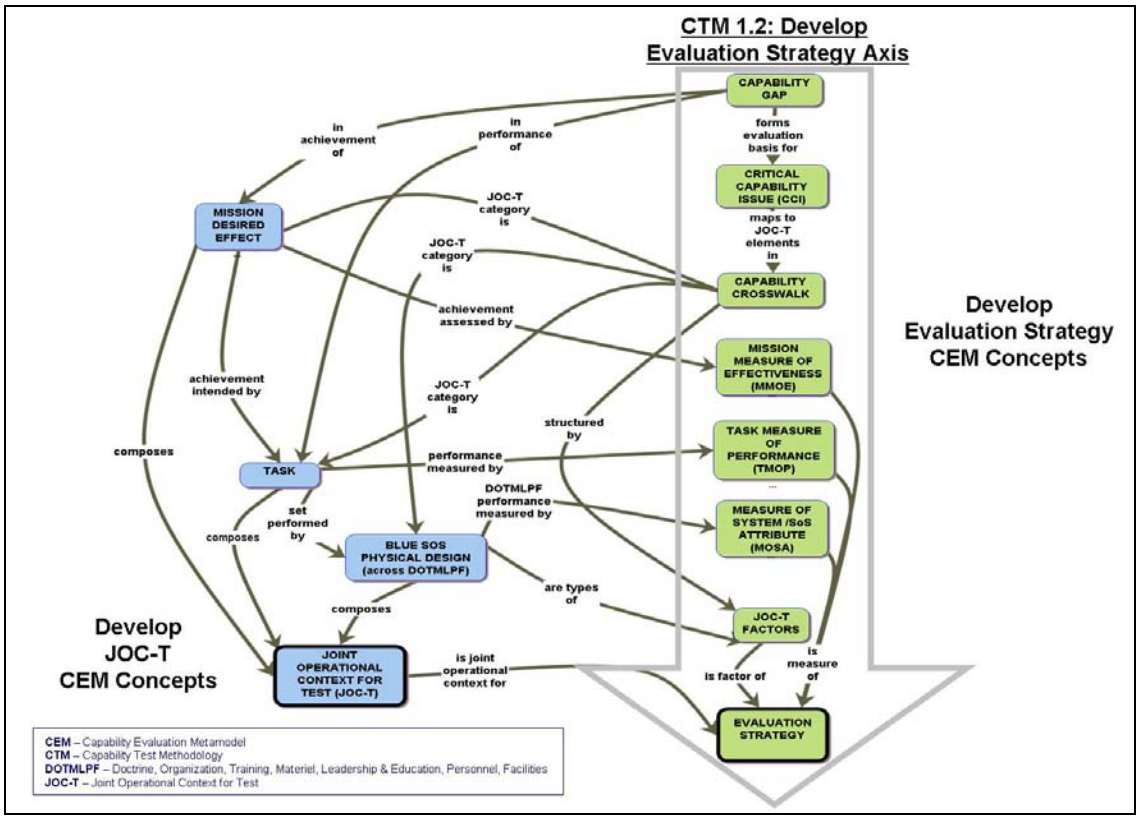



**Figure B-5. CEM Representation of CTM 1.2.3, Establish Test & Evaluation Strategy Framework**

### B.3.3.1 Develop Evaluation Framework

In ***CTM 1.2.3.1, Develop Evaluation Framework***, the analyst identifies and refines the test factors and levels expected to affect SoS performance, and the measures that will be used to evaluate and assess SoS performance. The main process is to identify and define the factor test space and the measure framework that will be used to evaluate the capability.

The development of the evaluation framework is a critical process that affects all subsequent analyses. It is critical to obtain subject matter expert (SME) and key stakeholder input here. It consists of two primary steps as shown in Figure B-6: (1) identify test factors and levels and (2) identify test measures. The primary inputs to this process include the JOC-T products from which many of the factors can and should be drawn (DoDAF products, Blue and threat system of system descriptions, appropriate joint mission products, identification of conditions); the CCI, which focus the framework development by defining the appropriate joint capability, missions, tasks, and conditions; and the system/SoS descriptions, which describe the Blue SoS across DOTMLPF. The primary outputs of this step are the initial JMe factor test space that captures the set of potential factors affecting JMe and the test measure framework that defines how the exploratory analysis and subsequent tests will evaluate JMe.

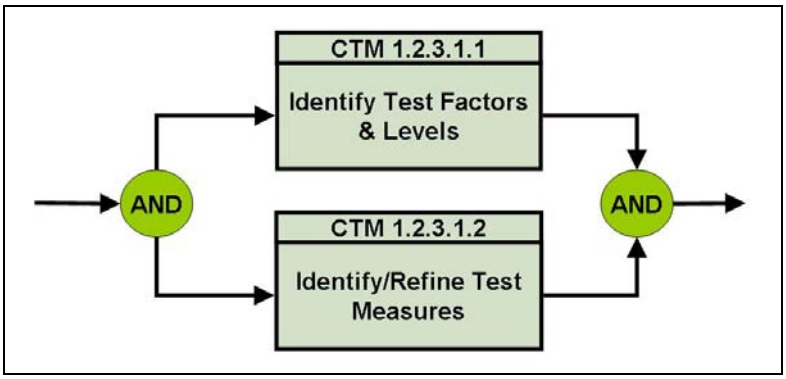



**Figure B-6. CTM 1.2.3.1, Develop Evaluation Framework**

The purpose of ***CTM 1.2.3.1.1, Identify Test Factors and Levels***, is to identify the relevant factors within the three dimensions of joint mission, capability SoS solution, and joint mission conditions; and to determine the possible ranges or levels through which the factors may vary.

In ***CTM 1.2.3.1.2, Identify/Refine Test Measures***, the team identifies the three types of measures that will be used to assess JMe: mission measures of effectiveness (Mission MOE), task measures of performance (Task MOP), and SoS attributes; determines the feasible regions or levels for the measures; and develops a structure to define the relationships between the measure types.

### B.3.3.1.1 Identify Test Factors and Levels – CTM 1.2.3.1.1

#### B.3.3.1.1.1 Inputs:

**Key Joint Capability Integration Development System (JCIDS) Documents:** These documents are the products of various steps within the JCIDS process. Not all documents will be appropriate for all SoS involved in a joint T&E.

- DOTMLPF Change Request (DCR)
- Initial Capabilities Document (ICD)
- Capability Development Document (CDD)
- Capability Production Document (CPD)

Department of Defense Architecture Products: This section provides a brief description of each of the relevant DoDAF products. All descriptions are taken verbatim from *DoD Architecture Framework, Version 1.5, Volume I: Definitions and Guidelines*, 23 April 2007. DoDAF products should be outputs of the Joint Operational Context for Test (JOC-T) development process, although not all of these products are required as part of the CTM. The following are the DoDAF operational view (OV) products that may provide key inputs to the evaluation framework development process. These products can be developed for either blue forces (BOV) or threat forces (TOV). See Annex H for more detailed information on DoDAF products.

- **OV-1, High-Level Operational Concept Graphic:** High-level graphical/textual description of operational concept.
- **OV-2, Operational Node Connectivity Description:** Operational nodes, connectivity, and information exchange need lines between nodes.
- **OV-4, Organizational Relationships Chart:** Organizational, role, or other relationships among organizations.
- **OV-5, Operational Activity Model:** Capabilities, operational activities, relationships among activities, inputs, and outputs; overlays can show cost, performing nodes, or other pertinent information.
- **OV-6a, Operational Rules Model:** One of three products used to describe operational activity - identifies business rules that constrain operation.
- **OV-6b, Operational State Transition Description:** One of three products used to describe operational activity - identifies business process responses to events.
- **OV-6c, Operational Event-Trace Description:** One of three products used to describe operational activity - traces actions in a scenario or sequence of events.
- **OV-7, Logical Data Model:** Documentation of the system data requirements and structural business process rules of the operational view.

The following are the DoDAF systems and services view (SV) products that may provide key inputs to the evaluation framework development process. These products can be developed for either blue forces (BSV) or threat forces (TSV).

- **SV-1, Systems/Services Interface Description:** Identification of systems nodes, systems, system items, services, and service items and their interconnections, within and between nodes.
- **SV-2, Systems/Services Communications Description:** Systems nodes, systems, system items, services, and service items and their related communications laydowns.
- **SV-3. Systems-Systems/Services-Systems/Services-Services Matrix:** Relationships among systems and services in a given architecture; can be designed to show relationships of interest, e.g., system-type interfaces, planned vs. existing interfaces, etc.
- **SV-4a, Systems Functionality Description:** Functions performed by systems and the system data flows among system functions.
- **SV-4b, Services Functionality Description:** Functions performed by services and the service data flow among service functions.
- **SV-5a, Operational Activity to Systems Function Traceability Matrix:** Mapping of system functions back to operational activities.
- **SV-5b, Operational Activity to Systems Traceability Matrix:** Mapping of systems back to capabilities or operational activities.

- **SV-5c, Operational Activity to Services Traceability Matrix:** Mapping of services back to operational activities.
- **SV-6, Systems/Services Data Exchange Matrix:** Provides details of system or service data elements being exchanged between systems or services and the attributes of that exchange.
- **SV-7, Systems/Services Performance Parameters Matrix:** Performance characteristics of systems and services view elements for the appropriate time frame(s).
- **SV-8, Systems/Services Evolution Description:** Planned incremental steps toward migrating a suite of systems or services to a more efficient suite, or toward evolving a current system to a future implementation.
- **SV-9, Systems/Services Technology Forecast:** Emerging technologies and software/hardware products that are expected to be available in a given set of time frames and that will affect future development of the architecture.
- **SV-10a, Systems/Services Rules Model:** One of three products used to describe system and service functionality - identifies constraints that are imposed on systems/services functionality due to some aspect of systems design or implementation.
- **SV-10b, Systems/Services State Transition Description:** One of three products used to describe system and service functionality - identifies responses of a system/service to events.
- **SV-10c, Systems/Services Event-Trace Description:** One of three products used to describe system or service functionality - identifies system/service-specific refinements of critical sequences of events described in the operational view.
- **SV-11, Physical Schema:** Physical implementation of the Logical Data Model entities, e.g., message formats, file structures, physical schema.

The following are the DoDAF all view (AV) products that may provide key inputs to the evaluation framework development process. These products can be developed for either blue forces (BAV) or threat forces (TAV).

- **AV-1, Overview and Summary Information:** Scope, purpose, intended users, environment depicted, analytical findings.
- **AV-2, Integrated Dictionary:** Architecture data repository with definitions of all terms used in all products.

The following are the DoDAF technical standards view (TV) products that may provide key inputs to the evaluation framework development process. These products can be developed for either blue forces (BTV) or threat forces (TTV).

- **TV-1, Technical Standards Profile:** Listing of standards that apply to systems and services view elements in a given architecture.

- **TV-2, Technical Standards Forecast:** Description of emerging standards and potential impact on current systems and services view elements, within a set of time frames.

**Other Documents:**

- Approved Analyses of Alternatives (AoA) Plan and results, if available.
- Combatant Command (COCOM) operations plans/orders (OPLAN/OPORD).
- Universal Joint Task List (UJTL), found in CJCSM 3500.04D.
- Other test plans and/or relevant results from developmental testing (DT), operational testing (OT), and live-fire testing (LF).

**Critical Capability Issues (CCI):** CCIs are the direct output of *CTM 0.3.2, Develop Critical Capability and Operational Issues*.

#### B.3.3.1.1.2 Stakeholders

- Project Managers (PM)
- Capability Managers (CM)
- Service Test Agencies

#### B.3.3.1.1.3 Factor Identification

The purpose is to identify the overall test factor space for exploration. This is a critical step, as it is the foundation for both the exploratory analysis and the subsequent testing and experimentation. Ultimately, the goal of the factor exploration is to refine the possible JMe test space, likely to be characterized by an unmanageably large number of test factors and interrelationships, into a refined JMe test space. The refined JMe test space consists of those test factors and/or interrelationships that appear to have some significant impact based on simulation output and analysis, review by subject matter experts, and input from COCOMs on critical concerns. A critical preliminary step in this analysis is to identify the possible factors in the JMe.

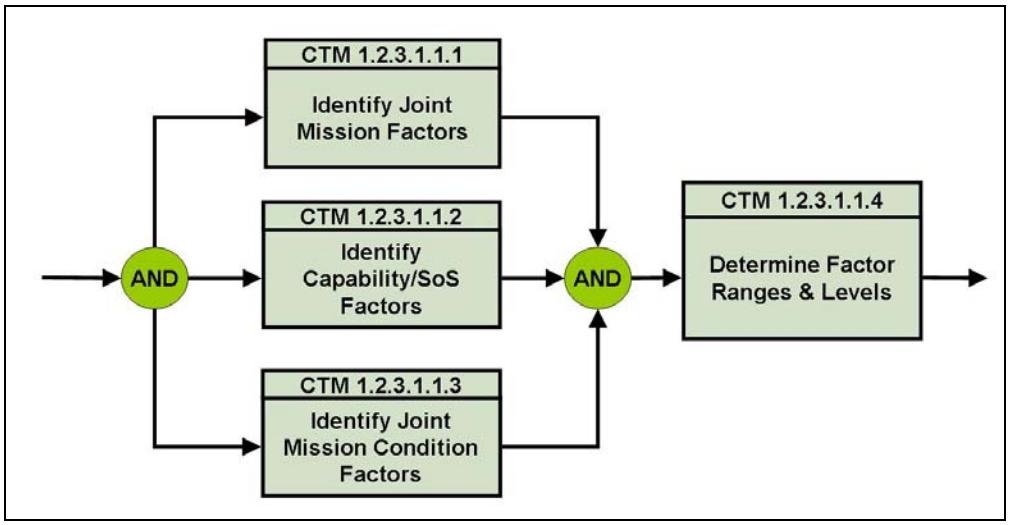



**Figure B-7. CTM 1.2.3.1.1, Identify Test Factors and Levels**

#### B.3.3.1.1.4 <u>Checklist for Factor Identification</u>

- Identify the joint mission and tasks.
  - Identify the joint mission objectives.
  - Identify the joint mission desired effects.
  - Identify the joint mission end state.
  - Identify the joint tasks.
  - Identify possible combinations of missions and tasks.
  - Identify the source.
- Identify blue SoS factors:
  - Identify blue SoS across DOTMLPF.
  - Identify the SoS attributes associated with each.
  - Identify the possible levels/ranges for each of the attributes/factors, if applicable.
  - Identify the combinations of Blue SoS to be considered for testing.
  - Identify the source.
- Identify threat SoS factors:
  - Identify threat SoS across DOTMLPF.
  - Identify the SoS attributes associated with each.
  - Identify the possible levels/ranges for each of the attributes/factors, if applicable.
  - Identify the combinations of threat SoS to be considered based upon authoritative guidance.
  - Identify the source.
- Identify physical environment factors:
  - Identify the key physical environment factors.

- Identify the possible levels/ranges for each of the factors, if applicable.
- Identify the combinations of physical environment factors to be considered for the subsequent analysis.
- Identify the source.

- Identify civil environment factors:
  - Identify the civil environment factors.
  - Identify the possible levels/ranges for each of the factors, if applicable.
  - Identify the combinations of civil environment factors to be considered based upon authoritative guidance.
  - Identify the source.
- Update the capabilities crosswalk.

#### B.3.3.1.1.5 Factor Dimensions

Factors are the independent variables of a joint capability that, if changed, may have an impact on JMe (the response or dependent variable). The JMe test space is characterized based upon factors from three dimensions:

- **Joint mission** dimension, which includes the objectives, desired effects, and end state factors;

- **Capability SoS solution** dimension, which includes system/SoS configuration options across DOTMLPF resources; and

- **Joint mission conditions** dimension, which includes threat conditions (e.g., threat actions, threat order of battle, threat command and control structure, threat systems, threat force disposition) and environmental conditions (e.g., physical and civil environment).

The factor dimensions above should help the analyst think through and identify the possible list of factors; the categories themselves are not critical in their own right. It may be helpful, at the mission level, for the analysis team to think of the various mission levels (e.g., tactical, operational/campaign, strategic), in order to explore that dimension fully.

The primary challenge in identifying factors is scoping the initial set to an appropriate set of factors based upon JOC-T, CCI development, etc. The analysis team must use the previous scoping efforts from the JOC-T and CCI development as an aid to factor identification. There may be a tremendous number of factors that can affect the SoS JMe, but identification of all of those factors is both problematic and excessive. The JOC-T and the CCIs can be used to identify key factors of interest, eliminate potential factors that do not fall within the experimental context of interest, and to set levels and ranges on the factors. Factor identification is more of an art than a science; therefore, it is hard to define procedures, heuristics, preset categories, etc., to ensure that all key factors are

identified. Use of the three dimensions above, along with the CCIs and JOC-T, is an appropriate starting point. In all cases, the analysis team should get input and feedback from key stakeholders to ensure that a comprehensive set of factors has been identified. For each factor identified, the analyst should document its source to ensure traceability throughout the capability test design process.

#### B.3.3.1.1.6 Factor Characterization

For each of the identified factors, the analyst should characterize the factor as either quantitative or qualitative. Quantitative factors are those whose potential levels are numerical, with values being either continuous (e.g., temperature, speed, and maximum effective range) or discrete (e.g., number of enemy artillery systems and number of system operators) in nature. Qualitative factors are those whose potential levels are not numerical, such as mission type (e.g., force entry operations, support operations, and stability operations), terrain type (e.g., desert, jungle, urban, and open), training level (e.g., well-trained, moderately trained, and untrained), binary (is the system present or not), or a type factor (one of a set of potential TTP). For all factors, the analyst should also identify the measurement units (e.g., miles per hour, kilobits per second, meters). Characterizing factors in this way is critical for the next step of determining levels and ranges, and for developing experimental designs for exploratory analyses and SoS experimentation.

Additionally, the analyst should characterize each factor as either controllable or uncontrollable. Controllable factors are those whose levels can be manipulated for live experimentation or the decision-maker, and could include such examples as the system configuration, number of systems, terrain type, and platform speed. Uncontrollable factors cannot be manipulated intentionally during live experimentation or the decision maker. Weather is a good example of an uncontrollable factor. It is important to note that the same factor may be controllable under certain experimental circumstances and not in others. For example, soldier experience level may be controllable in experiments for which the test team is able to choose or screen potential participants, but not for experiments for which the participant population is being provided by an external organization. Knowing the controllability of the factors serves three purposes:

- First, it identifies those factors that might be candidates for the constructive or virtual testing venues within the LVC-DE. Those venues permit manipulation of factors that may not be controllable in a live environment.
- Second, it sets potential bounds on the actual LVC test trial sets, which should not include factors that cannot be controlled during the experiment.
- Third, it identifies those factors that, regardless of their importance, cannot be affected by the decision maker in the implementation of the materiel or non-materiel solution.

It should also be noted that factors may be either controllable or uncontrollable in simulated environments as well; however, it may be premature to make that

determination in this step before potential models and simulations are identified. Nonetheless, this type of controllability is included in the table below for completeness, but may not be completed until later.

In order to keep track of the factor characterization, the team should use some sort of matrix as shown in table B-1. The table can either be integrated with the capability crosswalk or developed in isolation.

**Table B-1. Example Factor Characterization Tracking Matrix**

| Factor | Description | Units | Type | Numerical Type | Controllability | | |
|---|---|---|---|---|---|---|---|
| | | | | | By DM | In M&S | In Test Event |
| Training Level | The training level of the brigade staff | High, medium, low | Qualitative | N/A | Y | Y | N |
| Staff size | The number of personnel on the brigade staff | Number | Quantitative | Discrete | N | Y | Y |
| … | … | … | … | … | … | … | … |

#### B.3.3.1.1.7 Factor Ranges and Levels

In this step, the analyst attempts to set the range through which each factor may vary and/or identify the levels that each factor may attain. For a continuous, quantitative factor, setting the upper and lower limits on the factor range may be sufficient, since the levels that the factor may attain can be any value between the limits. The same may be true for discrete values as long as the intervals between the discrete values are consistent and can be defined. For certain discrete values and most qualitative values, the analyst may need to identify and define each possible level. By doing so, the range is also determined.

Setting the range/levels for quantitative factors is always challenging and, ultimately, somewhat subjective. One has to balance the desire to explore the greatest possible range of values with the time and resource limitations on the exploratory analysis. One technique that can be used is to set the limits at levels deemed realistic by the analysis team. Another technique is to set the limits approaching extreme, but not unrealistic or infeasible, values. In some cases, the limits may be known from previous testing and experimentation involving the same or similar systems. In such cases, those limits may be used. In other cases, the limits may be taken from documentation. For example, if the system under test already has a Capabilities Development Document (CDD), the analyst may use the threshold or objective values defined for system parameters in the CDD as the upper limit on the factor. The setting of levels is a different, but related, issue to that of setting ranges. The levels may depend on the type of experiment design to be used during the analysis. For categorical/quantitative factors, the levels will be defined by the categories of possible responses. In general, the identification of levels and ranges is more of an art than a science in many cases. This is true as well for the identification of the factors of interest in the first place.

#### B.3.3.1.2 Measure Framework Development – CTM 1.2.3.1.2

The other main aspect of the evaluation framework development is the development of the measures that will be used to support evaluation of CCI. Figure B-8 shows the steps involved in this process. The test measures form the basis of the capability measures framework. This should be an explanation of how the test will measure both technical and operational effectiveness of the SoS. The objective of test measures is to provide metrics that support evaluation to determine the SoS enhancement to the current force capabilities. Information gained by the evaluation will provide feedback to the acquisition lifecycle mangers, and help mitigate development and other programmatic risks. The evaluation strategy should indicate a decision point of focus, (e.g., Milestone [MS] B, MS C, Low Rate Initial Production [LRIP] decision, etc.) for the evaluation data. The following sections describe the three levels of measures for joint testing, including the determination of the feasible ranges and levels for the measures.

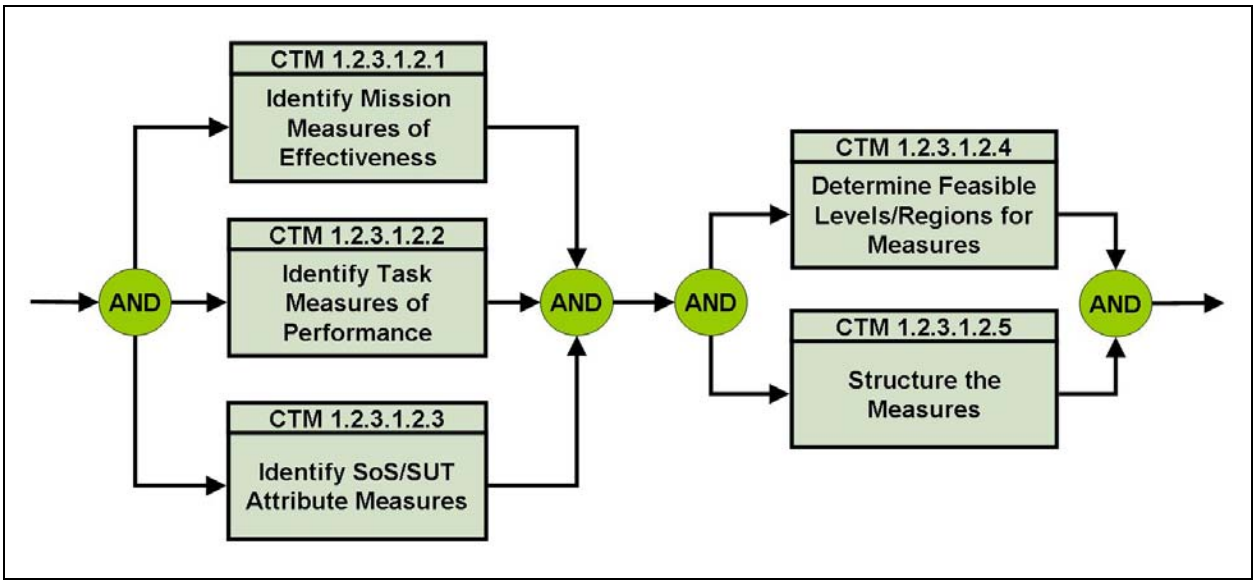



**Figure B-8. CTM 1.2.3.1.2, Identify/Refine Test Measures**

JMe itself is not a single response based upon factor levels and combinations. It is more likely a framework of multiple measures that, combined, provide insight into the effectiveness of a system of systems in a joint environment. Thus, it is important to identify the relationship of the measure levels.

##### B.3.3.1.2.1 Checklist

- Identify mission measure of effectiveness (Mission MOE):
  - Fully define the measure.
  - Identify the required data elements and define key terms.
  - Identify how the measure will be calculated and the units of measure.
  - Identify the desired fidelity of measurement.
  - Identify the expected feasible range of values and success criteria.

- Identify task measure of performance (Task MOP):
  - Fully define the measure.
  - Identify the required data elements and define key terms.
  - Identify how the measure will be calculated and the units of measure.
  - Identify the desired fidelity of measurement.
  - Identify the expected feasible range of values and success criteria.
- Identify measures of System/SoS attributes (MOSA):
  - Fully define the measure.
  - Identify the required data elements and define key terms.
  - Identify how the measure will be calculated and the units of measure.
  - Identify the desired fidelity of measurement.
  - Identify the expected feasible range of values and success criteria.
  - Identify the expected relationships between the measures.
  - Identify the expected relationships between measures at the same level.
  - Identify the expected relationships between measures of differing levels.

#### B.3.3.1.2.2. Measure Identification

The three levels of measures used to evaluate joint capabilities are:

- Mission measures of effectiveness (Mission MOE)
- Task measures of performance (Task MOP)
- Measures of SoS/SUT attributes (MOSA)

Figure B-9 shows the relationships of these three levels of measures and their impact on the CCIs/COIs and, ultimately, to the capability under evaluation.

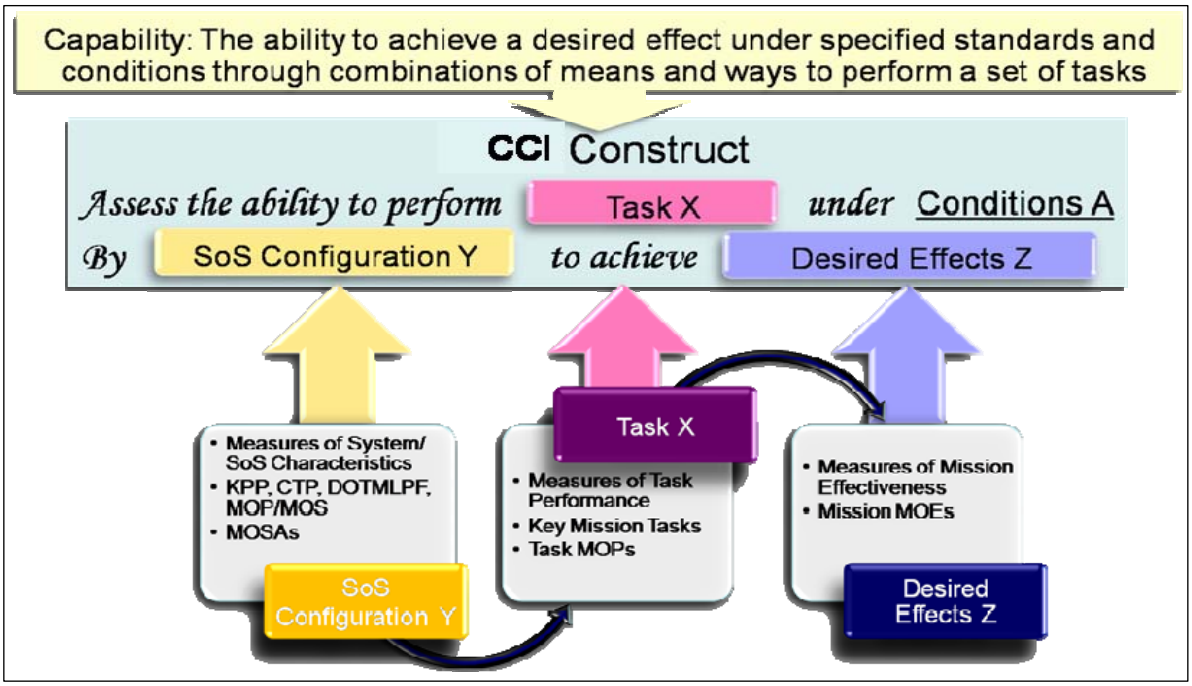



**Figure B-9. Relationship of the Measure Levels to the CCI/COI**

**B.3.3.1.2.2.1. Mission Measure of Effectiveness (Mission MOE)**
Mission MOEs are used to assess changes in system behavior, capability, or operational environment tied to measuring the attainment of a mission end state, achievement of a mission objective, or creation of a mission desired effect.

**Sources for Mission MOE** - The ICD specifies the need for a materiel approach to a capability gap derived from an initial analysis of materiel approaches executed by the operational user and, as required, an independent analysis of materiel alternatives. When trying to determine the desired effect and the measures of effectiveness for a system or system of systems it is important to reference this document (or its derivatives such as the CDD) as it defines the capability gap in terms of the functional area, the relevant range of military operations, desired effects, and time.

**Mission Desired Effects** - Desired effects identification/authoritative source description includes the ICD and the analytic agenda. The mission desired effect is the physical or behavioral state of a system that results from an action, a set of actions, or another effect. The desired effect for the system of systems may be different than the individual systems that make up the system of systems. However, the desired mission effect should be the result outcome, or consequence, that the commander desires to achieve by performing a mission, which could be a change to a condition, a behavior, or a degree of freedom change for either blue or threat forces. Mission desired effects are identified and derived from an authoritative source (i.e., analytic agenda, etc). Once the mission desired effects are defined as a condition for achieving an

associated strategic or operational objective, the Universal Joint Task List (UJTL) is used to determine the necessary tasks to accomplish the mission.

**Mission MOE Directions -** The Mission MOEs focus on mission accomplishment and mission utility for the SoS. The Mission MOEs should serve as the higher-level measures for the SoS or SUT. Sections 4b and 4d of the ICD should describe the capabilities in terms of desired outcomes, how they complement the integrated joint warfighting force, and indicate the minimum value at which the SUT/SoS capability will no longer be effective. When detailing the Mission MOEs for the SoS it is important to trace each specific MOE back to the appropriate desired effect as shown in Figure B-9. When detailing the Mission MOE for the SUT, the measures should be designed to correspond to the accomplishment of the systems mission, desired end state and the systems desired effects.

Mission MOE development is more of an art than a science. Whereas many of the Task MOP and MOSA may be found in the UJTL and required JCIDS documentation, respectively, Mission MOE may not be found in documentation. The analysis team must have a complete understanding of the mission desired effects and be creative in developing measures that will truly assess the achievement of those effects. Fortunately, mission-level measures are not a new concept in general. Such measures are often developed in support of analyses of alternatives (AoA) and other studies; however, the team must ensure that previously-developed measures are appropriate in the specific context under study before deciding to include them. These measures may be both qualitative and quantitative in nature, and should be developed in isolation of the other types of measure.

Determining the feasible ranges and levels for a Mission MOE can be quite challenging. These higher level measures rarely have threshold or objective values associated with them, often because the Mission MOE must be viewed in the context of a particular scenario or set of conditions before one can assess the goodness of the results. In rare cases, threshold and objective values may be delineated in authoritative documentation. If so, those can be used. Another technique is to use the current baseline SoS performance as a threshold against which to compare the performance of the new capabilities. However, in many cases, even these values may be unknown or unobtainable, particularly for unique measures developed specifically for a particular set of capability test events. Another technique is to query key operational stakeholders to determine what the expected and desirable ranges may be. This has the advantage of garnering user acceptance for the measures and thresholds. In some cases, it may be more appropriate not to assign an initial threshold for the measure range, but to use the measure responses to compare different SoS configurations under differing conditions. Thus, the measure would assume a comparative role instead of the role as a success/fail criterion for the SoS or capability.

#### B.3.3.1.2.2.2 Task Measure of Performance (Task MOP)

Task MOPs are used to quantify mission task accomplishment. Task MOPs are defined using the UJTL structure of task, condition, and standard and are based on joint force commander mission tasks deemed essential to mission accomplishment using specified conditions and standards.

**Sources for Task MOP -** The analysis of alternatives document contains mission tasks in the alternatives section. A task is an action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability or a directive statement used to assign a discrete action or set of actions to an organization that enables a mission or function to be accomplished. A single task may incorporate multiple individual actions. Joint tasks describe what tasks a system will be expected to perform to accomplish the mission. These performance tasks enable test personnel to evaluate the proposed systems based on specific requirements. The joint tasks are general statements of task performance requirements. Where possible, the tasks are described in quantitative terms. These tasks become the standards for comparing alternative systems.

Task identification can be logically broken down into several sequential steps:

1) Identification of key events;
2) Primary technical tasks and their entry and exit criteria;
3) Needed support tasks; and
4) Management tasks for accomplishment of primary and support tasks.

Some examples of mission tasks include:

- Conduct deep interdiction precision air strikes against heavily defended targets
- Conduct air strike operations in a stealth mode.

The primary authoritative source for joint tasks is the UJTL (CJCSM 3500.04D). This document should be the primary source for determining the key joint tasks required to achieve the mission desired effects, and the standards associated with the tasks. In some cases, particularly for a unique capability, there may be one or more required tasks that cannot be found in the UJTL. In those cases, the analysis should identify those tasks and document their source.

**Task MOP Directions -** A Task MOP is a criterion used to assess friendly actions tied to measuring task accomplishment. Critical parameters selected from the MOP will be key indicators of overall system or SoS performance. If a Task MOP is not met, it may prevent the associated MOE from being satisfied and thus put the program/project at cost, schedule, or performance risk. A Task MOP should be written as a criterion used to assess changes in

system behavior, capability, or operational environment. It is a directly measurable parameter that can be specifically tested and measured during the test or evaluation. The MOP is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect.

The initial source of information about Task MOP and associated criteria is the UJTL. For each joint task, the UJTL identifies measures for describing varying levels of joint task performance and discusses the process for developing criteria that define successful levels of performance. The UJTL considers standards, the combination of measures and criteria, and describes the process for developing these standards as follows:

*The standard for a joint task is set within the framework of the JFC's mission and in the context of the conditions, either most likely or worst case, that are linked to those missions. Thus, the standard(s) for a joint task can only be set when (1) the mission analysis is complete, (2) the conditions affecting the task have been identified and described, and (3) measures and criteria have been selected that reflect the task contribution to mission accomplishment. This means that standards are tied to missions. That is, just because a joint task has a particular standard on one mission does not mean that the same standard will apply to other missions. A task standard could be the same across missions, but it could also be different for each mission. (CJCSM 3500.04D CH 1, B-B-1)*

While the analysis team can use the measures defined in the UJTL, they must still look at the measures in the context of the mission by considering the mission desired effects and conditions before assigning the criteria or thresholds for assessing task performance. The UJTL includes additional, valuable discussions about selecting measures and criteria in its Appendix B, Enclosure B. There are a few techniques for determining feasible ranges and levels for Task MOP. One technique is to use the current baseline SoS task performance as a threshold against which to compare the performance of the new capabilities. However, in many cases, these values may be unknown or unobtainable, particularly for unique measures developed specifically for a particular set of capability test events. Another technique is to query key operational stakeholders to determine the expected and desirable ranges. This has the advantage of garnering user acceptance for the measures and thresholds. In some cases, it may be more appropriate not to assign an initial threshold for the measure range, but to use the measure responses to compare different SoS configurations under differing conditions. The measure would assume a comparative role instead of the role as a success/fail criterion for the SoS or capability.

#### B.3.3.1.2.2.3 Measure of System/SoS Attributes (MOSA)

The lowest level of test measures are system and SoS attributes, which measure SUT and SoS performance, and include critical test parameters (CTP) and key performance parameters (KPP). These are used to assess SoS and SUT performance and limitations based on specified system requirements and additional characteristics such as reliability, maintainability, human factors, and susceptibility characteristics. A list of commonly used system and SoS characteristics and attributes are provided as a reference in Annex M.

MOSA constitute many of the measures traditionally assessed in system T&E, and may include system-level measures of effectiveness and measures of performance. The analyst must develop these measures within the context provided by the CCI, by referencing the joint capability documents, consulting stakeholders, and conducting other appropriate background research.

When identifying a measure, the analyst must ensure that the measure is completely and unambiguously defined in a way that can be understood and interpreted similarly by anyone in the expected audience for the results. Additionally, as with the discussion above for identifying the factors, the analyst should determine and document whether the measure is quantitative or qualitative in nature. The analyst should identify the method of calculation (if quantitative) or determination (if qualitative) and the unit of measure; identify the data elements required to calculate or determine the measure response value (quantitative) or level (qualitative); define the key terms; and identify the desired fidelity of measurement (e.g., measured to the nearest hour).

The analysis team must avoid the temptation to identify a large number of measures just because they can. The reason for this will become especially evident during the exploratory and post-test analyses when the analyst must evaluate and make sense of all of the measures included in the evaluation framework. The measures chosen should only be those that are most important for evaluating JMe based upon mission, task, and system/SoS performance.

**Sources and Directions for System/System of System Attributes -** SoS performance attributes are attributes so significant they must be verified by testing and evaluation/analysis. The CDD and CPD state the operational and support-related performance attributes of a system(s) that provide(s) the capabilities required by the warfighter. Each attribute will be supported by an operationally oriented analysis to determine threshold and objective values. Below the threshold value, the military utility of the system(s) becomes questionable. In an evolutionary acquisition, it is expected that threshold values will generally improve between increments. Different attributes may come into play as follow-on increments deliver additional capability. An attribute may apply to more than one increment. The threshold and objective values of an attribute may differ in each increment. DoD components will, at a minimum, budget to achieve all stated thresholds.

***Key Performance Parameters (KPP)***

KPPs are those attributes or characteristics of a system that are considered critical or essential to the development of an effective military capability and those attributes that make a significant contribution to the key characteristics as defined in the Joint Operations Concepts (JOpsC). The determination of criticality should be vetted as part of the JCIDS process by the responsible parties and is validated by the MDA. KPPs should be identified in the ICD for the SUT and SoS along with specific qualities and attributes that the system/SoS must meet. KPPs are also identified in the CDD/CPD, and included verbatim in the acquisition program baseline (APB). The Joint Requirements Oversight Council (JROC) validates the KPPs for acquisition category (ACAT) I programs and any other program designated as "JROC interest." All KPPs must be verified by testing and evaluation. A KPP will normally be a rollup of a number of supporting attributes that may be traded off to deliver the overall performance required.

Only those few KPPs, recommended to be eight or fewer by CJCSM 3170.01 (E-4), that capture the minimum operational effectiveness and suitability attributes needed to achieve the overall desired capabilities for the system(s) during the applicable increment are considered to be system or SoS KPPs. Because of its critical nature, failure to meet a KPP threshold may result in a reevaluation, reassessment or termination of the program, or a modification of the content of production increments.

A KPP will normally be a rollup of a number of supporting attributes or key system attributes (KSA) that may be traded off to deliver the overall performance required.

***Key System Attributes (KSA)***

KSAs are those system attributes considered most critical or essential for an effective military capability but not selected as a KPP. KSAs provide decision makers with an additional level of capability prioritization below the

KPP but with senior sponsor leadership control (generally 4-star level, Defense agency commander, or Principal Staff Assistant). In the case of the mandated Sustainment KPP (Materiel Availability), the supporting Materiel Reliability and Ownership Cost KSAs require any changes to be documented in the subsequent update to the acquisition program baseline. KSAs do not apply to the net-ready KPP (NR-KPP).

The threshold value for an attribute is the minimum (e.g., effective range for a munition delivery system) or maximum (e.g., time to execute a task) acceptable value considered achievable within the available cost, schedule and technology at low to moderate risk. Performance that does not achieve the threshold value is not operationally effective or suitable. The objective value for an attribute is the desired operational goal achievable but at higher risk in cost, schedule and technology. Performance beyond the objective value does not justify additional expense. The difference between threshold and objective values sets the trade space for meeting the thresholds of multiple attributes. Advances in technology or changes in joint concepts may result in changes to threshold and objective values in future increments.

The attributes and their supporting rationale should reflect analytical insights identified by the supporting capabilities-based assessment (CBA) used to develop an ICD. The attributes should be directly related to the measures of effectiveness attributed to the capability as defined in the ICD. As a minimum, supporting analyses must include:

- AoA for potential acquisition category (ACAT) I programs and other programs as directed by the milestone decision authority (MDA);
- Cost-schedule-performance tradeoffs analysis;
- Capability cost tradeoffs analysis;
- Results of experimentation, testing and evaluation;
- Life -cycle supportability and affordability analysis;
- Lessons learned during the system development and demonstration (SDD) phase; and
- User feedback on fielded production increments

There is no DoD official definition of KSA, but an attribute is a quality or characteristic inherent in or ascribed to someone or something. Some attributes of software intensive systems may include interoperability, portability, reusability, robustness, generality, scalability, and flexibility. System attributes could be those attributes that either directly contribute to the achievement of a particular KPP or are part of the allocated roll-up into a capability that achieves a KPP. The following quotation provides a good summary description of performance attributes and KPP:

*The CDD and CPD state the operational and support-related performance attributes of a system that provide the desired capability*

*required by the warfighter, attributes so significant that they must be verified by testing and evaluation. The documents shall identify the specific attributes contributing most significantly to the desired operational capability, in threshold-objective format. Whenever possible, attributes should be stated in terms reflecting the capabilities necessary to operate in the full range of military operations and environment intended for the system. This will be used to guide the acquisition community in making tradeoff decisions between the threshold and objective values of the stated attributes. Operational testing will assess the operational effectiveness and suitability of the system and its ability to meet the production threshold values. (CJCSM 3170.01C, A-10 – A-11)*

***System/SoS Measures of Effectiveness and Suitability***

It is important for the analyst to understand the distinction between Mission MOE/Task MOP and system/SoS-level MOE/MOP. Throughout this document, Mission MOE and Task MOP indicate the measures at the joint mission and joint task level respectively and system/SoS MOE, MOP, and measures of suitability (MOS) refer to measures at the system/SoS level. For a system or SoS, there can be MOE, MOP, MOP, and MOS that address particular performance measures *at the system level*. The discussion in this section focuses on system/SoS MOE, MOP, and MOS. The purpose of the following description is to put system/SoS attributes in the context of the overall joint test. Methods for developing these measures will not be emphasized because their development is already understood by Service test agencies. The methods and processes associated with a joint test and evaluation do not replace those concepts.[2]

A system/SOS MOE is a criterion used to "[m]easure of the overall ability to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, supportability, survivability, vulnerability, and threat." (CJCSM 3170.01C, GL-19) The system/SoS MOE should be tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect due to a particular system or SoS. On the other hand, a measure of suitability (MOS) measures

*[t]he degree to which a system can be placed and sustained satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, environmental, safety and occupational health, human factors, habitability, manpower, logistics,*

[2] The reader may also consult the following service-level evaluator and analyst guides for more information about system-level measures: AEC (2008), Chapters 6, 7 and 10; AFOTEC (2007), pp 4-13 through 4-25; OPTEVFOR (2007), Chapters 2 through 7.

*supportability, logistics supportability, natural environment effects and impacts, documentation, and training requirements. (ibid)*

Table B-2 shows example operational effectiveness and suitability performance measures. The critical operational effectiveness and suitability parameters and constraints must crosswalk to those used in the AoA, and include manpower, personnel, training, software, computer resources, transportation (lift), compatibility, interoperability and integration, information assurance (IA), electromagnetic environmental effects and spectrum supportability, etc. They are focused on operational capabilities, not design specifications such as weight, size, etc. The list should be limited to critical measures that apply to capabilities essential to mission accomplishment and include all KPP. For each listed parameter, the analyst should provide the threshold and the objective values from the requirement document and reference paragraph. Keep in mind that if every critical operational issue is resolved favorably, the system should be operationally effective and operationally suitable when employed in its intended environment by typical users. However, successful achievement of these system-level measures does not guarantee success as the Task MOP and Mission MOE level.

**Table B-2. Example Measures of Effectiveness and Suitability**

| Operational Capability | Parameter | Capability Threshold | Capability Objective | Capability Reference |
|---|---|---|---|---|
| **Mobility** | Land Speed Miles per hour on secondary roads | xx miles per hour | xx miles per hour | Paragraph xxx |
| **Firepower** | Accuracy Main Gun Probability of hit/stationary platform/ stationary target | xxx probability of hit @ xxx range | xxx probability of hit @ xxx range | Paragraph xxx |
| **Supportability** | Reliability Mean Time Between Operational Failure | xxx hours | xxx hours | Paragraph xxx |

***System Critical Technical Parameters (CTP)***

Some COIs have supporting CTPs and thresholds. CTPs are measurable critical system characteristics that, when achieved, allow the attainment of desired operational performance capabilities. However, individual attainment of these attributes does not guarantee that the COI will be favorably resolved. The judgment of the operational test agency (OTA) is used to determine if the COI is favorably resolved.

Table B-3 shows example CTPs. CTPs are not user requirements; they are technical measures derived from desired user capabilities. Failure to achieve a CTP should be considered a reliable indicator that the system is behind in the

planned development schedule or will likely not achieve an operational requirement. The analyst should limit the list of CTP to those that support COI. The system specification is usually a good reference for the identification of CTP.

Next to each technical parameter, the analyst should list a threshold for each stage of development. For most technical parameters, the listed thresholds should reflect growth as the system progresses toward achieving the desired capabilities. Also, list the decision supported after each event to highlight technical performance required before entering the next acquisition or operational test phase. Ensure technical parameters are included for technical interoperability.

**Table B-3. Example Critical Technical Parameter**

| Supported Operational Capability (include ICD/CDD/CPD reference) | Technical Parameter | Developmental Stage Event | Threshold Value | Decision Supported |
|---|---|---|---|---|
| Main Gun Probability of Hit, 94 % at 1,500 meters (CDD. para. xxx.x) | Auxiliary sight Bore sight accuracy | System Demo Test-Accuracy Test | +/- 5 mils | Milestone B |
| | | Prod Readiness Test-Accuracy | +/- 3 mils | MS C (Low-Rate Initial Production Decision) |
| | | Prod Qual Test | +/- 1 mil | FRP |

### B.3.3.1.2.3 Determine Feasible Levels/Regions for Measures

In this step, the analyst attempts to determine a feasible level or region for each measure. If a published standard or threshold value is provided, then the feasible level or region is that standard or threshold value. If none is provided, the analyst may be required to go back to the capability manager or operational commander to determine the value(s).

A feasible level has one value. For quantitative measures, the value must be numeric and specify a level of accuracy. For qualitative measures, the value must be stated and described in a manner that is measurable and well defined.

A feasible region will have two levels for each measure. For a continuous, quantitative measure, an upper and lower limit must be identified. The same may be true for discrete values as long as the intervals between the discrete values are consistent and can be defined. For certain discrete values and most qualitative values, the analyst may need to identify and define each possible level. By doing so, the range is also determined.

#### B.3.3.1.2.4 Structure the Measures

In addition to identifying measures of each type, the analyst must develop a measure framework that captures the relationship between the measures at different levels. This framework provides the basis for determining JMe based upon particular combinations of factors, and is thus a critical step. The analyst should attempt to define the framework initially; however, subsequent analyses may expose previously unidentified relationships or indicate that previously identified relationships do not exist.

Due to the highly complex nature of the joint mission environment, the relationships between identified measures will likely be quite complicated (many-to-many) and very difficult to determine. Even with an accurate mapping of the measures to each other, success of lower-level measures may not necessarily correspond to success of the higher-level measures to which they relate. Lower-level measures may be viewed as success in isolation; however, higher-level measures may require more than just that all lower-level measures are successful. The success of the higher-level measure may require that the lower levels combine in a way that creates higher-level success through interactions, or the analysis team may not have captured one or more key lower-level measures that have a significant impact on the higher-level measure. Nonetheless, it is still important to capture the relationships so that the resulting output can be put into context within the overall joint mission environment (JME).

The complexity of the measure framework leads to at least two particular challenges. The first challenge is defining the relationship between measure responses for the purpose of screening potential factor combinations. In the subsequent exploratory analyses, the analysis team will use experimental design and the measure framework to narrow down the original set of factors and ranges to achieve a smaller set of factors from which to choose test trials. With only one measure response, such a task would be fairly straightforward; however, with multiple responses, and measures at multiple levels, the analysis of the outputs becomes quite complicated.

The analysis team must decide how they will evaluate the impact of factors based upon the multiple measure responses. There are a few techniques that can be used to accommodate multiple responses. First, the team can create a single measure as a weighted combination of multiple measure responses. The weights in such a technique are critical, as they will have a tremendous impact on the results. Weights should represent the importance to the ultimate decision maker; thus, the team must query stakeholders to ensure that they accurately capture the importance of the measures. Using weights usually requires a transformation of the measure responses into a common utility or value scale in order to ensure an "apples to apples" comparison. Another technique is to select the most important measure and consider only that in the analysis. Stakeholder analysis is essential to ensuring the correct measure has been identified and chosen. The team can

evaluate the responses for each measure and develop a rule set to determine whether the factor or factor interaction is of interest (e.g., if the factor combination has a significant impact on three of the four measures, that factor combination constitutes a factor combination of interest). The latter technique may also be used to prioritize the factor combinations. For instance, the highest-priority factors would be those that affect all of the measures, the next highest priority factor combinations would be those that affect all but one of the measures, etc.

The second challenge is determining what constitutes a factor combination "of interest" for screening purposes. The analysis team may screen for those factor combinations that seem to increase or decrease JMe, or the team may be interested in those combinations that lead to no change in JMe at all. The latter may be the case when the addition or removal of a particular system or capability is part of the factor space. The team may want to know that adding a particular system had no impact on the JMe. However, in most cases, the team should be focused on those factor combinations that have an impact. In most cases, the team should be interested in factor combinations that have a significant effect on the JMe, whether it is positive or negative. In the end, what the team is interested in will depend upon the questions being asked for the test and will likely be influenced by key stakeholders.

#### B.3.3.2 Explore Factor Selection – CTM 1.2.3.2

The purpose of ***CTM 1.2.3.2, Explore Factor Selection***, is to evaluate the measure responses from the exploratory runs to refine the JMe factor test space and to identify factor combinations to explore during subsequent iterations, as required. This process is discussed in detail in annex C.

#### B.3.3.3 Synthesize Exploratory Analysis – CTM 1.2.3.3

The purpose of ***CTM 1.2.3.3, Synthesize Exploratory Results***, is to integrate all of the analyses conducted during the multiple exploratory iterations to draw overarching insights about the refined JMe factor test space and the measure framework. This process is discussed further in Annex C.

### B.3.4 Develop Risks and Mitigation – CTM 1.2.4

#### B.3.4.1 Purpose

The purpose of ***CTM 1.2.4, Develop Risks and Mitigations***, is to identify and characterize risks, particularly those associated with the assumptions and modeling capabilities used during the exploratory analysis, and to develop a risk mitigation plan.

#### B.3.4.2 Overview

Once the TES framework has been developed, the analyst will complete an initial assessment of risks and possible mitigations in ***CTM 1.2.4, Develop Risks and Mitigations***. In particular, the focus of the discussions in this handbook is on those risks associated with the framework development and modeling limitations and assumptions

identified during the previous steps. The responsibility for risk and mitigation development goes beyond the analysis team, whose role is the focus of this chapter. Figure B-10 shows the steps involved in this process.

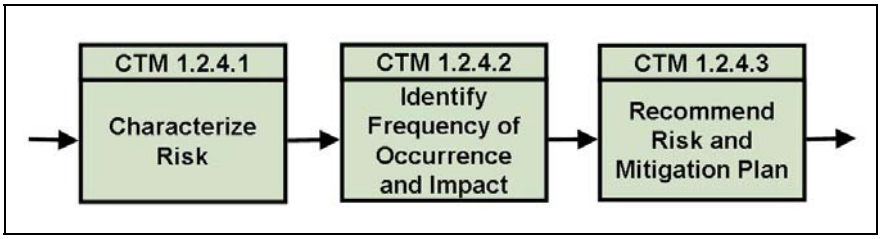



**Figure B-10. CTM 1.2.4, Develop Risks and Mitigations**

In ***CTM 1.2.4.1, Characterize Risk***, the analysis team identifies potential risk areas based upon the exploratory analysis and consolidates them for addition to the risk management plan.

***CTM 1.2.4.2, Refine Initial Test and Evaluation Strategy Outline***, involves the identification of the expected frequency (or probability) of occurrence, and the impact (or consequence) of the risk event.

In ***CTM 1.2.4.3, Recommend Risk Management Plan***, the analysis team develops methods to mitigate risk and helps develop the risk management plan.

### B.3.4.3 Inputs

#### B.3.4.3.1 Exploratory Analysis Findings and Insights

These are the findings and insights from all of the exploratory analysis iterations, grouped into the following categories: factor main effects, factor interactions, factor range, measures, and measure relationships. In this section, these will be used to help identify potential risks.

#### B.3.4.3.2 Refined Test and Evaluation Strategy

The Test and Evaluation Strategy (TES) captures the efforts conducted within ***CTM 1.3, Develop Evaluation Strategy***, by documenting some of the key outputs of the process. Risk management is part of *CTM 1.3* and should be added to this T&E strategy document.

### B.3.4.4 Checklist

- ☐ Characterize risk
- ☐ Identify risks from exploratory analysis modeling assumptions
- ☐ Identify risks from exploratory analysis model capabilities and limitations
- ☐ Identify risks from exploratory analysis modeling uncertainty
- ☐ Identify risks from exploratory analysis measure framework
- ☐ Identify frequency of occurrence and impact

- Identify frequency of occurrence of the risks
- Identify the impact of the risks
- Recommend risk management plan

### B.3.4.5 Process

#### B.3.4.5.1 Risk Characterization

Overall, risk mitigation is the process that identifies, evaluates, selects, and implements options in order to set risk at acceptable levels given system/SoS constraints and objectives. Often, the program will have a risk management plan (RMP). If so, the analysis team should refer to it and use the documented risk-management strategy to identify, evaluate, and track risk areas. Risk can be associated with all aspects of a program. Overall risk components include a future root-cause, which, if eliminated or corrected, would prevent a potential consequence from occurring; a probability (likelihood) assessed at the present time of that future root-cause occurring; and the consequence of that future occurrence. The focus of this section is identifying potential future root-causes. The next section will address the other two components.

In this handbook, the focus of the discussion is on the evaluation thread of the CTM; therefore, this chapter will address risk only from the evaluation thread perspective. The risk mitigation planning discussed here is only one component of the overall program and test risk strategies.[3]

The following are some potential contributors to risk that should be identified:

- **Modeling Assumptions:** Because a model is an abstraction of reality, assumptions must be made to represent scenarios and factors of interest. Some of these assumptions may be sources of risk and should be noted as such. For instance, the test trial sets recommended by the team are based upon insights drawn from the modeling results. Those model results, in turn, may have been significantly influenced by some of the major assumptions made during the modeling process. Those should be noted.
- **Model Capabilities:** For any given capability, each potential model may represent that capability differently, even for higher resolution models where the required abstractions are less. The model capabilities themselves will affect the results, insights, and recommended test trial sets.
- **Modeling Uncertainty:** Stochastic modeling involves uncertain inputs. The team must use statistics and other techniques to draw inferences about the distributions of outcomes. Each technique chosen has strengths, weaknesses, and assumptions associated with it. Some of those may present varying levels of risk. For instance, simply assuming independence, stationary, or normality for the output data may not be warranted, leading to possible inference errors; in such

[3] The reader may also consult AEC (2008), Chapter 13, for techniques to assess risk.

cases the analyst may choose to use something other than classical statistical methods, such as time-series analysis or nonparametric methods.

- **Measures:** The measures included in the measure framework are ultimately what determine JMe. Additionally, how the team defined relationships between the measures in the measure structure to evaluate JMe is significant. If the team made assumptions about the importance or unimportance of particular measures, those assumptions may contribute to risk.

#### B.3.4.5.2 Frequency and Impact Assessment

In this step, the analysis team attempts to assess, for each risk identified, the expected frequency (or probability or likelihood) that the future root cause will occur and the impact of the consequences of such an occurrence.

In some cases, determining frequency of occurrence may be straightforward based upon the analyses conducted. Normally, the analysis team has done some sensitivity analysis as part of the overall exploration processes to gain a deeper understanding of the impact of some of the assumptions made. Additionally, the use of appropriate statistical techniques to explore the data often provides a direct measure of the probability that certain assumptions (or hypotheses) are correct. In other cases, assessment of frequency is a judgment call based upon team expertise.

The determination of the impact of potential consequences may also be directly assessed as part of the exploratory analyses. For instance, if weather is determined to have a significant impact on the JMe during exploration via M&S, but cannot be controlled in the actual test, the team can provide a quantitative assessment of impact based upon the DOE results. If high winds severely degrade the JMe of the capability, the test planning team will have an understanding of what the impact will be. As a result, they may plan for a location or time of year when the winds are expected to be calm. The team may have to make a subjective judgment call to assess the impact of certain risks.

#### B.3.4.5.3 Risk Mitigation

During this step, the analysis team should assist in the development of the RMP by making recommendations concerning risk mitigation. Risk mitigation is deciding what to do about each of the risks assessed as important to the system/SoS and documenting the planned response. This is an iterative process, and the risk analysis should be reviewed and updated throughout the CTM. Some risks may be considered so potentially damaging that the PM chooses to avoid them completely, or may seek to transfer the risk to another party more capable of accommodating the risk. Some risks may be accepted with no further actions (low risks), but other risks may be accepted simply because there is no credible alternative. This second class of accepted risks will require that contingency actions be developed in case they occur. Many risk-handling choices seek to mitigate the probability of the risk event or to lower the consequence to an acceptable level.

#### B.3.4.6 Tools and Techniques

- Risk assessment techniques
- Risk management techniques

#### B.3.4.7 Output - Risk Management Plan (RMP)

The RMP documents the identified risks, expected frequency of occurrence, consequences or impact of occurrence, and the risk-mitigation techniques to be taken. This is not the primary responsibility of the analysis team, as the RMP encompasses the entire program. However, the analysis team will contribute to the RMP based upon the risks inherent to the evaluation strategy.

## B.4 Tools and Techniques

- Stakeholder analysis
- Background research
- DoDAF documentation development
- Capability crosswalk and other tracking matrices
- Measure development
- Decision-analysis techniques and multi-attribute-decision-making techniques (MADM)

## B.5. Products

### B.5.1 Initial Factor Set

The initial factor set contains the factors identified in each of the three dimensions, with each factor including a description; a characterization of the factor as either quantitative or qualitative and either controllable or uncontrollable; the range through which the factor may vary for the purposes of experimentation; its source; and any other pertinent information deemed relevant by the analysis team. This factor set may be revised throughout the CTM as new information becomes available via stakeholders, or subsequent modeling and experimentation.

### B.5.2 Measure Framework

The initial measure framework contains the measures chosen in each of the three measure levels (Mission MOE, Task MOP, and System/SoS Attributes), with each measure including a complete and unambiguous description; a characterization as either quantitative or qualitative; the method of calculation or determination; the unit of measure; the data elements required to calculate or determine the measure response value or level; definitions of key terms; and the desired fidelity of measurement. Additionally, the framework should include the relationship between measures at the same level and measures at different levels. As with the initial factor set, the measure framework may be revised throughout the CTM as new information becomes available via stakeholders, or subsequent modeling and experimentation.

### B.5.3 Capabilities Crosswalk

The capabilities crosswalk depicts the key elements of the evaluation strategy. The capability crosswalk matrix is a mapping of CCI to mission desired effects, tasks, and system/SoS configurations to set the stage for developing capability evaluation strategies. The capabilities crosswalk is used to document both the factors and the measure framework. See Annex E for more details about the capabilities crosswalk.

# ANNEX C - EXPLORATORY ANALYSIS

## C.1 Purpose

This annex describes the methods and processes for Capability Test Methodology (***CTM) 1.2.3.2, Explore Factor Selection***, which is used to conduct an initial exploratory analysis of independent variables (test factors) impact on dependent variables (test measures). The results of the exploratory analysis are intended to refine the joint mission effectiveness (JMe) factor test space and to identify significant test factor combinations to use during live, virtual, constructive distributed environment (LVC-DE) test and evaluation (T&E).

## C.2 Overview

The analysis team needs a methodology to explore a greater range of possible outcomes before committing to an approach that produces only a narrow scope of results. An exploratory analysis approach is an analytical tool that can provide the analysis team with potential insights across a broader range of potential outcomes than more narrowly-focused analyses. These insights include, but are not limited to, the exploration of unintended consequences, identifying trade-offs in variables and constraints, enhancing the intuition about a scenario and ultimately providing a good and robust solution in the midst of great uncertainty in combat models (Horne, 1998). The exploratory analysis approach is an attempt to help the analysis team think through complicated issues by illuminating the consequences of various assumptions, reinforcing or challenging intuition, and illustrating alternatives that might otherwise not have been considered. It will be utilized prior to the decision to focus resources on the LVC-DE and to guide the overall effort.

Exploratory analysis is the fundamental theme of the metamodel-test-metamodel approach. Its purpose is three-fold (represented in figure C-1):

1) To explore a wide range of possible factors and levels that might affect joint mission effectiveness (JMe), referred to as the initial JMe factor test space;
2) To identify those combinations of factors that have the greatest impact on JMe, referred to as the refined JMe factor test space; and
3) To recommend potential factor combinations of interest from the refined JMe factor test space for subsequent test events, referred to as potential test trial sets.

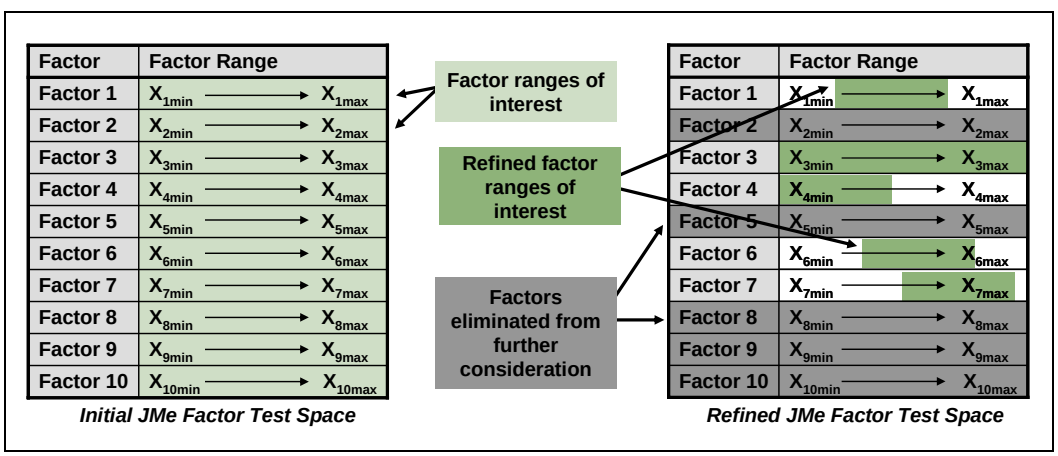



**Figure C-1. Exploratory Analysis Representation**

JMe test space refinement is accomplished by examining the test factors and factor interrelationships within the context of the measure framework, in order to identify factors of importance, factor levels of impact, and key interrelationships.

In general, the exploratory analysis takes a broad view across all of the potential factors and levels to identify which factors, and factor combinations, have the greatest impact on the effectiveness of the joint mission. It is a learning process during which the analysis team gains insights about the characterization of the factor space. This exploratory approach is intended to provide information such that the analyst can make decisions to formulate the experiment more clearly. Rather than predicting, optimizing, or tuning, the focus is on identifying significant factors and interactions and finding regions, ranges and thresholds where interesting things happen (Lucas, 2002).

These refined factors and interrelationships will provide the basis for the development of the test. Such a process allows for the simulation of the test in order to provide statistical output that will validate the significance of test factors and interrelationships. The resulting analysis will indicate where models and simulation need to be changed, as well as indicate other factors or interrelationships that may need to be modeled.

Depending on input from combatant commands (COCOM) on critical concerns, the outcome of the modeling and simulation (M&S) runs and/or re-runs, subject matter expert (SME) review, the JMe test space will be finalized. This final JMe test space will provide the basis for the development of the specific test vignettes and test trial sets, as well as LVC-DE infrastructure design and implementation for actual LVC testing.

This test space refinement through exploratory analysis occurs within ***CTM 1.2.3, Establish Test and Evaluation Strategy Framework***. Figure C-2 shows the sub-processes for developing the test and evaluation strategy (TES).

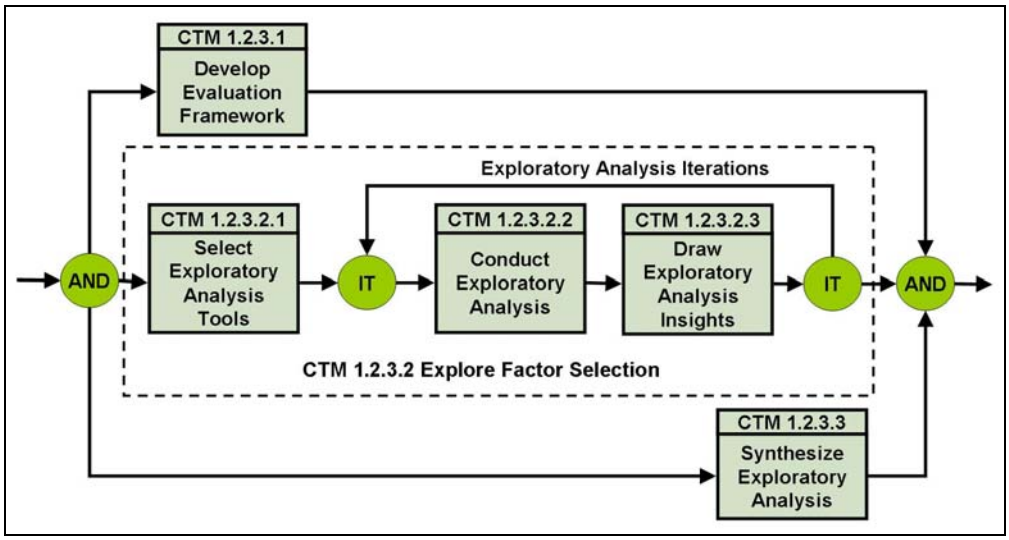



**Figure C-2. CTM 1.2.3, Establish Test and Evaluation Strategy Framework**

In ***CTM 1.2.3.1, Develop Evaluation Framework***, the analyst identifies and refines the test factors and levels that are expected to affect SoS performance, and the measures that will be used to evaluate and assess SoS performance. This process is discussed in Chapter 2 in the *Analyst's Handbook for Testing in a Joint Environment* (Analyst's Handbook). Additional detailed information is also available in Annex B of the Analyst's Handbook.

***CTM 1.2.3.2, Explore Factor Selection***, is an iterative loop using the metamodel-test-metamodel approach to evaluate test factors impact on test measures. This is the main focus for this annex.

In ***CTM 1.2.3.2.1, Select Exploratory Analysis Tools***, the analysis team identifies requirements for exploratory analysis tools; identifies potential tools to meet those requirements; assesses the tools; and selects the desired toolset. This process is discussed in section C.3.

***CTM 1.2.3.2.2, Conduct Exploratory Analysis***, contains the bulk of the analysis of the data. In this step, the analyst reduces and organizes the exploratory analysis data, evaluates the individual measure responses, and evaluates the measure relationships. This process is discussed in section C.4.

In ***CTM 1.2.3.2.3, Draw Exploratory Analysis Insights***, the analysis team consolidates the findings from the previous step to re-characterize the JMe, including both the factor test space and the measure framework. This process is discussed in Section C.5.

In ***CTM 1.2.3.3, Synthesize Exploratory Results***, the exploratory results is analyzed to draw overarching insights about the refined JMe factor test space and the measure framework. This process is discussed in section C.6.

**Note**: The level of effort in conducting an exploratory analysis of the test space is dependent on the resources and time available. A simple exploratory analysis may consist of a meeting of SMEs and key stakeholders to agree on priorities. A complete detailed exploratory analysis process may include multiple iterations of modeling and simulation, analysis, and consolidation of results that can quantitatively justify factor reduction and test space refinement.

An example of a repetitive model-test-model exploratory analysis approach is included as Enclosure (1) to this annex.

### C.2.1 Inputs

**Data:** The primary inputs to this step are the data developed from the M&S and other data collection events. This data should be consolidated and accessible by the team for the subsequent steps of the exploratory analysis. This data will be analyzed as part of this process.

**Data Management Plan (DMP):** The DMP identifies the procedures that will be followed for accessing and handling the data; dealing with sensitive information; modifying the data files; and storing the data. Procedures outlined in this plan provide the basis for how the data will be managed throughout the conduct of the analysis.

**Exploration Strategy Outline:** The exploration strategy outline contains a summary of the tools and techniques to be used, including the mapping of the tools/techniques to the evaluation framework; the chosen experimental designs and other analytical methods; the computing resources required; the scenarios and vignettes, including a mapping of the vignettes to the analytical methods; and the overall exploratory analysis approach to be used. It will be used here to understand the overall context within which the data, results, findings, and insights will be considered.

**Refined Evaluation Framework:** The refined evaluation framework captures any refinements to the original evaluation framework based upon assumptions, tool and technique capabilities, and insights gained during the exploratory analysis process thus far. This evaluation framework will be revised during this process as new insights about JMe are developed.

### C.2.2 Action checklist

- ❑ Prepare exploratory analysis tools.
- ❑ Identify exploratory analysis tool requirements.
- ❑ Identify statistical software requirements.
- ❑ Identify database software requirements.
- ❑ Identify other specialized software requirements.
- ❑ Identify potential exploratory tools and techniques.
- ❑ Assess exploratory tool and technique capabilities.
- ❑ Select exploratory analysis tools and techniques.
- ❑ Conduct exploratory analysis.
- ❑ Reduce and organize exploratory data.
- ❑ Evaluate individual measure exploratory responses.
- ❑ Evaluate exploratory measure relationships.
- ❑ Draw exploratory analysis insights.
- ❑ Consolidate exploratory findings.
- ❑ Re-characterize JMe of interest.

## C.3 Instructions for Analysis Tool Selection – CTM 1.2.3.2.1

The analysis team must identify, assess, and select the analysis tools to be used, based upon their ability to facilitate the required analyses. The steps involved in this process are shown in Figure C-3. For purposes of this discussion, analysis tools include statistical software, database software, and other related tools required to reduce and analyze the data.

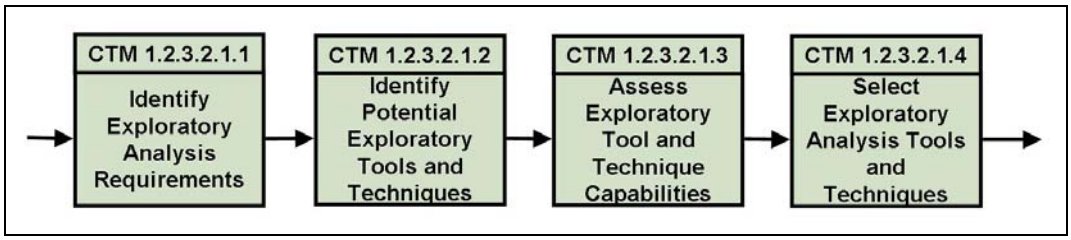



**Figure C-3. CTM 1.2.3.2.1, Select Exploratory Analysis Tools**

***CTM 1.2.3.2.1.1, Identify Exploratory Analysis Tool Requirements***, involves the identification of analytical tool requirements, which will be derived, in part, from an identification of the analytical techniques the team plans to use during the analysis. These steps are discussed in Section C.3.1, Analysis Tool Requirements.

***CTM 1.2.3.2.1.2, Identify Potential Exploratory Tools and Techniques***, involves the identification of potential analysis tools, including statistical software, database software, and other tools as appropriate. These steps are discussed in Section C.3.2, Analysis Tool Identification.

In ***CTM 1.2.3.2.1.3, Assess Exploratory Tool and Technique Capabilities***, the analysis team examines each potential tool to determine its general characteristics and its capability to provide the required analytical outputs. Details of this process can be found in Section C.3.3, Analysis Tool Assessment.

Finally, in ***CTM 1.2.3.2.1.4, Select Exploratory Analysis Tools and Techniques***, the analysis team compares their requirements against the potential tools and selects the set that will facilitate the most robust analysis of the JMe factor test space. This process is discussed in Section C.3.4, Analysis Tool Selection.

### C.3.1 Analysis Tool Requirements - CTM 1.2.3.2.1.1

**(1) <u>Purpose</u>**. The analysis team should identify the exploratory analysis tool requirements that will serve as the primary criteria used to identify potential tools, and against which the identified tools will be assessed. In addition to analytical capabilities, the analysis team may consider such requirements as tool availability and user expertise; however, the key here is to give the analysts a process for selecting appropriate tools instead of defaulting to what they have on hand or normally use. Merely defaulting to tools that the team is comfortable with may eliminate promising analytical techniques and may result in limited insights. In general, a particular suite of tools will not fit all cases.

**(2) <u>Discussion</u>.** The development of requirements should begin with an identification of the analytical requirements. The analyst should based the requirements on the set of test factors; the number of test factors, the types of factors (quantitative or qualitative), the level of values for each test factor, the proposed design of experiment (DOE), and any other issues which may impact requirements. Test factors should be identified as either SoS decisional factors (those materiel and non-materiel factors that impact an acquisition decision) or conditional factors.

The analyst should have an estimate of the resources that will be available for LVC-DE testing. Limited resources will restrict the number of test factors that can be considered, and thus may require a larger number of test factors to be a lower priority and set at constant values. Time can also impact the number of test factors in an LVC-DE. Shorter time periods for testing will decrease the number of test runs and limit test factors. The analyst may also need to understand acceptable levels of risk that may be considered in the identification of analysis requirements. If a factor is selected to be eliminated from the test design, the risk of removing that test factor must be considered, given its correlation to the dependent variables.

**(3) Tools and Techniques.** If the previous development of the evaluation framework provides the team with an incomplete understanding of the actual problem space, then the team should consult the joint operational context of test (JOC-T) efforts to understand the overall context of the test. The analyst may also wish to consult with SMEs and/or the test customers to determine priorities and levels of risk.

**(4) Inputs.**

- **List of Test Factors**. The test factors are those independent variables considered for the capability under test. The list should be labeled and sub-divided as SoS or conditional test factors. See Annex B for more information on selecting test factors.

**(5) Actions.**

- **Determine exploratory analysis requirements.** The analyst will determine the goal and objectives of the exploratory analysis. Typically the goal is to reduce the test space to a reasonable number of test factors that can be considered in an LVC-DE. This should be quantified when possible in terms of expected output (what is a reasonable set of test factors)

- **Determine decision criteria for analysis.** The analyst will determine what the decision criteria will be that will justify the type of analysis tools needed in the exploratory analysis process.

- **Prioritize the test factors**. The analyst will determine a priority list of SoS decision factors and a separate list of conditional factors. Prioritization may be based on decisional analysis techniques or may be based on SME and/or customer judgment.

**(6) Products.**

- **List of decision criteria.** The decision criteria will be developed into a list that can be used to score potential analysis tools to be used in the exploratory analysis process. Weights should be given to each criterion if one is more important than another.

- **Prioritized list of test factors**. Two lists will be developed: one list prioritizes the SoS factors and a second separate list prioritizes conditional factors. Those lower priority factors may be set at a fixed level and not be considered in the final LVC-DE.

### C.3.2 Analysis Tool Identification - CTM 1.2.3.2.1.2

**(1) Purpose**. This step involves the identification of particular analysis tools in accordance with the requirements developed in the previous step. The analysis team must attempt to examine the greatest possible range of potential tools. Although a more detailed evaluation of each resource will be conducted later, the team should conduct a high-level assessment of all the potential tools available. The analyst should screen out tools that clearly are not viable options for the current analyses.

**(2) Discussion**. The analysis team must attempt to examine the greatest possible range of potential assets. Although a more detailed evaluation of each resource will be conducted at a later time, the team should conduct a high-level assessment of all the potential resources available.

When identifying potential tools and techniques, the analysis team should use their understanding of the evaluation framework (factors and measures) to search for models and simulations capable of representing the capabilities, SoS, and environment to be tested. It is not necessary to go into a great level of detail here; the goal is to identify tools and techniques that might be potential options. It is also important to note that the team should consider tools and techniques that might only represent a particular aspect of the problem space. Since more than one tool will likely be used, a single tool need not have the capability of representing everything of interest. The team should also try to identify lower-resolution or less complex models to explore a greater space of possible outcomes. Larger or more complex models will require greater time for setup and execution. Alternatively, the analysis team may decide to examine the JMe space with lower-resolution models to identify those regions of interest that should be further explored with higher-resolution models.

Existing software packages allow the analyst to identify the factors and map them to specific parameters in the model; select appropriate DOE; specify the levels through which the factors will vary; specify the number of runs; and distribute runs within a computing cluster or across multiple processors. Such tools, however, are designed to interoperate with particular, pre-identified models. They can often be modified to

accommodate new models, but this must be identified early (in this step) to plan for the modification of the software.

If the previous development of the evaluation framework provides the team with only an incomplete understanding of the actual problem space, then the team should consult the JOC-T efforts to understand the overall context of the test.

There are two primary classifications of models and simulations: physical and interactive simulations. In the broad sense, physical simulations refer to a simulation in which physical objects are substituted for the real thing. These physical objects are often chosen because they are smaller or cheaper than the actual object or system. There are various degrees of abstraction from reality that correlate to the level of resolution of the physical simulation. The interactive simulation is a special kind of physical simulation, often referred to as human-in-the-loop (HITL) simulation, in which physical simulations include human operators. Fundamentally all simulations will generally fall into one of these two categories.

There are numerous examples of simulations in different areas of interest. These include but are not limited to:

- City simulators/urban simulations
- Engineering, technology or process simulations
- Military simulations
- Robotics simulators

For purposes of this guide, the analyst is primarily focused on the simulations classified as military. This subset of military simulations is quite extensive. The number of force on force models, transportation and logistics models, and political-social behavior models covers a wide spectrum of resolutions. These will be discussed further. The analyst must select the appropriate simulation tool based on the study needs. The analyst must evaluate the needs of the study with the modeling capabilities of the simulation and the scalability of the simulation.

The analyst must make choices on the resolution needed to meet the study the needs. It is important to recognize that once the simulation is chosen the scenario or model must be built and this is very much an art. The art is in deciding:

- What elements and activities to include in a model
- Which dimensions to make variables, constants, stochastic, constraints
- What to assume when creating relationships among the variables (linear vs. nonlinear, deterministic vs. stochastic, etc.)
- How in a step-by-step process to remove the unrealistic assumptions we necessarily make in building the initial version (Bonder, 2002).

In addition to models and simulations, the team should identify other tools and techniques that may be used to examine the factor test space. In particular, they should identify

techniques for gathering stakeholder input. Techniques of this sort include questionnaires, interviews, and professional military judgment (PMJ) panels. They should also consider software packages that will allow the analyst to identify the factors and map them to specific parameters in the chosen models; select an appropriate DOE; specify the levels through which the factors will vary; specify the number of runs; and distribute runs within a computing cluster or across multiple processors. Such tools, however, are designed to interoperate with particular, pre-identified models. They can often be modified to accommodate new models, but this must be identified early in this and the subsequent tests in order to plan for the modification of the software.

**(3) Tools and Techniques**. There are many analysis tools and, once the study analysis requirements are understood, a simple matrix can be developed comparing the different analysis tools with the key criteria. The key criteria may differ slightly from study to study. However, a standard set of criteria will almost always apply. Key criteria include:

- Cost
- Open source / licenses
- User interface
- Analyst familiarity
- Comparison of statistical capabilities such as the ability to do:
- Descriptive statistics
- ANOVA
- Regression
- Nonparametric statistics
- Quality control
- Time series analysis
- Other required analyses
- Computing resource requirements
- Size of datasets it can handle
- Import/export capabilities to other formats and computing platforms
- Specific tools that is present without modification/customization (e.g., logistic regression, time-series analysis)
- Whether additional modules can be added on later if requirements change

Ideally military simulations should be as realistic as possible. The ideal simulations are designed in such a way as to provide measurable, repeatable results that can be confirmed by observation of real-world events. As a general principle, the most reliable data are produced by actual observation and the most reliable theories are based on it. This is also true in military analysis, where analysts look towards live field exercises and trials as providing data that are likely to be realistic and verifiable. The analysis team will quickly determine the cost of obtaining very specific, detailed, and accurate input data are often too great to make it a feasible effort.

Simulations are an abstraction of reality. Military analysts have always struggled with the dilemma of balancing appropriate level of abstraction with the available resources to meet the needs of the study (see Figure C-4). Various tools and techniques exist based on the level of abstraction chosen to address the issue at hand. There are various categories of tools that should be examined as part of the selection process. These categories or types of simulation range from manual (implying very little computer involvement) to computer-assisted (implying HITL or virtual simulation, or fully computerized) (Taylor, 1983). Any user of simulations must always remember they are only an approximation of reality, and hence only as accurate as the model itself.

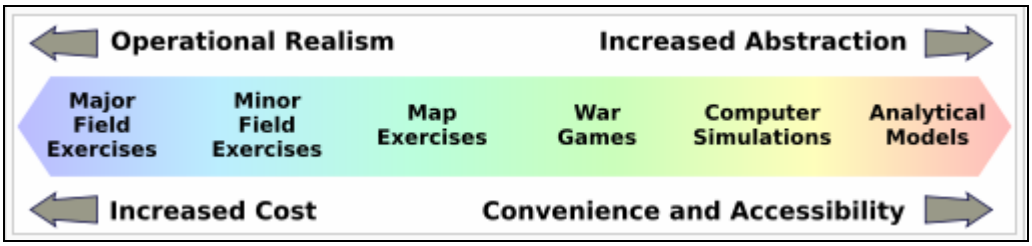



**Figure C-4. Range of Military Simulations (Taylor, 1983)**

The following is a list of some potential categories of military models and simulations. This is not an exhaustive list, it is a general overview. Some examples are included; for a more complete listing, the user should consult the DoD and service Modeling and Simulation Resource Repositories (MSRR) via the following link: http://www.msrr.dmso.mil/.

1) **Agent Based Models (ABM)** – Simulations based on the idea that it is possible to represent, in computerized form, the behavior of entities which are active in the world, and that it is thus possible to represent a phenomenon as the result of interactions of an assembly of agents with their own operation autonomy (Ferber, 1999). Examples of existing ABM include Map Aware Non-uniform Automata (MANA) and Pythagoras.
2) **Constructive Combat Simulations** – Simulations vary widely in the size of the represented force structure, generally are loosely categorized as force-on-force attrition models. These models do not require user interaction during the model run, since they involve simulated units and personnel operating simulated systems. Examples of existing constructive simulations include the Army's Combined Arms Analysis Tool for the XXIst Century (COMBAT$^{XXI}$) and the Navy's Naval Simulation System (NSS).
3) **Human-In-The-Loop (HITL)** – Simulations inject human interactions into key points of a simulation to exercise motor skills, decision making skills or communication skills. Many constructive simulations can also be operated as HITL as well. Examples include the Army's Janus and One Semi-Automated Forces (OneSAF) Objective System (OOS).
4) **Virtual Simulations** – Simulation technology which allows users to interact with a computer-simulated environment, real or imagined. Most current virtual reality environments are primarily visual experiences, displayed either on a computer screen or through special or stereoscopic displays; some simulations include

additional sensory information, such as sound through speakers or headphones. There are numerous examples in all of the services of these types of virtual simulators for aircraft, tanks, and individual soldiers.

5) **Non-Combat Simulations** – Simulations that incorporate social or political issues. Social-political-military simulations take a different approach to their purely military counterparts. Since they are largely concerned with policy or social issues rather than battlefield performance, they tend to be less prescriptive in their operation. However, various mathematical techniques have arisen in an attempt to bring rigor to the modeling process (Bayliss, 2002).

The list gives a general categorization of today's existing simulations. Often the definitive line between the categories is blurred. Many of the simulations listed above cross the category boundaries and potentially model many aspects. The analysis team must weigh the pros and cons of each simulation as part of the tool and technique selection process.

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- ❑ **List exploratory analysis goal and objectives.** A written goal and list of objectives for conducting an exploratory analysis. This will direct the requirement for certain analysis tools and techniques.

- ❑ **List of decision criteria.** The list of decision criteria established in the prior step.

**(5) <u>Actions</u>**. The following actions should be taken by the analyst in this step:

- ❑ **Identify possible modeling and simulations.** The analyst will determine possible modeling and simulations that can be used for conducting exploratory analysis.

- ❑ **Identify possible tools and techniques for analysis.** The analyst will determine possible tools and techniques for conducting a statistical analysis of the M&S results.

**(6) <u>Products</u>**. The following products should be developed as a result of this step:

- ❑ **List of potential M&S tools.** This list will determine what M&S tools will be considered for exploratory analysis.

- ❑ **List of potential analysis tools and techniques.** This list will help to determine how to analyze the M&S outputs in order to make a determination on critical test factors.

### C.3.3 Analysis Tool Assessment - CTM 1.2.3.2.1.3

**(1) <u>Purpose</u>**. The assessment process provides the analyst measurable terms by which to evaluate the suite of potential M&S tools and techniques.

**(2) <u>Discussion</u>**. The assessment process provides the analyst measurable terms by which to evaluate the suite of potential tools. The assessment process includes an evaluation of the strengths and weaknesses of the simulation/model. The analysis team must also take into account the time available to conduct the study. The analysis team must assess the entire analysis lifecycle while assessing the various tools available for use. The availability of resources and the capabilities or skill set of each resource is a critical aspect of the assessment.

This is not always a simple process. In order to help the team identify the critical elements needed to complete a study, it is often useful to break the exploratory analysis into distinct phases. The analysis team can examine how effectively the tools or techniques will support each particular phase and aid in capturing those elements important to completing the exploratory analysis. In examining the different phases of the exploratory analysis, the analysis team essentially takes into account all the aspects of the tasks and the additional time necessary to complete the tasks in that phase. The objective is to budget the available time carefully by assessing what needs to be completed with the tools and techniques being utilized, and developing an understanding the capabilities and skills of the analysis team. This is critical to the successful completion of the exploratory analysis.

For the purpose of identifying tools, the exploratory analysis can be decomposed into four basic phases:

1) Model development
2) Execution of the model
3) Analysis of the results
4) Presentation of the results

When examining the model development phase, the assessment will evaluate the strengths and weaknesses of the tools. The primary focus will highlight the capabilities of the tool and ultimately look at what the tool can or cannot model. This is not the only consideration when assessing the model development phase. Other characteristics of the model must also be examined to include ease of use, cost, availability, and computing resources. These characteristics ultimately contribute to the overall time necessary to complete the required tasks. Another key consideration during this phase is the time required to develop the scenarios or vignettes balanced against the capabilities or skill set of the analysts in using the tool. High resolution models can be very resource-intensive when developing scenarios or vignettes. Fully understanding the resource requirements is essential to the assessment process.

The model execution phase includes several aspects. It is primarily focused on the actual execution of the model runs. The execution of the model phase begins with the development of the DOE. Once the DOE is established, the DOE must be transformed so the input factors are in a format that is readable and executable by the model. Each tool is somewhat unique with regard to input file format. The time required to transform the DOE into a useable input format is variable for different tools and must be examined. The resolution of the model combined with the number of design points and replications required in the DOE can also have a significant impact on the overall the time required to execute the desired DOE. The number of DOEs executed as part of the exploratory analysis plan must be accounted for in the overall time required for this phase. These aspects are important in assessing the tools and techniques.

The analysis phase often requires more time than most analysts expect. There is a direct correlation between time and the number of DOE input factors, replications, and the quantity of output data that must be analyzed. The model will output the raw data in a particular format. The output files must be transformed into a format that is useful to the analysis team. The format will depend on which data analysis tools were chosen by the analysis team. Following the transformation of the output data to usable form, the raw output data must be manipulated in manner that supports the measures. The transforming of the output files and manipulation of the raw output data into useful data can be a time-consuming process; time required for this process must be considered in the assessment process.

The final phase culminates with analysis team summarizing or drawing conclusions from the completed analysis. The overall study or analysis objectives come together in some form of a deliverable product to the decision maker. Although this phase by itself is not tool- or technique-dependent, the ease at which a particular tool or technique supports this process is very much model-specific.

**(3) <u>Tools and Techniques</u>**. The analyst, knowing the study analysis requirements, can create a matrix comparing the assessment tools with against the key criteria. The scoring can be a simple + / - approach or a more elaborate weighting scheme. This process will allow the analyst to have a more complete understanding of the limitations and capabilities of the potential tool suite. This approach will also allow the analyst to determine whether one tool will satisfy the study needs or if a combination of tools will be required.

The approach can be as simple as developing a matrix of the criteria, issues to be resolved, questions to be answered, or factors to be studied, with a weighting of their importance. The team can rank the potential set of tools with regard to satisfying the particular needs. This allows the analysis team to evaluate the possible approaches and tools holistically. Table C-1 shows a simple example tool-comparison matrix that ranks the importance of each of the criteria by number of checkmarks, and scores each of the tools against each criterion using “+” signs, with more being better for both the checkmarks and the “+” signs. The total score for each tool is calculated by multiplying the number of checkmarks for a particular criterion by the number of “+” signs, and then

summing across all criteria. This is a very simple example of a multi-attribute decision-making technique. More complex techniques exist; for the purposes of this comparison, simple techniques are recommended.

**Table C-1. Example Tool Comparison Matrix**

| Criteria | Weight | Tool #1 | Tool #2 | Tool #3 |
|---|---|---|---|---|
| Factor representation | ✓✓✓ | +++ | + | + |
| Achievable measures | ✓✓✓ | ++ | +++ | ++ |
| Data output | ✓✓ | + | ++ | +++ |
| Ease of use | ✓ | +++ | + | ++ |
| Cost | ✓ | + | ++ | + |
| Availability | ✓ | ++ | + | +++ |
| **Total Score** | | **23** | **20** | **21** |

It is important for the analyst to note that some tools may be included even if they score poorly in a comparison. For example, consider a team examining a complex command and control (C2) capability. They may consider a tool that models the C2 network well, but does not represent other key factors and scores poorly for the other criteria. However, if that model is the only one of those under consideration capable of representing the C2 network, the team may choose to include that model anyway, even if it scores poorly overall. Thus, the analyst must also ensure coverage of all the factors and measures by the suite of tools, not the capabilities of each tool in isolation. A table like the one shown is good for an initial cut on the relative ranking of the models. Additional evaluation is required to select the most appropriate set of tools and techniques to use. The next step will capture that final mapping of factors and measures to the tools being used.

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- ❑ **List exploratory analysis goal and objectives.** A written goal and list of objectives for conducting an exploratory analysis. This will direct the requirement for certain analysis tools and techniques.

- ❑ **List of decision criteria.** The list of decision criteria established in the prior step.

- ❑ **List of potential M&S tools.** This list will determine what M&S tools will be considered for exploratory analysis.

- ❑ **List of potential analysis tools and techniques.** This list will help determine how to analyze the M&S outputs in order to make a determination on critical test factors.

**(5) Actions**. The following actions should be taken by the analyst in this step:

- ❑ **Assess each M&S tool.** The analyst should use the list of decision criteria to assess each M&S tool for possible use in exploratory analysis.

- ❑ **Assess each analysis tool.** The analyst should use the list of decision criteria to assess each analysis tool for possible use in exploratory analysis.

**(6) Products**. The following products should be developed as a result of this step:

- ❑ **M&S tool comparison matrix.** A matrix of M&S tools rated against decision criteria.

- ❑ **Analysis tool comparison matrix.** A matrix of analysis tools rated against decision criteria.

### C.3.4 Analysis Tool Selection - CTM 1.2.3.2.1.4

**(1) Purpose**. The analyst will use the analysis tool assessments developed in the previous step to compare the potential tools and make a selection based on the results.

**(2) Discussion**. The analyst must determine what M&S tools are needed for exploratory analysis as well as the analysis tools and techniques for evaluating the M&S results. More than one of each may be selected to improve the validity of results, however additional tools will require additional analysis and resources. It is most likely the analysis team will select one statistical package and one M&S tool to meet the exploratory analysis requirements.

**(3) Tools and Techniques**. Upon selection of a tool or suite of tools, the analyst must begin examining the best approach to utilizing the selected tool set. The analyst will map each of the chosen tools and techniques to the factors it can model and the measure responses it can provide. The mapping of tools and techniques is essentially a subset of the selection process. It is a process to focus more clearly on what can be modeled, or the factors that can be evaluated against the response surface or output of the model. It will focus on the basic question of whether the output can provide insights to the questions being asked.

This mapping should be done systematically and captured in some form. At first, the mapping can be fairly simple as shown in Table C-2. In that example, the analyst identifies, for each tool and for each measure and factor, whether or not the tool can support it. The assessment is fairly high-level; in this case, either a yes, no, or qualified yes. Later, this same matrix can be used to provide greater detail about how each factor and measure will be captured by the model. Using a table such as this allows the analysis team to see, at a glance, the coverage that they will have for each of the factors and measures. This will help guide their future efforts, particularly deciding upon appropriate DOEs and the analysis approach. Any such tracking can be used along with more

complex assessments; Table C-2 is provided only to demonstrate the basic concepts of the step.

**Table C-2. Example Mapping of Tools and Techniques to Factors and Measures**

| Factors/Measures | Tool #1 | Tool #2 | Tool #3 |
|---|---|---|---|
| Factor #1 | Yes | Yes | Yes, indirectly |
| Factor #2 | Yes | No | Yes |
| … | … | … | … |
| Measure #1 | Yes | Yes, indirectly | Yes |
| Measure #2 | No | Yes | Yes |
| … | … | … | … |

**(4) Inputs**. The following inputs are required for this step:

- **M&S tool comparison matrix.** A matrix of M&S tools rated against decision criteria.
- **Analysis tool comparison matrix.** A matrix of analysis tools rated against decision criteria.

**(5) Actions**. The following actions should be taken by the analyst in this step:

- **Decide on M&S tool set.** The analyst will decide what M&S tools will support the exploratory analysis requirements.
- **Decide on analysis tool set.** The analyst will decide what analysis tools will support the exploratory analysis requirements.

**(6) Products**. The following products should be developed as a result of this step:

- **M&S tool set.** A list of M&S tools required for the exploratory analysis process. If more than one M&S tool is required, a mapping of tools to test factors will be included (see Table C-2 for example).
- **Analysis tool set.** A list of analysis tools and techniques required for the exploratory analysis process.

## C.4 Instructions for Conducting Exploratory Analysis - CTM 1.2.3.2.2

This process is a key step in the exploratory analysis. The team examines and analyzes the data to draw the insights necessary to achieve the primary goal of the exploratory analysis – to narrow the JMe factor test to a set of potential test trials. The team will reduce and organize the data, as well as evaluate the individual measure responses and their relationships. Figure C-5 shows the three primary steps of this process.

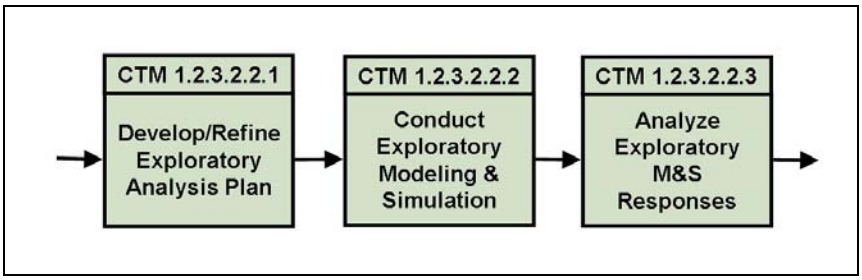



**Figure C-5. CTM 1.2.3.2.2, Conduct Exploratory Analysis**

In ***CTM 1.2.3.2.2.1, Develop/Refine Exploratory Analysis Plan***, the analysis team must develop or refine a plan for conducting exploratory analysis that will result in prioritization of SoS and conditional test factors. Details of this step can be found in section C.4.1.

***CTM 1.2.3.2.2.2, Conduct Exploratory Modeling and Simulation***, involves actual M&S of the capability under the JOC-T that will provide insights on factor sensitivity with the dependent test variables (Mission MOEs, Task MOPs, and MOSAs). This process is discussed further in section C.4.2.

In ***CTM 1.2.3.2.2.3, Analyze Exploratory Modeling and Simulation Responses***, the analyst must evaluate the exploratory M&S responses to determine which factors have the greatest impact on the dependent test variables. This process is discussed further in section C.4.3.

### C.4.1 Develop/Refine Exploratory Analysis Plan – CTM 1.2.3.2.2.1

**(1) <u>Purpose</u>**. To develop or refine an exploratory analysis plan that will identify the analytical methods and processes for exploratory analysis.

**(2) <u>Discussion</u>**. Once the team chooses the modeling tools and techniques, the team must decide how to employ the tools to provide insights about the JMe factor test space. A key step is to identify the analytical methods that will be used by the analysis team. The primary steps within this process are shown in Figure C-6. The purpose of this process is to ensure that the team approaches the analysis systematically in order to draw valid insights.

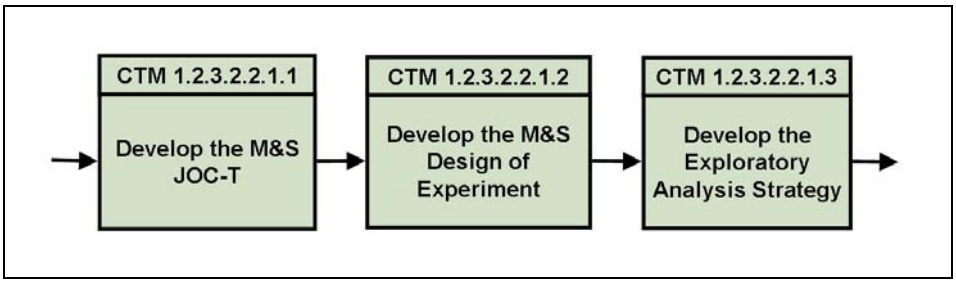



**Figure C-6. CTM 1.2.3.2.2.1, Develop/Refine Exploratory Analysis Plan**

In ***CTM 1.2.3.2.2.1.1, Develop the Modeling & Simulation JOC-T***, the analysis team develops the models to simulate the SoS in the scenarios and vignettes that are a part of the test JOC-T.

> The identification of factors, measures, tools, and experimental designs culminates in the development of an M&S scenario and vignettes. The combination of vignettes will be developed to explore the possible factor ranges within a solid DOE. The scenarios and vignettes should be developed primarily from JOC-T products, and tied to higher-level planning guidance. It is critical to the validity of the exploratory-analysis results that the M&S scenarios and vignettes draw from the same sources as the test scenario and vignettes that will be developed in preparation for the actual test.
>
> Scenarios utilized for large-scale studies are derived ultimately from direction provided in the Defense Planning Guidance (DPG). The DPG contains defense strategy and the guidance for key planning and programming priorities to execute that strategy. The DPG presents the Secretary of Defense's (SECDEF) strategic plan for developing and employing future forces. The DPG is prepared by the Office of Secretary of Defense (OSD) and published by 1 October in the odd years. The DPG is a principal product of OSD planning and reflects military advice and information recommended by the Chairman of the Joint Chiefs of Staff (CJCS); Service long-range plans and positions on policy and other matters advanced by Service Secretaries; and Combatant Commander appraisals of major issues and problems bearing on command missions. Services use the DPG to develop Service-approved large-scale scenarios that are the basis from which all studies are executed. These large-scale scenarios are often decomposed into smaller subsets to focus on the particular needs of a study.
>
> The analysis team is responsible for the development of an overall scenario upon which individual M&S vignettes will be based. The analysis team must ensure the scenario will address the key issues the team wishes to explore as part of the analysis. The team must also ensure the basis of the vignettes can be traced back to joint and Service approved scenarios. The JOC-T effort addressed in Section 4.4, Joint Operational Context for Test, should serve as the intermediate step between the higher-level DPG and Service-specific scenarios and the M&S scenario used during the exploratory analysis. The JOC-T should be the primary reference for the analysis team as they develop the scenario (or scenarios) to serve as the M&S context. By tracing to JOC-T, the team ensures the scenario used will be consistent with the scenarios developed for the actual test. It is essential the M&S and test scenarios draw from the same authoritative sources; otherwise, the results of the exploratory analysis may be called into question when compared to the test results. Ideally the analysis team should try to involve future members of the test environment and white cell teams, who will be responsible for developing the actual test scenarios and vignettes. The scenario chosen can have a tremendous impact on the system/SoS and capability performance. Different systems/SoS may contribute differently to the mission based upon the particular scenario. The team must ensure the validity of the M&S scenario.

Rarely will the analysis team conduct the exploratory analysis with a fully-developed, large-scale scenario. These are often too complex and can be difficult for the analysis team to model at the onset of an effort. These scenarios often lack the fidelity required to model the scenario into a simulation. To reduce the complexity and focus the efforts on a particular set of questions and issues, the analyst will develop smaller subsets of a large scenario, known as vignettes. The analysis team decomposes the large scenario into a series of smaller vignettes, ensuring the measure framework and JMe factor test space are being addressed. The development of the individual vignettes is tied closely to the chosen experimental designs that will capture all of the factors in the evaluation framework. The team must ensure the vignettes address all of the factor variations required by the DOE to be used. Vignette development is a thought-intensive process which requires expert input and validation. As with M&S scenario development, the analysis team should try to involve future members of the test environment and white cell teams in this process.

While this discussion has been focused on the development of scenarios and vignettes for M&S, the reader should note that the same scenarios and vignettes should be used for the other models and techniques as well. For instance, when asking questions of a key stakeholder or subject matter expert (SME), the team should place the questions in the context of a scenario or vignette, since these can have a tremendous impact on the responses provided. Doing so facilitates the synthesis of M&S results and the results obtained via other analytical methods.

Once the analysis team has developed the vignettes, they should map the vignettes to the DOE and other analytical methods. By mapping the vignettes to the analytical methods, the team is also indirectly mapping the vignettes to the JMe factor test space. This mapping provides traceability back to the evaluation framework, and, ultimately, to the test issues and JOC-T. Traceability is essential to the evaluation strategy development and makes the subsequent results more defensible.

In ***CTM 1.2.3.2.2.1.2, Develop the Modeling & Simulation Design of Experiment***, the analysis team identifies and selects experimental designs to explore the problem space. More detail on DOE processes can be found in annex D.

In general, when utilizing the exploratory analysis approach, the analysis team desires designs that can look at a large number of factors, isolate interactions, identify non-linearities, and find thresholds where responses change dramatically. DOE provides the capability to achieve those goals by systematically varying the factor levels and observing the resulting responses. The goal is to identify and develop a design, or set of designs, that can efficiently uncover new insights and effectively communicate findings to decision-makers (Hughes, 1997). The analyst draws upon the original factor set, which contains all of the factors of interest, with their characterization, ranges, and levels. That factor set, in conjunction with the mapping of the factors to tools from the previous step, provides a logical connection between the previous work and the development of a set of experimental designs. The team may choose to

develop a single experimental design per tool to be used or possibly even multiple designs for use with a single tool. The team should keep track of which factors are addressed by which DOE to ensure coverage of all of the factors of interest.

The choice of DOE depends upon many considerations, for example, the types of insights desired; interest in main effects, or two- or more-way interactions; the number of factors and their type (quantitative/continuous, quantitative/discrete, qualitative); simulation run times; and computational budget. There are many experimental-design techniques; the following list is by no means comprehensive:

- Full factorial designs
- Fractional factorial designs
- Box-Behnken designs
- Central composite designs
- Latin squares
- Taguchi robust designs
- Mixture designs
- D- and A-optimal designs
- Nearly orthogonal Latin hypercubes
- Fractional factorial controlled sequential bifurcation

Figure C-7 shows a flowchart (from Annex D) that can be followed for factor classification and initial design selection. Annex D contains more detail concerning the use of experimental designs, and includes discussions of the processes involved; the strengths, weaknesses, and assumptions of particular experimental design techniques; common pitfalls, and methods for analyzing the output of the designs. The discussions in this section and in the annex are appropriate not only for the exploratory analysis, but also for the actual test design in later CTM processes.

Experimental design may not be appropriate for all tools selected for use in evaluating the JMe factor test space. For instance, it is critical that the analysis team query key stakeholders concerning their opinions of the importance of the factors. Not only does this facilitate buy-in from the potential decision makers, it also provides an opportunity to evaluate factors that might not be suitable for models and simulation. In such cases, the team may choose to use techniques such as questionnaires, interviews, and expert panels to gather their input. Clearly, such data-collection instruments do not lend themselves to traditional experimental design; therefore, the team must identify other analytical methods. For instance, multi-attribute decision making (MADM) can be used to model stakeholder preferences based upon both qualitative and quantitative input. Value-focused thinking is another technique for structuring stakeholder objectives and evaluating the importance of factors. Depending upon the particular decision-analysis model chosen, the team can craft the data-collection instruments, interviews, and panels properly to obtain the required data. Decision-analysis techniques are only one set of potential analytical methods that can be used in addition to experimental design. Depending upon the

type of issues being addressed, other models and techniques may provide unique means of evaluating and refining the JMe factor space. As before, the team should keep track of which factors are addressed by which analytical methods to ensure coverage of all of the factors of interest.

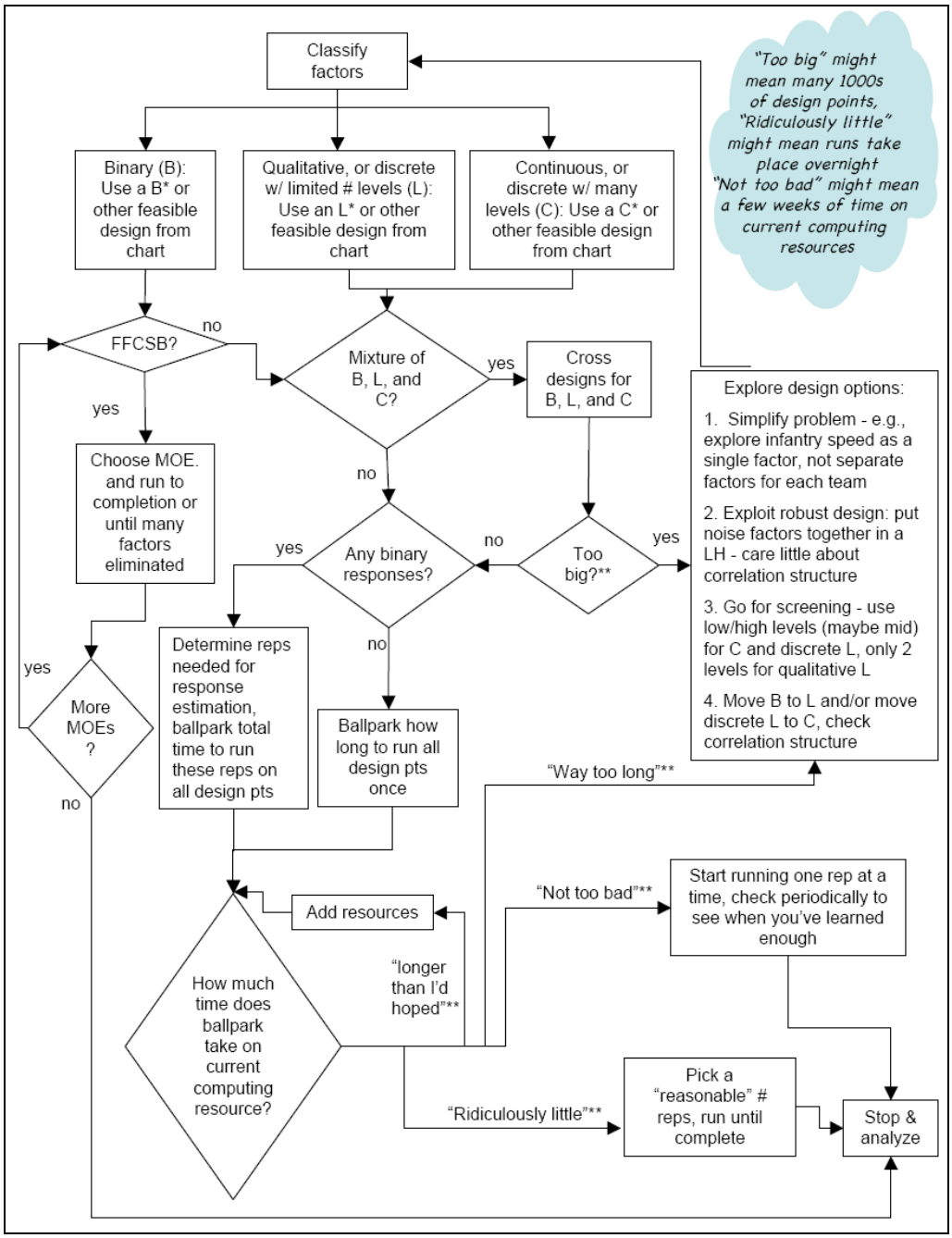



**Figure C-7. Flowchart of Initial Design Selection Process**

In ***CTM 1.2.3.2.2.1.2, Develop the Exploratory Analysis Strategy***, the analysis team develops a strategy for model-test-model that will achieve the exploratory analysis goal and objectives.

Throughout the exploratory-analysis planning process, the analysis team will begin to develop and refine the overall approach to be used for conducting the exploratory analysis. As the team identifies a suite of tools and techniques, the analytical methods to be used, and the scenario and vignettes, they begin to formulate a holistic exploratory approach they will use to conduct the analysis.

The analysis team should have a clear understanding of the measures or factors that can and cannot be modeled after completing ***CTM 1.2.3.2.1, Select Exploratory Analysis Tools***. It is rare, if not impossible, for the analysis team to be able to capture the entire evaluation framework identified initially. The JMe factor test space and/or measure framework may require modification based upon what can or cannot be modeled (if this is the first iteration), or based upon any previous iterations. In nearly all cases, the analysis team will need to make a set of assumptions based on what cannot be modeled in the evaluation framework. The team should clearly document the assumptions made and logic behind the decisions. These assumptions are necessary, but inevitably will be the basis upon which the eventual outcome of the analysis may be challenged. This further amplifies the importance of documenting the assumptions. Ultimately, the assumptions fill the gaps and allow a complete framework.

After the initial identification of tools and techniques, the team may have to identify the factors and measures that will not be evaluated. It is important that the team not remove such factors and measures completely, but document the modification. The initial evaluation framework was developed based upon the analytical requirements. Since the analytical requirements have not changed, but the analytical capabilities have, it is important to make the distinction between the two frameworks.

After subsequent iterations, the team will begin to narrow the JMe factor space by removing factors that have demonstrated little or no significant impact, or by identifying narrower ranges of factor levels to specific ranges of significant impact. In some cases, the analysts may realize that the factor ranges were not wide enough, particularly if the extremes of the ranges produce interesting effects. The team may need to revisit the techniques they used to identify the range boundaries and expand them if the ranges remain in the feasible region. This is the goal of the exploratory analysis. The team may also gain insights about the measures themselves that could require a modification to the measure framework. The team should maintain a record of the sequential modifications to the evaluation framework to provide traceability from the original evaluation framework ultimately to the recommended test-trial sets.

With the tools and techniques chosen, the team can begin to determine the analytical flow. The early or initial stages of a study are often not conducive to utilizing a single model or tool that captures the entire framework. The analysis team often must make choices on how to address the evaluation framework using multiple models, experimental designs, and runs.

During the processes of identifying and selecting tools and techniques and choosing analytical methods, the analysis team will begin to develop an overall approach to conducting the exploratory analysis. Potential approaches may include one or a combination of the following methods:

1) Running multiple models in parallel, focusing on different subsets of factors;
2) Running multiple models in parallel, focusing on the same subset of factors in order to compare the results and provide cross-validation of the models;
3) Running multiple models sequentially, using the results of a particular model initially to provide inputs for another model for subsequent runs;
4) Running a single model sequentially, narrowing the JMe factor test space of interest during each subsequent run.

In almost all cases, they will likely use a combination of approaches to provide a robust examination of the factor space. The approach used will depend upon the desired insights, model capabilities, and the resources available (including time). Key to this step is the development of an approach that leverages the strengths of the models given the factors and measures of interest. In any case, it is unlikely that a single design can be used to explore a complex JMe factor space thoroughly; some sequential runs are likely to be required.

The analysis team can now put together an overall outline that provides an overview of the plan or roadmap for achieving the desired outcomes of the exploratory analysis. The exploratory analysis outline should capture the tools and techniques to be used, including the mapping of the tools/techniques to the evaluation framework; the chosen experimental designs and other analytical methods; the computing resources required; the scenarios and vignettes, including a mapping of the vignettes to the analytical methods; and the overall exploratory analysis approach to be used.

This is a working document that can and should be updated between iterations. An evaluation at the conclusion of each phase or iteration of the study is necessary to determine if the desired outcome was achieved. The strategy outline should be updated to reflect any necessary changes needed to address gaps or shortcomings in the phase or iteration.

**(3) <u>Tools and Techniques</u>**. Various tools and techniques can be considered for the exploratory analysis. Some are listed as:

- Alternative identification, evaluation, and comparison techniques
- Models
- Simulation
  - City simulators/ urban simulations
  - Engineering, technology or process simulations
  - Military simulations
    - Agent-based models

    - Constructive combat simulations
    - Human-in-the-loop simulations
    - Virtual simulations
    - Non-combat simulations
  - Robotics simulators
- Experimental design
  - Full factorial designs
  - Fractional factorial designs
  - Box-Behnken designs
  - Central composite designs
  - Latin squares
  - Taguchi robust designs
  - Mixture designs
  - D- and A-optimal designs
  - Nearly orthogonal Latin hypercubes
  - Fractional factorial controlled sequential bifurcation
- Decision analysis techniques
  - Multi-attribute decision making (MADM)
  - Survey development and design
  - Interview techniques
  - Profession military judgment (PMJ) panels
- High performance computing clusters (HPCC) and techniques
- Scenario development and generation tools

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- ❑ **List exploratory analysis goal and objectives.** A written goal and list of objectives for conducting an exploratory analysis. This will direct the requirement for certain analysis tools and techniques.

- ❑ **M&S tool set.** A list of M&S tools required for the exploratory analysis process. If more than one M&S tool is required, a mapping of tools to test factors will be included (see Table C-2 for example).

- ❑ **Analysis tool set.** A list of analysis tools and techniques required for the exploratory analysis process.

- ❑ **Completed OV-1** with mission and task information included.

- ❑ **Mission decomposition.** List and mapping of the mission statement, mission objectives, mission desired effects, and known Mission Measures of Effectiveness (Mission MOE).

- ❑ **Task decomposition.** List and mapping of joint and Service tasks that will be performed by the capability under test.

- ❑ **Conditional factors.** List of threat and environmental condition factors that may be considered in the evaluation strategy. Include a determination of levels for each factor.

- ❑ **SoS factors.** List of test factors for the SoS with identified levels (values) for each factor.

- ❑ **List of test measures sub-divided by Mission MOE, Task MOP, and MOSA.** Include identified standards, threshold values, priorities, weightings, and relationships.

**(5)** **Actions**. The following actions should be taken by the analyst in this step:

- ❑ **Collect inputs.** The analyst will collect the inputs that define the evaluation framework; need the independent and dependent variables that are a part of the capability T&E.

- ❑ **Develop M&S test scenario.** The analyst will ensure the overarching test scenario (or scenarios) is developed in the models from authoritative guidance and the JOC-T products.

- ❑ **Develop M&S vignettes.** The analyst will ensure required vignettes within the scenario context are developed in the models in order to ensure that the required factors and their levels are addressed.

- ❑ **Map vignettes to factors.** The analysis team will map the vignettes to the test factors and their levels in order to ensure that all of the factors are addressed in the vignettes.

- ❑ **Develop experimental designs.** The analysis team will identify and select experimental designs to explore the problem space.

- ❑ **Develop other methods.** The analyst will identify other analytical techniques that may be required; particularly for tools that are not conducive to experimental design.

- ❑ **Identify general analysis design.** The analyst will capture the team's plan for the sequential or simultaneous execution of model runs and experimental designs. This step is continually refined as the exploration strategy is developed.

- ❑ **Create exploratory analysis strategy.** The analysis team consolidates the exploratory analysis design into a single document that addresses the tools and techniques, experimental designs and other analytical methods, scenarios and vignettes, and analysis approach to be used.

**(6) Products**. The following products should be developed as a result of this step:

- ❑ **Exploratory analysis strategy.** The exploratory analysis strategy contains a summary of the tools and techniques to be used, including the mapping of the tools/techniques to the evaluation framework; the chosen experimental designs and other analytical methods; the computing resources required; the scenarios and vignettes, including a mapping of the vignettes to the analytical methods; and the overall exploratory analysis approach to be used. It is a compilation of numerous outputs from the exploratory analysis planning process.

### C.4.2 Conduct Exploratory Modeling and Simulation – CTM 1.2.3.2.2.2

**(1) Purpose**. The process involves actual M&S of the capability under the JOC-T that will provide insights on factor sensitivity and critical factor levels that impact the dependent test variables (Mission MOEs, Task MOPs, and MOSAs). The analyst must configure and populate the appropriate models and data collection instruments, and use them to collect data.

**(2) Discussion**. This is the execution step of the exploratory analysis. In this process, the analysis team will conduct model runs and other data-collection events in accordance with the exploration strategy outline to obtain the data required for the subsequent analyses. The steps involved in this process are shown in Figure C-8.

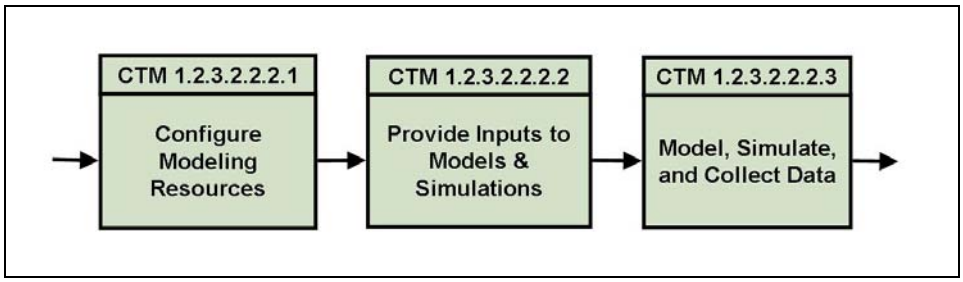



**Figure C-8. CTM 1.2.3.2.2.2, Conduct Exploratory Modeling and Simulation**

In ***CTM 1.2.3.2.2.2.1, Configure Modeling Resources***, the analysts acquire, configure, and test the required computing resources to ensure that the resources are in place and working prior to the model runs.

In this process, the team configures computing resources required for the model runs and data collection. In many cases, due to the large number of factors being explored, exploratory runs may number in the hundreds, thousands, or more, depending upon the number of factors and the complexity of the design. Such experiments are often

run on high-performance computer clusters, if available. The effort required to configure such computing resources can be quite significant.

In many cases, this step may begin with the acquisition of the required hardware and software. The hardware must be set up and the software installed and tested. Even simulation experiments conducted at a single workstation require prior planning to ensure the computing capability works as required during execution. For more complex configurations, prior planning is a necessity and may require information technology (IT) expert assistance. The team must adhere to DoD, Service, and organization IT guidelines and regulations. This may require advance notice and the submission of appropriate paperwork and forms. The requirements are more stringent if the analysis team is conducting model runs using data that are classified SECRET or higher, or if the team is operating the models in a distributed manner across a network.

***CTM 1.2.3.2.2.2.2, Provide Inputs to Models and Simulations***, entails setting the factors and model runs as required in the DOE and as planned in the Exploratory Analysis Strategy.

In this step, the team instantiates the vignettes into the models; develops the necessary input files to vary the parameters through the required design points according to the experimental design; and configures the software to provide the required data outputs. Instantiation of the vignettes includes the development of unique modeling constructs to ensure that the model represents the factors of interest. Developing the input files requires the analysis team use the specified DOE to create input files with each row of data that specifies a trial with the factors set at various levels. Multiple iterations or runs of each trial may be a part of the DOE, which may require additional configuration of the input files.

Ideally, the analysis team will have access to a tool designed specifically for creating DOE input files. Existing software packages allow the analyst to identify the factors and map them to specific parameters in the model; select appropriate DOE; specify the levels through which the factors will vary; specify the number of runs; and distribute runs within a computing cluster or across multiple processors. Such tools, however, are designed to interoperate with particular, pre-identified models. They can often be modified to accommodate new models, but this must be identified early (in the tool identification and selection step) to plan for the modification of the software. For these types of DOE software products, the analysis team may have to create input files using the tool that will be used to execute the simulation models. That effort will occur in this step.

Many government and commercial simulations have a built-in functionality for varying parameters within a DOE and specifying multiple runs of design points. If so, that capability can assist the analysis team as well. For such simulations, this step will involve the specification of the design and factor levels, as well as the selection of the number of runs for each factor configuration, or design point.

In addition to preparing the input files, the team must configure the model to provide the data output required to evaluate the measures. Some of the DOE software packages discussed earlier provides the capability to do this. Most simulations themselves also allow the user to specify the data to be collected from each run. In addition to selecting the type of data to be collected and the intervals at which the data should be collected, the user should also identify the format, file type, and destination drives for the output files.

In ***CTM 1.2.3.2.2.2.3, Model, Simulate, and Collect Data***, the analysis team conducts model runs, executes other data collection instruments, and collects the data that will be analyzed in the next process. Once the models and other data-collection instruments are prepared and the computing resources are configured, the model runs and data collection are executed according to exploratory analysis strategy, and the data are collected.

**(3) Tools and Techniques**. The M&S tools have been previously described. Additional tools for DOE and developing input files are described in Annex D.

**(4) Inputs**. The following inputs are required for this step:

- **Exploratory analysis strategy.** The exploratory analysis strategy contains a summary of the tools and techniques to be used, including the mapping of the tools/techniques to the evaluation framework; the chosen experimental designs and other analytical methods; the computing resources required; the scenarios and vignettes, including a mapping of the vignettes to the analytical methods; and the overall exploratory analysis approach to be used. Thus, it is a compilation of numerous outputs from the exploratory analysis planning process.

**(5) Actions**. The following actions should be taken by the analyst in this step:

- **Acquire, configure, test computing resources.** The analysis team will acquire, configure, and test the required computing resources to ensure that the resources are in place and working prior to the model runs.

- **Develop model inputs.** The analysis team will develop and configure input files for the models that utilize the DOE as planned in the exploratory analysis strategy.

- **Conduct model runs.** The analysis team executes the M&S runs on all chosen models according to the specified DOE and the exploratory analysis strategy. Depending upon the number of factors, computational load, and complexity of the models, this process may take hours, days, or even longer to execute. While the team need not watch the simulations execute, particularly if they are running the models without graphical user interfaces (GUIs), they should continue to check the progress to identify potential run-time errors or degraded performance. In

some cases, the analyst may have to vary each of the factor levels manually between DOE design points or specify different input files between sets of runs.

- **Collect data.** The analysis team will collect and consolidate the data from the disparate sources. Throughout the data-collection process, the team should check the data being developed to ensure quality and identify whether the data elements collected meet the requirements. In some cases, the team may have to make adjustments to the data-collection to ensure that quality data are developed. The data should be consolidated conveniently to facilitate follow-on data reduction and analysis.

**(6) Products**. The following products should be developed as a result of this step:

- **Exploratory analysis data.** The main product of this step is the data developed from the M&S and other data-collection events. These data should be consolidated and accessible by the team for the subsequent steps of the exploratory analysis.

### C.4.3 Analyze Exploratory Modeling & Simulation Responses – CTM 1.2.3.2.2.3

**(1) Purpose**. The purpose is to evaluate the measure responses from the exploratory runs.

**(2) Discussion**. In this step, the analysis team analyzes the individual responses for each measure at each of the three levels (Mission MOE, Task MOP, and Measure of System/SoS Attributes). For each of these measures, the analyst will explore the response region to identify key findings. Such an analysis will include an examination of the impact of the factor levels, and their interactions, on the measures. At a minimum, the team should develop regression or similar models for both expectation and variability. The latter is important for understanding which factors contribute most to the variability within the measure responses, which can later be used with the models of expectation to determine whether the factor should be examined in subsequent analyses and/or included within the recommended test trial sets. The team should try to identify factor regions that produce unexpected or surprising results. This may point the team to factor combinations that should be explored further. The steps involved in this process are shown in Figure C-9.

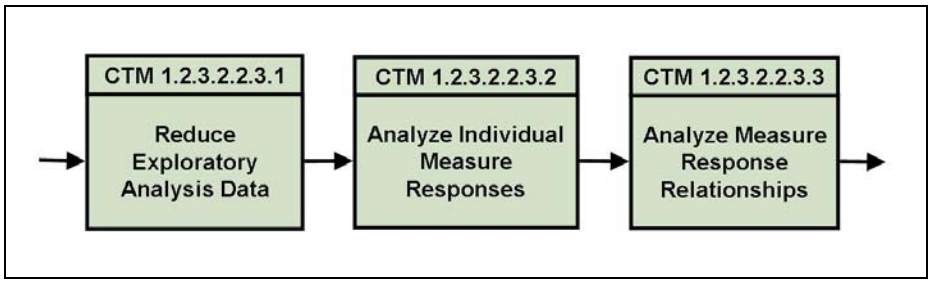



**Figure C-9. CTM 1.2.3.2.2.3, Analyze Exploratory Modeling & Simulation Responses**

In ***CTM 1.2.3.2.2.3.1, Reduce Exploratory Analysis Data***, the analysis team takes the consolidated data from the data collection efforts and manipulates the data to transform them into the elements needed for the analysis.

> Data reduction is the process by which the analysis team manipulates the data set to prepare it for subsequent analyses. The data available to the team will likely have been collected from disparate sources, such as multiple models and simulations, as well as questionnaires, interviews, expert panels, etc. The data will be of varying quality and in different forms and may require further reduction and reconstruction to be useful for analysis. Thus, the goal of this step is to get the data into a consistent and usable form.
>
> Data outputs from the models must be transformed into the particular data elements defined for each of the measures. If possible, the calculations required to convert a particular set of data elements into measures of interest may be conducted here as well. Additionally, the factor levels themselves may also have to be transformed for the analysis. The qualitative data may have to be transformed in some way as well, through binning, assigning them to particular qualitative levels, or converting them to numerical values (ordinal or nominal). Data coding can be a useful data-reduction technique, and is particularly critical for questionnaire data in which the qualitative responses must be converted, or for transforming qualitative factor levels for numerically-based analyses. Additionally, data coding ensures consistency in identifying missing data so that they can be easily recognized later. Some statistical packages provide automated means to make data coding easier, particularly when the codes are assigned to a consistent set of qualitative levels. For some data types, such as interview data, the coding must be done by hand using analyst or expert judgment.
>
> The reduced data will then have to be organized into appropriate databases and statistical software packages to facilitate the follow-on analysis. Throughout the manipulation of the data, the team must ensure that they adhere to the data management-plan to ensure that no data are lost and that changes to data are tracked.

***CTM 1.2.3.2.2.3.2, Evaluate Individual Measure Responses***, contains the bulk of the analysis of the data. In this step, the analysis team examines and analyzes the data to draw the insights necessary to achieve the primary goal of the exploratory analysis – to narrow the JMe factor test to a set of potential test trials. In this step, the team will evaluate the individual measure responses for each of the three levels of measures, to include a determination of effects of the factors on those measures.

> A test factor may impact system performance, SoS performance, task performance, and/or mission effectiveness. Each factors impact may be on one or all of the above. Therefore, the analysis must first focus on each factors impact on each measure individually.
>
> The analysis of SoS/System attribute measures (MOSA) provides insights into how each test factor impact SoS or System performance. System/SoS level data will be

used to produce quantitative and qualitative exploratory responses on key performance parameters (KPP), system/SoS measures of effectiveness (MOE), system/SoS measures of suitability (MOS), and system/SoS measures of performance (MOP).

The analysis of task measures of performance (Task MOP) provide insights on how each test factor impact joint and service tasks. The analysis team will produce quantitative and qualitative exploratory analysis results at the task level, using statistical and operational measures reflecting mission-level task outcomes. These measures evaluate mission task accomplishment by the SoS in the joint mission environment. This level of measures is one of the two levels unique to joint test and evaluation, and is therefore essential to the evaluation of joint mission effectiveness (JMe).

The analysis of mission measures of effectiveness (Mission MOE) provides insights on how each test factor impacts JMe. The analysis team will produce quantitative and qualitative exploratory analysis results at the mission level, using statistical and operational measures reflecting mission-level desired effects. These measures are used to assess changes in system behavior, capability, or operational environment tied to measuring the attainment of a mission end state, achievement of a mission objective, or creation of a mission effect. This level of measures is the other of the two levels unique to joint test and evaluation, and is therefore essential to the evaluation of JMe.

In ***CTM 1.2.3.2.2.3.3, Evaluate Measure Response Relationships***, the analyst examines the relationships between the measures responses both within a particular measure level and between levels.

While the previous analyses of individual measure responses are not simple tasks, they are fairly straightforward because they involve only one measure at a time. Examination of multiple measures and their relationships across measure levels adds a new level of complexity and is the particular aspect that makes the exploratory and post-test analyses for a joint test event unique. Unfortunately, there are no standard techniques for doing such analyses, but there are a variety of techniques from which to choose, each having its own strengths and weaknesses.

This step is critical because the analysis team must narrow the JMe factor test space using the entire measure structure. Unfortunately, this means that the analyst cannot just focus on the response characteristics of a single measure, but must consider all measures deemed important enough to be included in the evaluation framework. As previously discussed during the evaluation framework development process, the team should have chosen only those measures important to the evaluation of JMe. Thus, inclusion in the evaluation framework implies importance, and no measure should be ignored without justification.

There are two main aspects to the analysis of measure relationships. The first is an examination of the interactions and correlations among measures at the same level and between measures from different levels. The second is the development of techniques to consider the multiple measure responses from all three levels systematically to refine the JMe factor test space. The first necessarily pre-cedes the second, since the relationships between the levels must be understood before techniques can be developed to refine the JMe factor test space.

**(3) <u>Tools and Techniques</u>**. There is a wide range of available analysis tools that can be used for exploring complex factor and measure spaces, including traditional statistics and regression, classification and regression trees (CART), contour and interaction plots, response-surface methodologies (RSM), and decision-analysis techniques (particularly for SME input). These data-analysis and data mining techniques are critical given the large amount of data likely to be generated. The analyses conducted in this step should be viewed as a learning process within which the analyst is trying to learn about the factor space. It is important as well to point out that the analytical results should be compared against SME input.[1] More detail about selected analytical techniques and their application can be found in Annex B, Develop Evaluation Strategy.

**Measure Relationships at the Same Level**: One technique the analysis team can use to explore the relationships between measures at the same level is to look at all of the measure responses by DOE design point and record, for each measure, whether that measure is above or below the mean value of the responses, calculated using all model runs and DOE design points. This requires that the analyst set upper and lower bounds on the mean; only if the measure response is outside of those upper and lower bounds will it will be designated as falling above or below mean. Otherwise, without bounds, all measure responses will fall either above or below the mean, even if the deviation is insignificant. The analyst may set the bounds by using standard deviations or another method, such as those used in statistical quality control and control charts (such as X-bar and R charts). In the latter technique, if an observation is within X-bar plus or minus R the process is deemed "in control," and otherwise "out of control." The resulting table may look something like Table C-3.

**Table C-3. Example Qualitative Examination of Multiple Measures at the Same Measure Level**

| DP | M1 | M2 | M3 | M4 | M1/2 | M1/3 | M1/4 | M2/3 | M2/4 | M3/4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | + | + | 0 | + | ++ | +0 | ++ | +0 | ++ | 0+ |
| 2 | 0 | + | + | + | 0+ | 0+ | 0+ | ++ | ++ | ++ |
| 3 | + | 0 | - | - | +0 | +- | +- | 0- | 0- | -- |
| 4 | + | 0 | 0 | + | +0 | +0 | ++ | 00 | 0+ | 0+ |
| 5 | - | 0 | 0 | - | -0 | -0 | -- | 00 | 0- | 0- |
| … | | | | | | | | | | |

[1] Consult service-level evaluator and analyst guides for more information about analytical techniques and general discussions about statistical analysis results, using statistical and operational measures developed as part of the techniques: AEC (2008), Chapter 9 and Appendices C and E; OPTEVFOR (2007), Chapter 9.

Each row in the table represents a separate design point (DP). Each column in the left half of the table represents a measure (e.g., measure #1 is designated M1). A "+" in the column indicates that the measure response fell above the upper bound; a "-" indicates that the measure response fell below the lower bound; and a "0" indicates that the response fell between the bounds. The columns on the right half of the table are all of the possible pairs of measures (e.g., the combination of measures #1 and #2 is represented by M1/2). For each of the combinations, the results from the left half of the table are recorded together (e.g., both M1 and M2 fell above the upper bound for DP 1, so M1/2 has a "++" in the column). Using this technique, there are nine possible combinations of "+", "-", and "0". The team can then count, for each pair of measures across all design points, the number of times each combination appears. If the counts appear evenly distributed across the possible combinations, then the team may conclude that there are no interactive relationships between the measures. If a majority of the design points fell into the "++" and "--" categories then the team may conclude that there appears to be a positive relationship between the two measures. If a majority of the design points fell into the "+-" and "-+" categories, then the team may conclude that there appears to be a negative relationship between the two measures. This technique can be extended beyond pairs of measures to triplets, etc. Currently, this technique only provides a qualitative understanding of the potential relationships, but may be helpful for identifying areas of further exploration. This is a relatively new idea; therefore, there are no statistical underpinnings to the technique, although there appears to be potential for that.

There are other, statistically-based techniques that consider interactions and correlation between multiple measures. One technique is to calculate the correlation between the measures using standard equations. Correlation will give the team an understanding of the degree of linear relationship between the measure responses, but will not capture non-linear relationships.

Another technique that considers multiple responses and is appropriate for DOE is the Advanced Response Surface Methodology (ARSM), which is discussed further in Annex G. Other potential techniques and methods to examine the results are discussed there as well.

**Measure Relationships at Different Levels**: When exploring the relationships between multiple measure levels, the analysis team should draw upon the measure structure developed previously in the evaluation framework. With the measure structure as a starting point, there are multiple techniques that may be used to examine the relationships.[2] One qualitative technique is an extension of the one discussed above in conjunction with Table C-3. If the previously-discussed technique has already been used for each of the measure levels, the team can easily extend those results as shown in Table C-4. In this table, each of the measure responses is recorded in the same way for the lower-level measures (represented in this case by T1-T3 for

[2] The reader may also consult AEC (2008), pp 7-9 through 7-11 for a discussion of measure relationships and techniques to aggregate measures.

Task MOP), as well as the results for the higher-level measure of interest (represented in this case by M1 for Mission MOE). For each of the possible pairings between a lower-level measure (T1-T3) and the higher-level measure (M1), the results are recorded together in the right half. Using this technique, there are again nine possible combinations of "+", "-", and "0" and the team can conduct a similar analysis as that discussed previously. The technique can be extended to combinations between a higher-level measure and combinatorial groupings of lower-level measures. Again, this technique only provides a qualitative understanding of the potential relationships, but may be helpful for identifying areas of further exploration. This is a relatively new idea; therefore, there are no statistical underpinnings to the technique, although there appears to be potential for that.

**Table C-4. Example Qualitative Examination of Multiple Measures at Different Measure Levels**

| DP | T1 | T2 | T3 | M1 | T1/M1 | T2/M1 | T3/M1 |
|---|---|---|---|---|---|---|---|
| 1 | + | + | 0 | + | ++ | ++ | 0+ |
| 2 | 0 | + | + | + | 0+ | ++ | ++ |
| 3 | + | 0 | - | - | +0 | 0- | -- |
| 4 | + | 0 | 0 | + | +0 | 0+ | 0+ |
| 5 | - | 0 | 0 | - | -0 | 0- | 0- |
| … | | | | | | | |

Another technique that can be used is to apply regression to fit a model for a particular higher-level measure response (e.g., a Mission MOE). For instance, if the analysis team treats the Mission MOE response as the response to be modeled and treats the lower-level measures (e.g., Task MOP) as the regressors, they can fit a model that will indicate the impact of each of the Task MOP on the Mission MOE response. Using only a regression, though, will not be sufficient for understanding the relationships since it does not account for possible correlation between the Task MOP (multicollinearity), which may exist in many cases. Thus, the analysis team should use standard techniques to deal with multicollinearity by developing a cross correlation matrix between the Task MOP and removing those regressors that are heavily correlated with others.

One particular drawback to this technique is that the team must make the case that the Task MOP will occur before the Mission MOE, and thus this is a natural division of dependent (Mission MOE) and independent (Task MOP) variables. Thus the team must be careful in their selection of measures to examine in this way by choosing those tasks and mission effects that meet that assumption. For instance, in a call-for-fire scenario, the component tasks of the call-for-fire mission will occur before the mission effects of destroying (or not) the target. On the other hand, call-for-fire tasks would not necessarily occur before mission effects such as survivability of the firing unit. These relationships should be evident in the measure structure developed earlier, which should be consulted before executing this technique.

If the analysis team is able to directly manipulate task performance in M&S and measure the Mission MOE as output, they may have a third technique available. Instead of varying the larger set of factors in the JMe factor test space, they can choose to vary the task performance (and thus the Task MOP) directly by treating the Task MOP as the factors. The team can apply a DOE to the analysis and directly measure the impact of Task MOP on Mission MOE. The same is true between the other possible levels as well. In fact, since the other possible level-to-level analyses will involve the Measures of System/SoS Attributes (MOSA) as the lower level measures, this technique may be available in even more cases. As discussed previously, MOSA are often the factors that are varied in a simulation model anyway, instead of being the measures that are evaluated. The team may have already captured many of the MOSA as factors to be varied in the model, and so the team may be able to explore the relationships between the levels through the original DOE.

Another caveat to the use of correlation is that it does not measure causality, which is ultimately what the team would like to know. It can, however, assist the team in identifying causality by pointing out situations where it might exist, and where a causal mechanistic link can be sought.

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- ❑ **Exploratory analysis data.** The data developed from the M&S and other data-collection events. These data should be consolidated and accessible by the team for exploratory analysis.

- ❑ **Exploratory analysis strategy.** The exploratory analysis strategy contains a summary of the tools and techniques to be used, including the mapping of the tools/techniques to the evaluation framework; the chosen experimental designs and other analytical methods; the computing resources required; the scenarios and vignettes, including a mapping of the vignettes to the analytical methods; and the overall exploratory analysis approach to be used. Thus, it is a compilation of numerous outputs from the exploratory analysis planning process.

**(5) <u>Actions</u>**. The following actions should be taken by the analyst in this step:

- ❑ **Reduce exploratory analysis data.** The analysis team will process the data to transform it into a usable form for analysis.

- ❑ **Prepare analysis tools.** The analyst will ensure the analysis tools and techniques are available and ready for use. This includes the input of analysis data into the tools.

- ❑ **Evaluate individual measures.** The analyst will evaluate each measure for factor impacts. Results will be recorded in matrix form that will provide insights on measure responses.

- **Evaluate measure relationships.** The analyst will evaluate the measure relationships both within a particular measure level and between levels. Results will be recorded in matrix form that will provide insights on measure responses.

**(6) Products**. The following products should be developed as a result of this step:

- **Qualitative matrix of factor impacts on single and multiple measures.** This is a matrix of factor impacts on the measures. Those factors that are statistically significant are positive (+1) or negative (-1) depending on the direction of the measure response. Those not significant are neutral or zero in value. Multiple measures can be recorded as in Table C-3 or as a multiplier of the numerical values.

- **Qualitative matrix of factor impacts on multiple measures across levels.** This is a matrix of factor impacts on the measures across levels. Those factors that are statistically significant are positive (+1) or negative (-1) depending on the direction of the measure response. Those not significant are neutral or zero in value. Multiple measures can be recorded as in table C-4 or as a multiplier of the numerical values.

## C.5 Instructions for Drawing Exploratory Analysis Results - CTM 1.2.3.2.3

**(1) Purpose**. The purpose is to develop findings that will reduce the test space and/or identify factor combinations to explore during subsequent iterations, as required.

**(2) Discussion**. The analysis team must consolidate the findings from the previous step to recharacterize the JMe, including both the factor test space and the measure framework.

In this process, the team consolidates all of the findings from the analysis of the data and draws key insights to refine the JMe factor test space and the measure framework. The two main steps of this process are shown in Figure C-10.

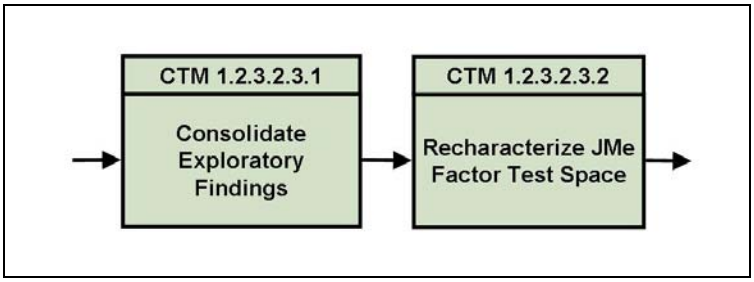



**Figure C-10. CTM 1.2.3.2.3, Draw Exploratory Analysis Results**

In ***CTM 1.2.3.2.3.1, Consolidate Exploratory Findings***, the analysis team brings together all of the key findings from the analysis and draws insights that will facilitate the refinement of the evaluation framework.

After analyzing the data, the analysis team should have a significant set of key findings concerning the effects of factors, and their interactions, on the measure framework; regions of unusual or unexpected results; and the relationship between the measures, both within and between levels. These findings will be combined in this step and organized to develop the insights required to refine the evaluation framework.

Findings and insights may be grouped into the following categories:

- Factor main effects: This category includes findings related to the significance of the factor main effects with respect to the measure framework. The findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn about which factors are most significant in terms of their main effects and should be grouped as controllable or uncontrollable.
- Factor interactions: This category includes findings related to the significance of the factor interactions with respect to the measure framework. The findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn about which factor interactions are most significant and should address whether interactions involving factors whose main effects were not significant will be carried forward to the next step.
- Factor range: This category includes findings related to the significance of the particular factor ranges explored. This may include findings of unexpected or surprising results within or at the boundaries of the factor range. As before, the findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn about whether the factor ranges for those factors moving forward should be narrowed or expanded in subsequent steps.
- Measures: This category includes findings related to the measures themselves. This may include findings that particular measure responses showed little variability across the experimental designs or that the calculation methods used for certain measures were ineffective at achieving their intent. Insights should be drawn about whether particular measures should be removed from the framework or whether the calculation methods for particular measures should be adjusted to provide more meaningful results.
- Measure Relationships: This category includes findings related to the relationships between the measures, both within and between levels. This may include findings that key assumptions made when structuring the measures were not validated by the results or that the results indicated unexpected relationships. Insights should be drawn about whether the measure structure should be adjusted by altering or adding new linkages within the original measure structure.

In addition to these categories, the team will likely make changes to the evaluation strategy outline if subsequent iterations are to be performed. The team may identify

requirements for a particular modeling capability that had not been considered or may see the need to adjust the exploratory analysis approach for future iterations.

***CTM 1.2.3.2.3.2, Recharacterize JMe Factor Test Space***, involves the actual refinements to the JMe factor test space and the evaluation framework based upon the insights from the exploratory analysis. In this final step, the analysis team uses the insights developed during the previous step to refine the evaluation framework used to evaluate JMe. Refinements may include the removal of insignificant factors for subsequent analyses, adjustments to factor ranges, removal of or changes to measures within the measure framework, and adjustments to the measure structure.

**(3) Tools and Techniques**. Once the analysis team has explored and gained insights into the relationships between measures on the same level and between levels, they must decide how they will consider the multiple responses for the purpose of refining the JMe factor test space.

One set of techniques can be drawn from the decision-analysis field, specifically multi-attribute decision making (MADM) and related techniques. In these methods, weights are assigned to each of the measures according to their importance. The measure responses are then converted to a common scale. A total score is then developed by multiplying the weight of each criterion by the scaled value of the responses and the products are summed. The particular methods used to assign weights and convert responses vary according to the specific MADM technique used. The result is a single value that represents a combined measure of JMe. MADM techniques require significant expert and decision-maker input, as the subjective assignment of weights, and the methods used to convert individual measure responses to a common scale, have a significant impact on the resulting total values.

The advantage of these techniques is that the team can now refine the JMe test factor space based upon the responses of the single response measure. There are some significant disadvantages however. First is the subjectivity of the methods. Even with significant stakeholder input and extensive sensitivity analyses, the subjectivity may make the results less defensible. Second is that many decision makers are uncomfortable with combining critical measures into a single value that has no intrinsic natural meaning. Third, multiple levels of measures complicate the model. Many MADM techniques are capable of dealing with different levels of measures; however, the relationships between the levels must be well-understood to construct the hierarchy, which is seldom the case. Additionally, there are assumptions about preferential independence that must be met when using an additive model as described above.

Another technique is to identify the number of measures for which each factor and factor interaction was significant. Table C-5 shows an example that tracks the number of measures (out of five) for which 16 factor main effects were significant. This technique allows the analysis team to prioritize the factors according to the number of measures for which the factors were significant. The team can then choose to eliminate the factors that fall into the lowest-priority categories.

**Table C-5. Example Tracking Matrix for Factor Significance by Number of Measures**

| All 5 Measures | 4 Measures | 3 Measures | 2 Measures | 1 Measure | None |
|---|---|---|---|---|---|
| Factor 3<br>Factor 4<br>Factor 10<br>Factor 13 | Factor 1<br>Factor 11<br>Factor 16 | Factor 2<br>Factor 5<br>Factor 12<br>Factor 15 | Factor 9<br>Factor 14 | Factor 8 | Factor 6<br>Factor 7 |

Of course, this method assumes that all measures are of equal importance. Therefore, it is likely that the team will want to develop a similar matrix for each of the levels of measures and then develop rules for deciding which factors will move forward to the next iteration of the exploratory analysis. As part of this, the team should also capture the specific factor-level range that showed the most significance, through methods such as CART. This technique should not replace the need for the analysis team to explore the measure responses for regions of interest, but may be an acceptable technique for deciding which factors are most important.

Regardless of the technique chosen, the analyst should consider the controllability of the factor, particularly before deciding to include the factor in the recommended test trial set. Since such factors cannot be controlled during experimentation, they cannot be factors in the test DOE. However, if the team concludes that an uncontrollable factor has a significant impact on JMe, that fact should be noted and carried forward to the test planning stage. For instance, if the training level of the participants appears to be a significant factor, but cannot be controlled in the experiment, the test planning and analysis teams will want to be aware of that insight. Having this insight available when designing the actual test event or analyzing the post-test results, may help put the results in context, particularly if the training level of the participants can be assessed prior to the test event.

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- ❑ **Exploratory analysis goal and objectives.** A written goal and list of objectives for conducting an exploratory analysis. This will direct the requirement for certain analysis tools and techniques.

- ❑ **Qualitative matrix of factor impacts on single and multiple measures.** This is a matrix of factor impacts on the measures.

- ❑ **Qualitative matrix of factor impacts on multiple measures across levels.** This is a matrix of factor impacts on the measures across levels.

**(5) <u>Actions</u>**. The following actions should be taken by the analyst in this step:

- ❑ **Consolidate findings.** The analyst will evaluate the separate measures and measure relationships to establish findings from the exploratory analysis. This

will include findings related to factor main effects, factor interactions, and factor ranges.

- ❑ **Recharacterize factor test space.** The analyst will evaluate the findings to draw insights on how to refine the test space. This includes the ability to recommend certain SoS factors as lower priority and set at stressing levels. Also includes recommending conditional factors set at specific levels to reduce the test space.

**(6) Products**. The following products should be developed as a result of this step:

- ❑ **Consolidated list of findings.** These are the findings and insights from the particular exploratory analysis iteration, grouped into the following categories: factor main effects, factor interactions, factor range, measures, and measure relationships.

- ❑ **Recommendations on test space reconfiguration.** The refined evaluation framework captures the refinements to the original evaluation framework based upon the analysis results and insights. Recommendations can be for prioritizing SoS and conditional factors, combining factors at specified levels, or setting factors at stressing levels that will have the greatest influence on measure responses.

## C.6 Instructions for Synthesizing Exploratory Results - CTM 1.2.3.3

**(1) Purpose**. The purpose is to integrate all of the analyses conducted during the multiple exploratory iterations to draw overarching insights about the refined JMe factor test space and the measure framework.

**(2) Discussion**. Prior to reaching this process, the team may have iterated the exploratory analysis multiple times to thoroughly explore the JMe test factor space. The decision to conclude the iterations is a subjective one based upon the analysis team's judgment and the exploratory analysis goal. That judgment should be based upon whether the team feels that the analysis conducted has been sufficient to get an order-of-magnitude notion of results and effects. Additionally, the team should be satisfied that the JMe factor test space has been refined sufficiently to provide the test planning team a reasonably small set of factors from which to design subsequent tests. This may require some knowledge about the test events to be conducted during the capability lifecycle. The final result should be a final JMe test factor space consisting of potential test trial sets of interest for subsequent testing. The steps that make up this process are shown in Figure C-11.

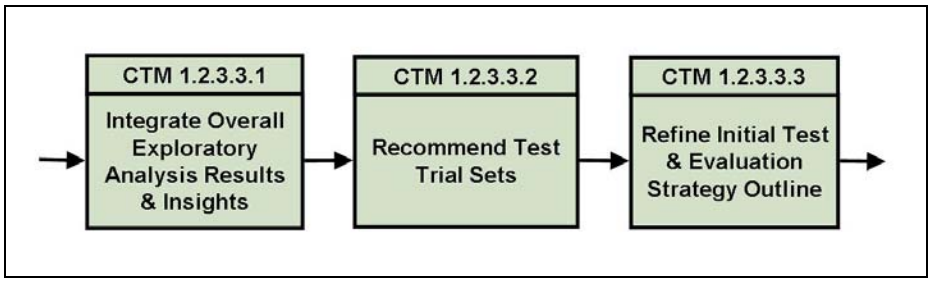



**Figure C-11. CTM 1.2.3.3, Synthesize Exploratory Results**

In ***CTM 1.2.3.3.1, Integrate Overall Exploratory Analysis Results & Insights***, the team gathers, consolidates and organizes the findings from the multiple iterations of the exploratory analysis. Just like consolidating the findings for each exploratory analysis iteration, the findings and insights may be grouped into the following categories:

- Factor main effects: This category includes findings related to the significance of the factor main effects with respect to the measure framework. The findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn that address which factors are most significant in terms of their main effects and are grouped as controllable or uncontrollable.
- Factor interactions: This category includes findings related to the significance of the factor interactions with respect to the measure framework. Again, the findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn that address which factor interactions are most significant.
- Factor range: This category includes findings related to the significance of the particular factor ranges explored. This may include findings of unexpected or surprising results within or at the boundaries of the factor range. As before, the findings may be specific to particular measures, or across the framework as a whole, depending upon the analysis techniques chosen. Insights should be drawn that address how the factor ranges were narrowed or expanded between iterations.
- Measures: This category includes findings related to the measures themselves. This may include findings that particular measure responses showed little variability across the experimental designs or that the calculation methods used for certain measures were ineffective at achieving their intent. Insights should be drawn that address whether particular measures were removed from the framework or whether the calculation methods for particular measures were adjusted to provide more meaningful results.
- Measure Relationships: This category includes findings related to the relationships between the measures, both within and between levels. This may include findings that key assumptions made when structuring the measures were not validated by the results or that the results indicated unexpected relationships. Insights should be drawn that address whether the measure structure was adjusted by altering or adding new linkages within the original measure structure.

For each of the categories, overarching insights should be developed in a way that will facilitate the development of a recommended test trial set.

***CTM 1.2.3.3.2, Recommend Test Trial Sets***, involves the development of a recommended test trial set from which subsequent test events will draw when designing their test.

The overarching insights developed in the previous step will help guide the analysis team in the development of a recommended test trial set for future test events. The test trial set should include recommended independent test factor variables and dependent test measure variables. The independent test factor variables should be derived from the controllable factors remaining in the refined JMe factor test space. In conjunction with the identification of the factors, the ranges through which the factors should be varied should also be identified. The dependent test measure variables should be taken from the refined measure framework, and should include a complete, concise, and unambiguous description; whether the measure is quantitative or qualitative in nature; the method of calculation (if quantitative) or determination (if qualitative) and the unit of measure; the data elements required to calculate or determine the measure response value (quantitative) or level (qualitative); definitions of key terms; and the desired fidelity of measurement.

In ***CTM 1.2.3.3.3, Refine Test and Evaluation Strategy***, the Test and Evaluation Strategy (TES) is refined based upon the results of the exploratory analysis. This TES is documented in the Data Analysis Plan and in the capability crosswalk matrix. This process closes the feedback loop of the exploratory analysis back into the develop an evaluation strategy process.

**(3) <u>Tools and Techniques</u>**. The tools and techniques for synthesizing the exploratory results may be based on analyst preference on how to present the results. Certain decision techniques, affinity diagrams (for relating similar insights and findings), and stakeholder analysis may be considered.

**(4) <u>Inputs</u>**. The following inputs are required for this step:

- **Exploratory analysis goal and objectives.** A written goal and list of objectives for conducting an exploratory analysis. This will direct the requirement for certain analysis tools and techniques.

- **Consolidated list of findings.** These are the findings and insights from the particular exploratory analysis iteration, grouped into the following categories: factor main effects, factor interactions, factor range, measures, and measure relationships.

- **Recommendations on test space reconfiguration.** The refined evaluation framework captures the refinements to the original evaluation framework based upon the analysis results and insights. Recommendations can be for prioritizing

SoS and conditional factors, combining factors at specified levels, or setting factors at stressing levels that will have the greatest influence on measure responses.

- ***Data Analysis Plan (DAP).*** The DAP is a document that will be developed by the analyst to capture the evaluation strategy and to explain how the analysis will be conducted. At this stage in the process, the DAP will probably be a template shell of a plan that will need to be filled in specific information pertaining to this test and evaluation.

- ***Draft Test & Evaluation Strategy.*** A draft of the TES for the capability under test.

**(5) <u>Actions</u>**. The following actions should be taken by the analyst in this step:

- **Integrate overall exploratory analysis results.** The analyst will gather, consolidate and organize the findings and insights from the multiple iterations of the exploratory analysis.

- **Develop exploratory analysis insights.** The analyst will develop insights based on the exploratory analysis that recommend certain SoS and conditional test factors be set at stressing levels so as to not be considered in future test and evaluation. This action will require presenting the results to key stakeholders for approval.

- **Recommend test trial sets.** The analyst will recommend a set of test trials for future LVC-DE testing. The recommendations will be based on exploratory analysis insights on prioritized test factors.

- **Refine initial test and evaluation strategy.** The analyst will revise the TES and DAP to reflect new test trial sets and test factors for the new DOE.

**(6) <u>Products</u>**. The following products should be developed as a result of this step:

- ***Test Trial Matrix.*** A matrix of test trials that should include recommended independent test factor variables, with the ranges through which the factors should be varied, and the dependent test measure variables, with all of their required elements.

- ***Revised Test & Evaluation Strategy.*** The draft TES will be revised to reflect the results and decisions from the exploratory analysis process.

- ***Data Analysis Plan (DAP).*** The DAP should be revised to reflect the exploratory analysis insights and approved changes to the TES. At this point, the DAP can incorporate new information on test design and measures evaluation.

*This page intentionally left blank.*

ENCLOSURE 1 TO ANNEX C
EXAMPLE EXPLORATORY ANALYSIS MODEL – TEST – MODEL PROCESS*

# Team 6: Exploring the Design Space of Command and Control Capability Evaluation Strategies


**TEAM 6 MEMBERS**
David Dryer, Ph.D. – Lead
Mark Fiebrandt – Co-Lead
Joint Test and Evaluation Methodology, US

Norman Reitter
Concurrent Technologies, US

Susan Sanchez, Ph.D.
Naval Postgraduate School (NPS), US


## Introduction

The goal of Team 6 activities at IDFW 17 was to conduct a capability evaluation strategy refinement process for testing in a joint environment (TIJE). These activities were supported by a prototyped automated capability test methodology (CTM) evaluation thread builder and agent-based simulation (ABS) tools developed by Joint Test and Evaluation Methodology (JTEM). The use case scenario for this data farming team exercise was based on a command and control (C2) joint capability area (JCA), focused on joint fires (JFIRES) and close air support (CAS) mission task threads developed as part of a JTEM 2008 test event. An ABS model was written in Visual Basic as part of an Excel spreadsheet designed to model both Blue and Red agents in a limited joint mission environment. The model allowed flexibility to consider materiel and non-materiel factors in the design of experiment (DOE) that focused on C2 processes and decisions. The exit criteria for the team included planning a C2 critical capability issue evaluation strategy design, exercising the evaluation strategy using an ABS, and conducting an initial capability analysis of the evaluation strategy at mission and task levels. This effort explored the use of automated capability analysis and simulation tools for TIJE by JTEM and its partners. Specific objectives for Team 6 at IDFW 17 were:

- Conduct a capability evaluation strategy refinement process for TIJE
- Prototype an automated CTM evaluation thread builder and ABS tool
- Utilize a "use case" scenario based on a C2 JCA, focused on JFIRES and CAS mission task threads developed as part of a JTEM 2008 test event
- Consider materiel and non-materiel factors in the DOE process
- Evaluate methods for analysis of factors across multiple measures

* Based on work conducted at an International Data Farming Workshop and resulting article written for the IDFW draft report.

## Approach

The overall approach for our C2 capability evaluation exploration involved applying CTM 1.3 Develop Evaluation Strategy processes to recommend refinements for capability-focused test and evaluation (T&E) designs (see figure C-1-1). After identifying and collecting evaluation inputs relevant to the case study's capability concepts (CTM 1.3.1), a critical capability issue was developed (CTM 1.3.2) to structure an iterative exploration of the case study's design space, which occurred as part of establishing the use case's T&E strategy framework (CTM 1.3.3). This iterative exploration involved a cycle of developing and refining efficient DOE metamodels, exercising these metamodels with an ABS model, and conducting analysis on selected ABS simulation response measures.

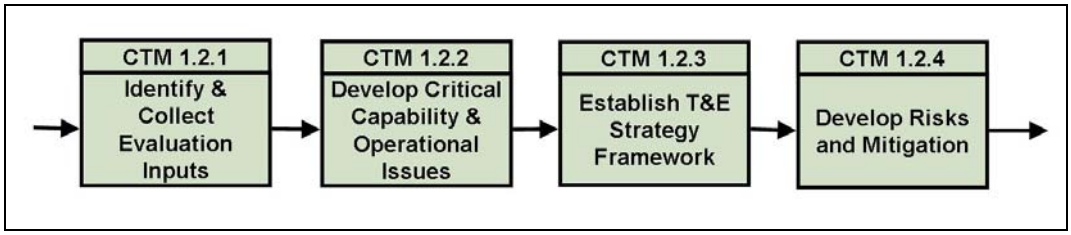



**Figure C-1-1. CTM 1.2 Develop Evaluation Strategy Processes**

The use case critical capability issue is shown in figure C-1-2. The phrasing is a reflection of this definition of capability: The ability to achieve a desired effect under specified standards and conditions through combinations of means and ways to perform a set of tasks (Chairman of the Joint Chiefs of Staff, 2007). We used a critical joint issue template of "Assess the ability to perform [Task(s)] by [System(s)/System of System(s)] under the following [Condition(s)] to achieve [Mission Desired Effect(s)]." Key concepts in this capability-centric issue phrasing include task, system of systems (SoS), condition (threat and environment), and mission desired effect.

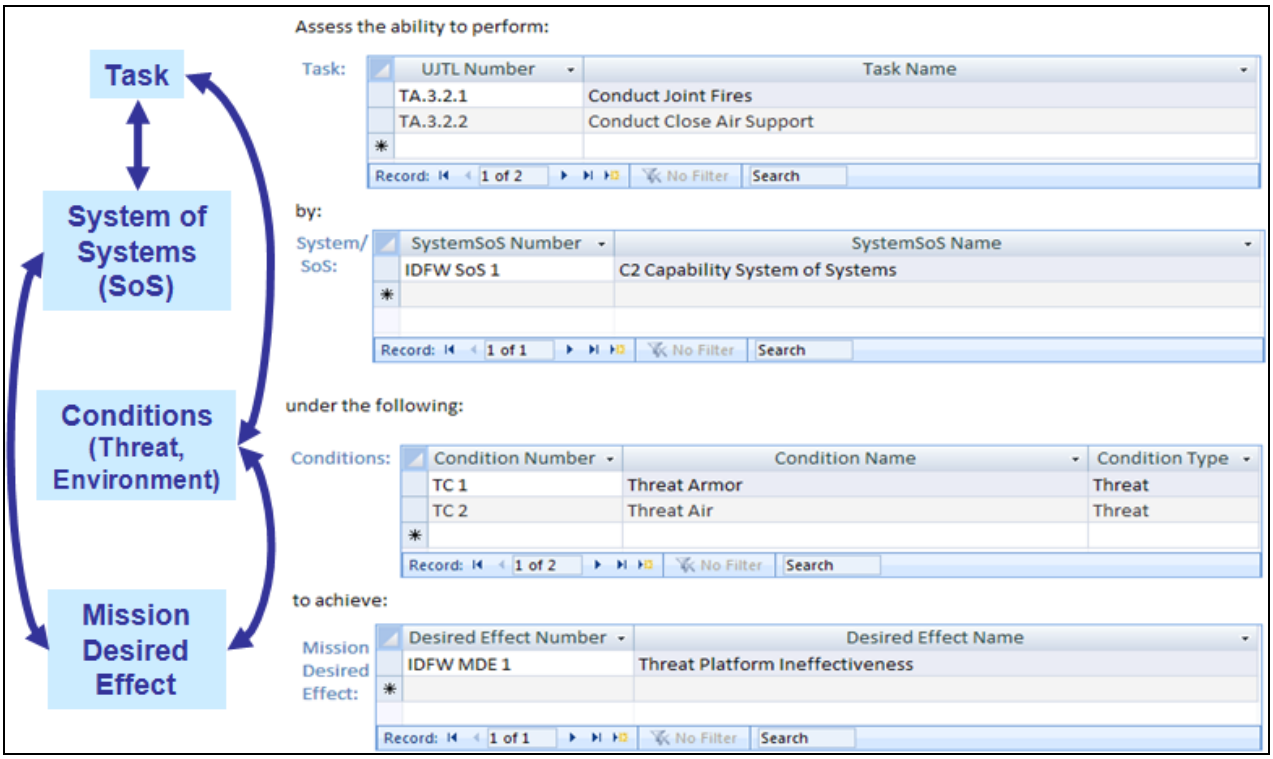



**Figure C-1-2. Use Case Critical Joint Issue**

Based on this critical capability issue, a joint operational context for test (JOC-T) was developed using key capability concepts. An overview of this capability JOC-T is shown in figure C-1-3 as a DoD Architecture Framework (DoDAF) High Level Operational Graphic (OV-1) product. The OV-1 has three main sections.

The upper part represents an SoS means and ways implementation across materiel and non-materiel aspects. The icons represent various performers and the lines represent interaction types relevant to this particular capability assessment. SoS Sensor, C2, Global Information Grid network infrastructure, and Engagers are represented. Threat air and ground targets are also represented. Lines between SoS performers are communication interactions. Lines between SoS and threat performers are sensing and engagement interactions.

The middle part of the OV-1 represents the tasks that the SoS is to perform. In this case, two joint tasks are represented in task swimlanes with a top-level activity sequence chain. Both Conduct Joint Fires and Conduct Close Air Support have similar chains involving Detect, Identify, Decide, Deconflict, Deliver, and Assess activities. Of course, each activity chain has its own set of performers in a SoS implementation.

The lower part of the OV-1 represents an intended mission desired effect to be achieved by the SoS performing the set of tasks. The mission desired effect is threat platform (air and ground) ineffectiveness.

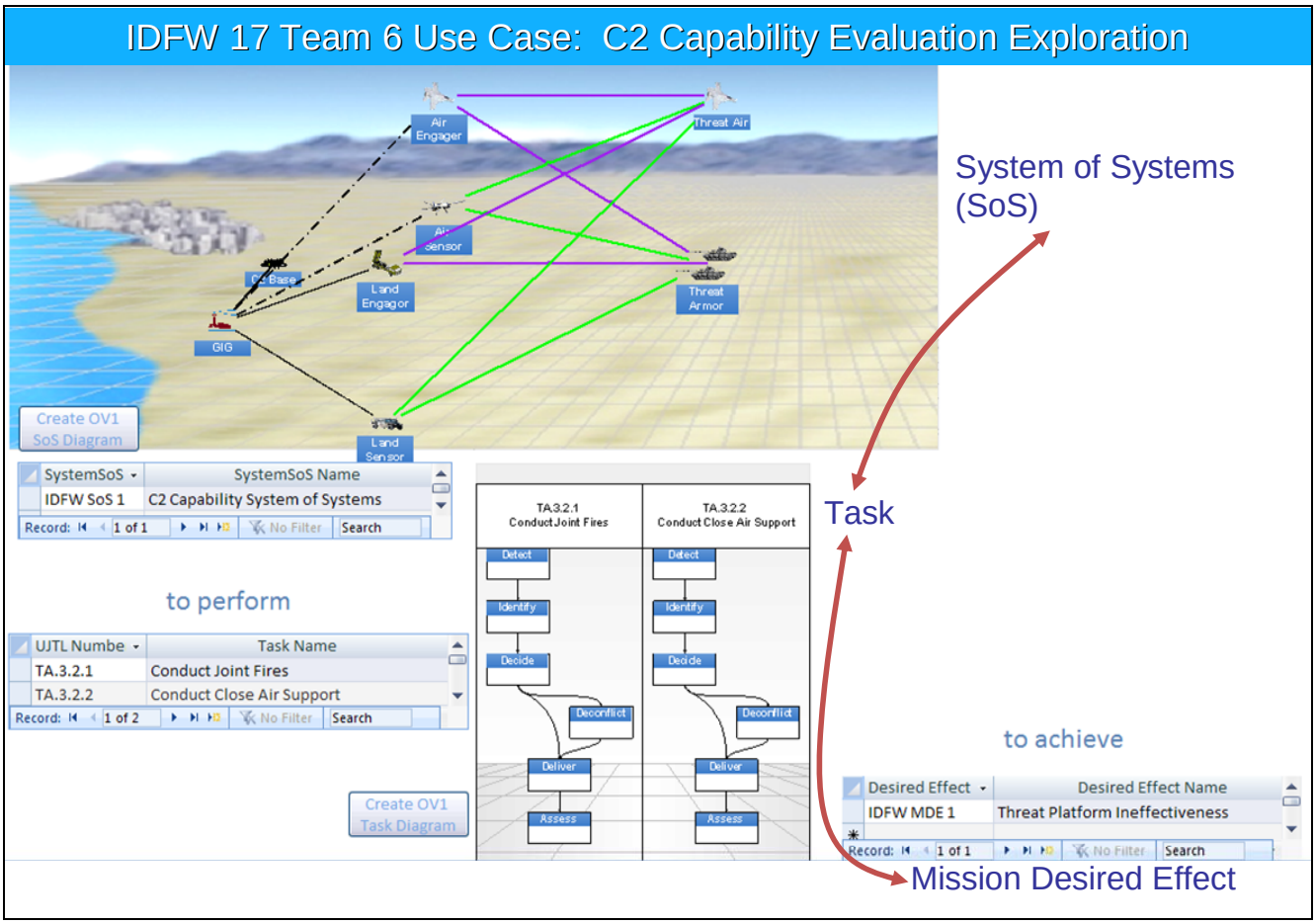



**Figure C-1-3. Use Case High Level Operational Graphic (OV-1)**

Taking the use case critical joint issue and JOC-T, an initial T&E design space was created involving DOE dependent variable measures and independent variable factors. The use case measures were based on the CTM's measures framework shown on the left side of figure C-1-4. The measures framework has three levels of measures: Mission Measure of Effectiveness (Mission MOE), Task Measure of Performance (Task MOP), and Measure of System/System of Systems Attribute (MOSA). These measures relate to the definition of capability in the following ways. Mission MOEs measure the ability to achieve mission desired effects. Task MOPs measure task performance. MOSAs measure system-level and SoS-level performance attributes across the materiel and non-materiel aspects of an SoS implementation. MOSAs can involve system-level effectiveness and suitability measures.

Multiple dependent variable measures were developed at each level as shown on the right side of figure C-1-4. The Mission MOEs focused on threat system combat ineffectiveness by counting threat kills and calculating cumulative ineffectiveness time of threat systems in the joint operations area (JOA). The Task MOPs calculate various times of C2 activities within joint close air support (JCAS) task and JFIRES task kill chains. The MOSAs were focused on non-materiel C2 personnel attribute performance relating to call for fire (CFF) deconfliction and decision procedures.

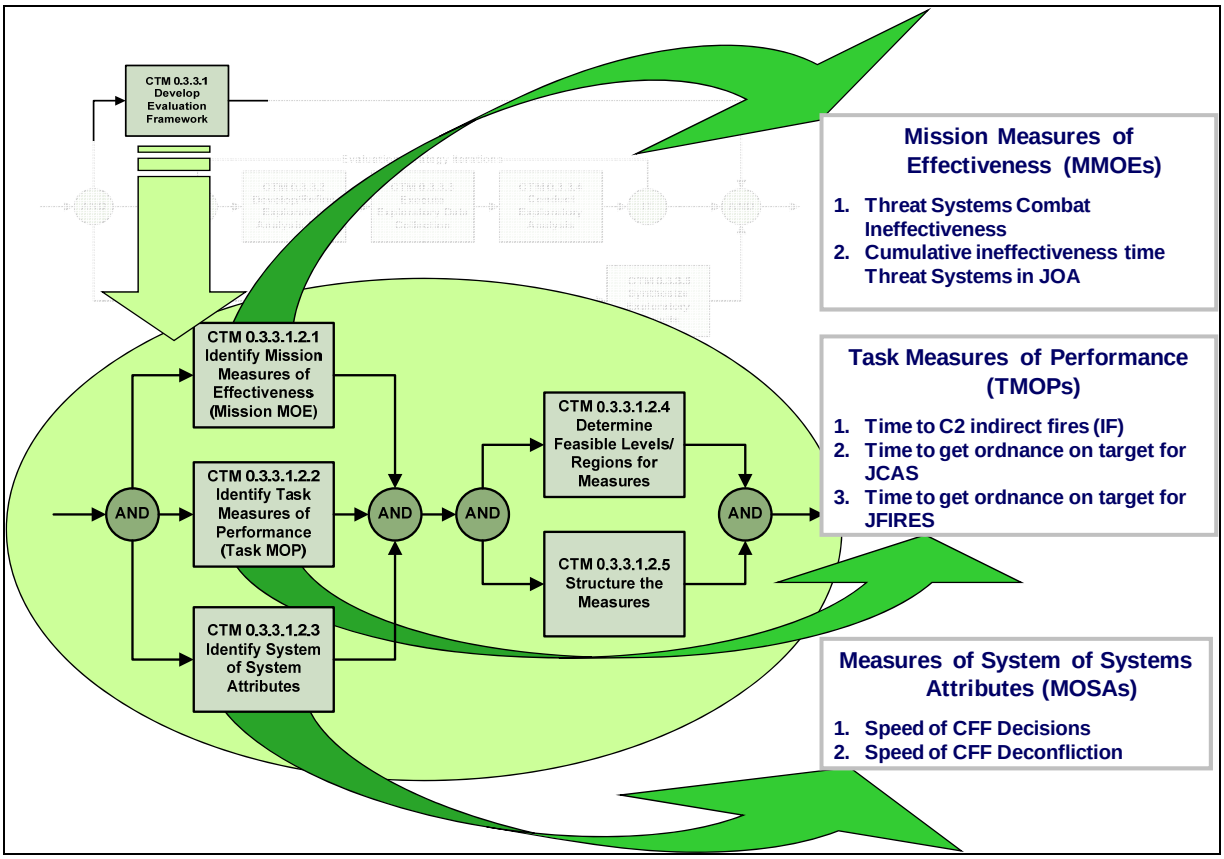



**Figure C-1-4. CTM Dependent Variable Response Levels and Use Case Measures**

## ABS Model and Scenario

Legacy ABSs that have been used in previous IDFW workshops did not simulate C2 processes very effectively nor include non-materiel factors in the evaluation strategy. To include these characteristics in an ABS model required the development of one by the JTEM Operations Research team. The result was an Excel-based model that simulated friendly and threat units in a JOA that resulted in direct and indirect fires. Direct fires occurred by ground units that were capable of moving, sensing, and shooting, and that did not require a CFF process to clear an airspace with a joint force. Indirect fires occurred when ground units with no firing capability or air units with firing capability sensed a target. Indirect fires required a CFF process to clear an airspace.

The ABS model was able to simulate the C2 steps associated with the CFF process (See figure C-1-5).

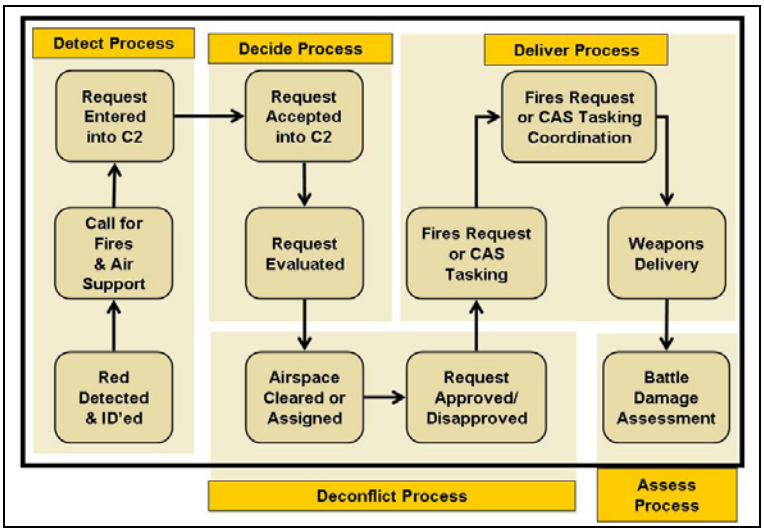



**Figure C-1-5. Call For Fires (CFF) Process**

Macal and North (2006) state a fundamental feature of an agent is the ability to make independent decisions; be active rather than purely passive. The friendly and threat units in this ABS model actively engaged each other following a series of procedural rules and variation of seventeen different SoS and conditional factors that include materiel attributes and non-materiel doctrinal procedures. Figure C-1-6 displays the capability crosswalk matrix with categorical and continuous factor levels used in the model. To simplify the experimental design, each factor was assigned two levels or values in which to evaluate their impact. Some level of complexity did exist with dependencies among multiple factors. For example, the value of factor 12 (Restricted Operating Zone [ROZ] slack time) had no impact if factor 11 (ROZ expiration) was set at "no". In this case, the solution was to treat the two factors as one factor with three levels:

- ROZ expiration with slack time of one
- ROZ expiration with slack time of two
- No ROZ expiration

One other dependency existed with the presence of a civilian zone, factors 15–17. This three-factor dependency needed to be treated as one combined factor with five levels. The implications of these multi-level factors will become apparent in the follow-on discussion of the experimental designs used in the workshop.

| Crosswalk Dimension | Crosswalk Sub-Dimension | Factor | Levels | Factor Type |
|---|---|---|---|---|
| System of Systems | Materiel | 1. Global Information Grid (GIG)? | Yes/No | Categorical |
| | | 2. Blue Speed | 1/2 (Multiplier) | Continuous |
| | | 3. Blue Monitor | 1/2 (Multiplier) | Continuous |
| | | 4. Blue Fires | 1/2 (Multiplier) | Continuous |
| | Non-materiel Doctrine | 5. Multiple Trackers? | Yes/No | Categorical |
| | | 6. Expedited Call For Fire? | Yes/No | Categorical |
| | | 7. Call For Fire Type of Decision | Available/Closest | Categorical |
| | | 8. Expedite Move | Yes/No | Categorical |
| | | 9. ROZ Type | Restrictive/Permissive | Categorical |
| | | 10. ROZ Size | 1/2 (Multiplier) | Continuous |
| | | 11. ROZ Expiration? | Yes/No | Categorical |
| | | 12. ROZ Slack Time | 1/2 time increments | Continuous |
| | | 13. Multi-Service Wait time | 1/2 time increments | Continuous |
| Condition | Environmental | 14. Adverse Weather? | Yes/No | Categorical |
| | | 15. Civilian Zone? | Yes/No | Categorical |
| | | 16. Civilian Zone Size | 20 / 40 | Continuous |
| | | 17. Civilian Zone Location | Location 1 or 2 | Categorical |

ROZ: Restricted Operating Zone

**Figure C-1-6. Factor Capability Crosswalk**

The scenario consisted of a joint forcible entry mission where threat forces hindered the ability to expand lodgment and control key infrastructure in order to facilitate rapid force build-up in the JOA. One desired effect to mission effectiveness included destroyed or neutralized threat forces. The desired joint capability was an ability to efficiently command and control indirect fires of Joint Close Air Support (JCAS) and Joint Fires (JFIRES) tasks. This was considered to be the SoS under test.

The seventeen modeled factors were the independent variables to the model. Dependent variables consisted of multiple measures that follow the JTEM evaluation strategy to evaluate joint mission effectiveness, task performance, and key SoS attributes. Each level of measures was important to the evaluation of the joint capability when integrated into a joint mission environment. Figure C-1-4 identifies the three levels of measures and the specific measures that were evaluated within the ABS model.

**First Iteration: C2 Issue Resolution III "Quick Look" Exploration**

Our first iteration was to do a "quick look" analysis of the capability space using a resolution III fractional factorial design at two levels, with an additional five-level DOE crossing factor (Civilian Zone). While resolution III designs can explore a large number of factors with a relatively small number of runs, they confound main effects with two-factor interactions (Box, Hunter, & Hunter, 1978). Therefore, we could not estimate main effects separately from two-factor interactions, but could get an indication of factors which had main effect or two-way factor significance.

The first iteration resolution III DOE was:

$$2_{III}^{15-11}\,(16\text{ runs})\text{ x Civ.Zone }(5\text{ levels})$$
$$=80\text{ trials }\ (20\text{ reps }\ \text{each})$$

which used a 15 two-level factor design crossed with Civilian Zone as a five-level factor.

The threat systems combat ineffectiveness Mission MOE was initially analyzed using a full main effects fit as shown in Figure C-1-7. The left hand column shows Xs indicating subject matter expert (SME) estimates of which factors have important effects concerning the threat systems combat ineffectiveness Mission MOE. Blue Speed (two speed levels), ROZ Type (Restricted or Permissive level), GIG (On or Off), and Blue Fires (two range levels) were estimated as the four most important factors. In the first iteration resolution III DOE runs, the four factors estimated as important were all statistically significant with p-values (Prob> [t]) of less than .05. Factors which did not provide indications of significance are candidates to reduce the DOE factor space in the next iteration of analysis. Based on our first iteration "quick look", we reduce the Civilian Zone factor levels from five to three, since varying Civilian Zone location did not have a significant effect. We also gained insights about which factors in the design space had indications of main or two-way significance. A surprise was that the No Adverse Weather factor level indicated a negative, main/two-way effect on the Mission MOE, which is non-intuitive and required further investigation in subsequent iterations.

**Full Main Effects Fit:**

Sorted Parameter Estimates

| SME Estimate | Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|---|
| X | Blue Speed | 29.0625 | 1.409628 | 20.62 | <.0001* |
| O | ROZ Size | -26.1625 | 1.409628 | -18.56 | <.0001* |
| O | CFF Type of Decision[A] | -12.39375 | 0.704814 | -17.58 | <.0001* |
| X | ROZ Type[P] | 10.71875 | 0.704814 | 15.21 | <.0001* |
| O | Multiple Trackers?[No] | -5.81875 | 0.704814 | -8.26 | <.0001* |
| O | ROZ Slack Time | 9.3875 | 1.409628 | 6.66 | <.0001* |
| X | GIG?[Off] | -3.81875 | 0.704814 | -5.42 | <.0001* |
| O | Civilian Zone?[3] | -6.58125 | 1.409628 | -4.67 | <.0001* |
| O | ROZ Expiration[No] | -2.79375 | 0.704814 | -3.96 | 0.0001* |
| O | Expedite Move[No] | -2.63125 | 0.704814 | -3.73 | 0.0003* |
| X | Blue Fires | 4.8375 | 1.409628 | 3.43 | 0.0008* |
| O | Blue Monitor | 4.4375 | 1.409628 | 3.15 | 0.0020* |
| O | Expedited CFF?[No] | -2.21875 | 0.704814 | -3.15 | 0.0020* |
| ?O | Adverse Weather?[No] | -2.11875 | 0.704814 | -3.01 | 0.0031* |
| O | Multi-service Wait time | -1.5875 | 1.409628 | -1.13 | 0.2620 |
| O | Civilian Zone?[4] | -1.175 | 1.409628 | -0.83 | 0.4059 |
| O | Civilian Zone?[1] | 0.41875 | 1.409628 | 0.30 | 0.7669 |
| O | Civilian Zone?[2] | 0.0125 | 1.409628 | 0.01 | 0.9929 |

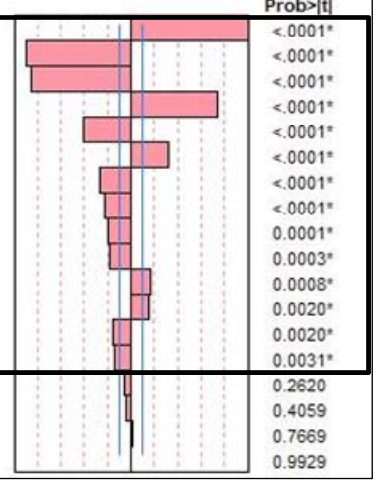


**Figure C-1-7. First Iteration "Quick Look" Resolution III Full Main Effects Fit**

**<u>First Iteration: Exploring Multiple Measures</u>**

As a secondary objective in the first analysis iteration, we wanted to investigate techniques for analyzing results across multiple measures. One method is to review results for each measure and then construct a relationship table to evaluate which factors impact which measures. Figure C-1-8 displays the results for the main effects on each of the mission measures of effectiveness, task measures of performance, and measures of SoS attributes as shown in the yellow squares. Squares where there are zeros indicates “no significant impact” based on the factor response. Squares with positive or negative one values indicate positive or negative “significant impacts” on the measure. Two way interactions across measures are seen in the right-hand columns. Those squares with a green background represent a direct relationship across the two relevant measures while those with a red background represent an indirect relationship. Those squares with a zero do not appear to have a relationship. This provides us insights on factors that are significant across multiple measures. In our results, we observe that every factor has the same

impact on both of the Mission MOEs. We also see the factors Blue Monitor, Multiple Trackers, and ROZ Slack Time each have a positive impact on every measure. A third observation is that the factors CFF Type of Decision and ROZ Expiration influence the Task MOPs and MOSAs in a different direction from the Mission MOEs. We may draw a conclusion that these five factors may be significantly important for follow-on analysis.

Green: Direct relationship
Red: Indirect relationship
No color: No relationship

MMOE: Mission Measure of Effectiveness
TMOP: Task Measure of Performance
MOSA: Measure of System of Systems attributes

| Factors | Values | MMOE1 | MMOE2 | TMOP1 | TMOP2 | TMOP3 | MOSA1 | MOSA2 | M1T1 | M1T2 | M1T3 | M2T1 | M2T2 | M2T3 | M1S1 | M1S2 | M2S1 | M2S2 | T1S1 | T1S2 | T2S1 | T2S2 | T3S1 | T3S2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GIG | Off or On | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| Blue Speed | 1 or 2 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Blue Monitor | 1 or 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Blue Fires | 1 or 2 | 1 | 1 | 0 | -1 | 0 | 1 | -1 | 0 | -1 | 0 | 0 | -1 | 0 | 1 | -1 | 1 | -1 | 0 | 0 | -1 | 1 | 0 | 0 |
| Civilian Zone | Off or On | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| Civilian Zone Size | Small or Large | -1 | -1 | -1 | 0 | 0 | -1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Multiple Trackers | No or Yes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Adverse Weather | No or Yes | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Expedited CFF | No or Yes | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CFF Type of Decision | Closest or Available | -1 | -1 | 1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Expedite Move | No or Yes | 1 | 1 | 1 | -1 | 0 | 1 | 0 | 1 | -1 | 0 | 1 | -1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | -1 | 0 | 0 | 0 |
| ROZ Type | Restrictive or Permissive | 1 | 1 | 1 | 1 | 1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 |
| ROZ Size | 1 or 2 | -1 | -1 | -1 | 0 | 0 | -1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | -1 | 1 | -1 | 1 | -1 | 0 | 0 | 0 | 0 |
| Multi-service Wait time | 1 or 2 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| ROZ Expiration | No or Yes | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ROZ Slack Time | 1 or 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Civilian Zone Location | 1 or 2 | 0 | 0 | -1 | 0 | -1 | -1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

**Figure C-1-8. Measures Relationship**

## Second Iteration: C2 Issue Resolution V Main Effect and Two Way Interaction Exploration

As a second analysis iteration, we used a resolution V fractional factorial design at two levels. When simulation resources and time allow the increased number of runs, resolution V designs are preferred over resolution III designs, since the higher resolution designs do not confound main effects and two-factor interactions (Box, Hunter, & Hunter). Resolution V designs allowed us to separately assess main effect and two way interactions using a stepwise regression model of response measures.

The second iteration resolution V DOE was:

$$2_V^{12-4}\ (256\ \text{runs})\ \text{x Civ.Zone}\ (3\ \text{level})\ \text{x ROZ Expiration}\ (3\ \text{level})$$
$$= 2304\ \text{trials}\ (3\ \text{reps each})$$

which used 12 two-level factors crossed with Civilian Zone and ROZ Expiration as 3-level factors.

The threat systems combat ineffectiveness Mission MOE was again used for second iteration analysis for factor significance. Classification and Regression Tree (CART) partitioning and stepwise regression models were used to identify factor main and two-way effects. The second

iteration CART is shown in Figure C-1-9 with the partitioning of factor levels producing "good" Mission MOE responses (larger threat attrition counts) on the left side and "bad" Mission MOE responses (smaller threat attrition counts) on the left side. The CART shows five levels of partitions to indicate the most significant factor partitions. Good factor levels of high Blue Speed, small ROZ size (1), and permissive ROZ type produce a Mission MOE mean attrition of ~42. Good factor levels of high Blue Speed (2), small ROZ size (1), and permissive ROZ type produced a Mission MOE mean attrition of ~42. At the other end of this Mission MOE effect spectrum, bad factor levels of low Blue Speed (1) and large ROZ size (2) produced a Mission MOE mean attrition of ~21.

The second iteration analysis also included a stepwise regression fit incorporating main and two-factor interactions as shown in Figure C-1-10. The left hand column shows the mapping of SME factor importance estimates to the significant (p-value < .05) regression fits for the threat systems combat ineffectiveness Mission MOE. One surprise was that the Blue Fires was significant as a two-factor interaction with ROZ type, not as a main effect. Another surprise was the number of significant two-factor interactions with ROZ Type, indicating second order effects between this non-materiel doctrine factor and other factors. Based on our second iteration analysis, we identified important decision factor, supporting decision and condition factor recommendations to further reduce the design space. Our approach was to combine key supporting decision and condition factor levels into a "meta" stressing factor which varied factor levels in the same "good" and "bad" Mission MOE directions. The factors selected for this stressing factor are further discussed in the third iteration analysis section.

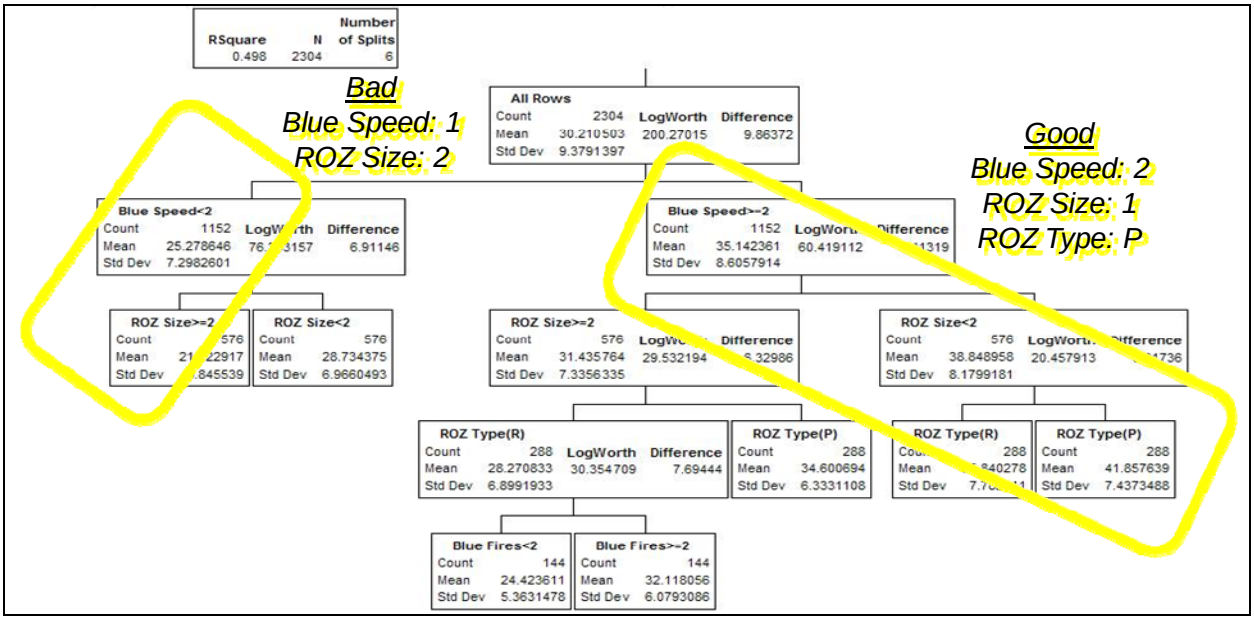



**Figure C-1-9. Second Iteration Classification and Regression Tree (CART)**

**Stepwise Regression Fit:**

Sorted Parameter Estimates

| SME Estimate | Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|---|
| X | Blue Speed | 9.8637153 | 0.202018 | 48.83 | 0.0000* |
| O | ROZ Size | -7.162326 | 0.202018 | -35.45 | <.0001* |
| X | ROZ Type[P] | 2.4657118 | 0.101009 | 24.41 | <.0001* |
| ?x | (Blue Fires-1.5)*ROZ Type[P] | -4.063368 | 0.202018 | -20.11 | <.0001* |
| X | GIG?[Off] | -1.751302 | 0.101009 | -17.34 | <.0001* |
| O | Civilian Zone?{Yes40-Yes20&No} | -1.855143 | 0.107136 | -17.32 | <.0001* |
| O | CFF Type of Decision[A] | -1.508247 | 0.101009 | -14.93 | <.0001* |
| O | Multiple Trackers?[No] | -1.313802 | 0.101009 | -13.01 | <.0001* |
| O | Adverse Weather?[No]*ROZ Type[P] | 1.2773438 | 0.101009 | 12.65 | <.0001* |
| x | (Blue Monitor-1.5)*ROZ Type[P] | 1.7161458 | 0.202018 | 8.49 | <.0001* |
| x | (Blue Speed-1.5)*ROZ Type[P] | 1.2421875 | 0.202018 | 6.15 | <.0001* |
| x | GIG?[Off]*ROZ Expiration{No-Yes1} | -0.6875 | 0.123711 | -5.56 | <.0001* |
| O | (Civilian Zone?{Yes40-Yes20&No}+0.33333)*(ROZ Size-1.5) | 1.1803385 | 0.214273 | 5.51 | <.0001* |
| x | (Civilian Zone?{Yes40-Yes20&No}+0.33333)*ROZ Type[P] | -0.57194 | 0.107136 | -5.34 | <.0001* |
| x | CFF Type of Decision[A]*ROZ Type[P] | 0.531684 | 0.101009 | 5.26 | <.0001* |
| O | Adverse Weather?[No]*CFF Type of Decision[A] | 0.5099826 | 0.101009 | 5.05 | <.0001* |

Surprise!

Surprise!

**Figure C-1-10. Second Iteration Resolution V Stepwise Regression Fit**

**<u>Third Iteration: C2 Issue Refined Design with Stressing Factor Exploration</u>**

In the previous steps, a set of lower priority factors were determined and set at levels that resulted in high and low stress levels (high and low values in the measure). As a third analysis iteration, we again used a resolution V fractional factorial design at two levels with nine of the original factors and a tenth factor that consisted of the remaining five factors set at most stressing and least stressing levels. This tenth factor is considered a "conditional" factor, each single factor represents conditions in the physical environment or conditions in the doctrinal environment.

The third iteration resolution V DOE was:

$$2_V^{10-3} \text{ (128 trials)} \quad \text{(3 reps each)}$$

which used nine two-level factors and a tenth most/least stressing factor.

The team applied CART analysis output with the stressing factors set at their stressing levels. The result was the stressing factor becoming the most statistically significant factor. The best response to the dependent variable was achieved when the stressing factor was set at its least level combination along with Blue speed, CFF Type of Decision, and ROZ type set at desired levels. The results can be seen in Figure C-1-11 with a least squares regression indicating those most significant factors. This implies the original 17 factors can now be reduced to a set of three materiel factors, three non-materiel factors, and one stressing factor, each with two levels. These results could then be presented to SMEs as recommendations for prioritizing factors that may be explored in future live, virtual, and constructive test and evaluation.

$2_V^{10-3}$ (128 trials) (3 runs ea.) - Factor 10 is 5 Factor Most/Least Stressing

Test Event Factor Priority Recommendations:
5 Most Stressing Levels
1 (Decision Factor)
2 (Decision Factor)
3 (Decision Factor)
4 (Decision Factor)
5 (Decision Factor)
6 (Decision Factor)

**Sorted Parameter Estimates**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| (5 Factor Least Stressing Levels) | 17.59375 | 1.122227 | 15.68 | <.0001* |
| Blue Speed | 30.8125 | 2.244455 | 13.73 | <.0001* |
| (Blue Fires-1.5)*ROZ Type[P] | -15.8125 | 2.244455 | -7.05 | <.0001* |
| ROZ Type[P] | 6.375 | 1.122227 | 5.68 | <.0001* |
| CFF Type of Decision[A] | -6.296875 | 1.122227 | -5.61 | <.0001* |
| GIG?[Off] | -5.90625 | 1.122227 | -5.26 | <.0001* |
| Multiple Trackers?[No] | -5.59375 | 1.122227 | -4.98 | <.0001* |

| Dimension | Sub Dimension | Priority/Factor | Levels | Factor Type |
|---|---|---|---|---|
| System of Systems | Materiel | 1. Blue Speed | 1/2 (Multiplier) | Continuous |
| | | 2. Blue Fires | 1/2 (Multiplier) | Continuous |
| | | 5. Global Information Grid (GIG)? | Yes/No | Categorical |
| | Non-materiel: Doctrine | 3. Restricted Op. Zone Type | Restrictive/Permissive | Categorical |
| | | 4. Call For Fire type Decision | A/C | Categorical |
| | | 6. Multiple Trackers? | Yes/No | Categorical |
| | | Most Stress 1. ROZ Size | 1 (Multiplier) | Continuous |
| | | Most Stress 2. ROZ Expiration | No | Categorical |
| | | Most Stress 3. Multi-Service Wait time | 2 (Additional Time Increment) | Continuous |
| Condition | Environmental | Most Stress 4. Adverse Weather? | Yes | Categorical |
| | | Most Stress 5. Civilian Zone? | Yes (Size = 40) | Categorical |

**Figure C-1-11: Stressing Level Results**

## Discoveries and Surprises

Our main discoveries during IDFW 17 related to key refinements of the CTM's 0.3 Develop Evaluation Strategy focus process. These refinements included promising automation for developing evaluation strategy measures and factors from critical operational issue and capability crosswalk constructs. The generation of resolution V designs for our hybrid sets of continuous and categorical factors proved very useful. These resolution V designs were generated using Sanchez & Sanchez (2005) fractional factorial composite designs.

Our team discovered surprises related to our capability exploration and our team dynamics in IDFW 17. While SME factor estimates proved to be significant, we were surprised at the number and composition of two-factor interactions. Some of the factors in the significant two factor interactions did not have corresponding main effects, which would increase the need to explore level interactions in further test excursions, such as live, virtual, and constructive test events.

## Conclusions

IDFW enabled us to derive conclusions concerning the relevance of data farming techniques to the CTM 0.3 Develop Evaluation Strategy process. Resolution V and crossing design are examples of preferred DOE techniques needed for assessing main effect and two-way interactions for hybrid capability sets of categorical and continuous factors. Initial low resolution DOE and data farming may provide first insights into significant measures. These insights may differ across multiple measures and require retaining uncertain factors in the subsequent designs. Iterative farming can provide additional prioritization of factors. Factors with more than two discrete levels require additional farming to assess their impact. Multiple

measures (dependent variables) add significant complexity to determining factors with highest impact. This requires evaluation of factors across multiple dependent variables (measures) by constructing relationship tables that provide insights on measure impacts.

Additional conclusions can be made related to effectiveness and suitability of the CTM. Exploratory analysis requires an iterative process that is consistent and repeatable for prioritizing factors. In this workshop, a set of factors were set at most and least stressing levels to compose one stressing factor. This allowed for a smaller set of factors that result in less requirements and resources for future testing. Alternative steps may be available for prioritizing factors. Automated tools for DOE and modeling can help to simplify the exploratory analysis process. Non-materiel factors can be equally important to testing an SoS.

**Issues for Further Investigation**

We anticipate further exploration and refinement of the CTM evaluation thread processes during future IDFW events. This includes the refinement of capability assessment structures and efficient DOE techniques, such as composite fractional factorials are key focus areas for further exploration and automation. Additional excursions may be made on the process for prioritizing test factors which can lead to a consistent and repeatable methodology for testing SoS in a joint mission environment.

**References**

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for Experimenters. John Wiley & Sons, Inc.: New York.

Chairman of the Joint Chiefs of Staff (2007). CJCSM 3170.01C, Operation of the Joint Capabilities Integration and Development System, Department of Defense: Washington D.C.

Macal, Charles M. and North, Michael J. 2006. Tutorial on Agent-Based Modeling and Simulation Part 2: How to Model with Agents, Proceedings of the 2006 Winter Simulation Conference, IEEE 1-4244-0501-7/06.

Sanchez, S. M. & Sanchez, P. J. (2005). Very Large Fractional Factorial and Central Composite Designs. ACM Transactions on Modeling and Computer Simulation, 15 (4), 362–377.

*This page intentionally left blank.*

# ANNEX D - DESIGN OF EXPERIMENTS (DOE)

This annex was developed for the *Analyst's Handbook for Testing in a Joint Environment* (Analyst's Handbook), but it is designed to stand on its own; therefore, it contains its own reference section.

A concise overview of the effective use of simulation experiments and efficient designs to support testing in joint environments is provided. It is assumed the reader has a basic knowledge of statistics and experimental design. A brief summary of basic ideas and notation is provided.

Annotated bibliographies are presented for those seeking additional detail.

**Contents**


D.1 Introduction ..... D-3

D.2 Choice of Design ..... D-3
- D.2.1 Basic Concepts ..... D-3
  - D.2.1.1 Definitions and Notation ..... D-3
  - D.2.1.2 Pitfalls to Avoid ..... D-6
  - D.2.1.3 Choosing Factors ..... D-7
  - D.2.1.4 Sample-Size Issues ..... D-8
  - D.2.1.5 Annotated Bibliography ..... D-8
- D.2.2 Potential Single-Stage Designs: Descriptions and Assumptions ..... D-8
  - D.2.2.1 $2^k$ Factorial Designs (Coarse Grids) ..... D-9
  - D.2.2.2 $m^k$ Factorial Designs (Finer Grids) ..... D-11
  - D.2.2.3 $2^{k-p}$ Resolution 5 Fractional Factorial Designs ..... D-12
  - D.2.2.4 Central Composite Designs ..... D-14
  - D.2.2.5 Nearly Orthogonal Latin Hypercube Designs ..... D-15
  - D.2.2.6 Robust Design Methods. ..... D-17
  - D.2.2.7 Annotated Bibliography ..... D-18
- D.2.3 Potential Sequential Methods: Descriptions and Assumptions ..... D-19
  - D.2.3.1 Response-Surface Methodology ..... D-19
  - D.2.3.2 Other Simulation Optimization Techniques ..... D-19
  - D.2.3.3 Modified Response-Surface Methodology ..... D-20
  - D.2.3.4 Group Screening ..... D-22
  - D.2.3.5 Annotated Bibliography ..... D-23
- D.2.4 Design Comparisons ..... D-23

D.3 Special Opportunities with Computer-Simulation Experiments ..... D-25
- D.3.1 Randomization ..... D-26
- D.3.2 Variance Reduction ..... D-26
- D.3.3 Effective Use of Computing Resources ..... D-27
- D.3.4 Annotated Bibliography ..... D-28

D.4 Examples: Designs, Analyses, and Insight ..... D-28

D.4.1 Potential Analytic and Graphical Methods .....D-29
D.4.2 Sample Designs.....D-31
D.4.2.1 Large 2k Fractional Factorial or CCD Experiments .....D-31
D.4.2.2 NOLH Design .....D-32
D.4.2.3 Robust Design.....D-33
D.4.2.4 Screening Design .....D-33
D.4.3 Gaining Insight .....D-34
D.4.4 Annotated Bibliography.....D-35

D.5 Conclusions .....D-35

References .....D-36

**List of Figures**

D-1. Confounded Factor Effects for Seizing the Objective.....D-7
D-2. One-at-a-Time Sampling for Seizing the Objective .....D-7
D-3. $2^2$ and $2^3$ Factorial Designs.....D-9
D-4. $2^2$ and $11^2$ Factorial Experiments for Seize-the-Objective.....D-11
D-5. Scatter Plot Matrix for a $5^4$ Factorial Design.....D-12
D-6. $2^{3-1}$ Fractional Factorial .....D-13
D-7. Construction of Central Composite Designs .....D-15
D-8. Random Latin Hypercube for Seize-the-Objective .....D-16
D-9. Scatter Plot Matrices for Selected Factorial and NOLH Designs .....D-17
D-10. Design Comparison Chart .....D-24
D-11. Flowchart of Initial Design Selection Process .....D-25

**List of Tables**

D-1. Simple Experimental Design in Natural and Coded Levels.....D-6
D-2. Terms for a $2^3$ Factorial Design.....D-10
D-3. Data Requirements for Factorial Designs .....D-12
D-4. Terms for a $2^{7-4}$ Fractional Factorial Design .....D-13
D-5. Data Requirements for 3-Level Designs.....D-15
D-6. Data Requirements for Nearly Orthogonal Latin Hypercube Designs.....D-16
D-7. Multiple Response Optimization Process through Use of CARTs .....D-21

## D.1 Introduction

This annex discusses the tools and techniques for design of experiment (DOE) that support testing of complex system of systems (SoS) for a joint capability. The primary focus is on the exploratory analysis process, where the goal is to identify a smaller, more manageable, set of key factor combinations from which test trial sets are chosen; this is accomplished primarily through experimental design in conjunction with modeling and simulation (M&S). The primary benefits include:

- An ability to pre-test many scenarios via simulation to reduce the number of live tests needed
- An ability to pre-screen a very large number of factors (i.e., SoS attributes) in the simulation to assist in identifying the key drivers of joint mission effectiveness (JMe).

It is assumed the reader has a basic knowledge of statistics and experimental design, a brief summary of basic ideas and notation is provided in section D.2.1. Brief descriptions of various types of single-stage and sequential designs are provided in sections D.2.2 and D.2.3, respectively, and a table comparing the characteristics and capabilities of these designs in section D.2.4. In section D.3, several special opportunities are presented that are available for simulation experiments that are generally not available in live tests. Briefly described are the different considerations regarding randomization (section D.3.1), the benefits of variance reduction techniques (section D.3.2), implementation issues that relate to using computing resources effectively (section D.3.3), and opportunities to use DOE as part of the debugging and simulation-development process (section D.3.4). In section D.4, descriptions of several analytic and graphical tools and techniques are provided; these are useful for gaining insight in the exploratory analyses. Annotated bibliographies are provided for those needing additional detail.

## D.2 Choice of Design

This section contains a brief summary of basic ideas in classical DOE, as well as an overview of useful types of designs.

### D.2.1 Basic Concepts

#### D.2.1.1 Definitions and Notation

One of the first things an experimenter or tester must do to design a good experiment is identify the experimental factors. In DOE, factors are the input (or independent) variables that may have some impact on experimental outputs (called responses in DOE parlance). In general, an experiment may have many factors, each of which might assume a variety of values, called levels of the factor in DOE. A primary goal of many DOEs is to identify which of the factors are really important for which responses, and which are not and can thus be dropped from further consideration. This greatly reduces the experimental effort and simplifies the task of interpreting the results. We would like to identify the nature of the impact on the responses (e.g., increasing, linear, quadratic) for the important factors; also whether the levels of some factors influence the effects that other factors have (called factor interactions).

To identify good or appropriate designs, it is often useful to classify the factors along several dimensions:

- Quantitative or qualitative. Quantitative factors naturally take on numerical values, while qualitative factors do not (though they might be assigned artificial values, such as 1 if a particular weapons package is present and 0 if it is not).
- Discrete or continuous (quantitative factors only). Discrete factors can have levels only at certain separated values; an example would be the number of supply ships deployed in a region, which would have to be a non-negative integer, presumably with some upper bound. Continuous factors can assume any real value, perhaps within some range, such as the speed at which a vehicle is operated.
- Controllable or uncontrollable. In a simulation or live test all factors are manipulated and controlled, but in reality factors might be controllable or not. For example, the degree or nature of enemy jamming of a communications system would be controlled in a test, but not in an actual fight. This can affect how the experimenter interprets the estimates of the effects of factors.
- Binary or not. When developing and analyzing a DOE, it can sometimes matter whether or not a factor is binary. Binary factors are naturally constrained to just two levels, like presence of a particular weapons package or not. Non-binary factors could take on more than two values, but might still be tested at only two levels, typically “low" and “high," or might be allowed to assume more than two levels in the experiment. Both binary and non-binary factors could be quantitative or qualitative, as well as controllable or uncontrollable. The more levels factors can take on in an experiment, the more work (computational or real) is required, but the more information can be gleaned. A basic goal of DOE is to figure out the specific factor-level combinations necessary to get the information desired (importance of factors and the nature of their effects) in the most efficient way, i.e., with no more experimental runs than absolutely necessary.

Throughout this annex, simulation model denotes any model that is evaluated using a computer. Simulation models come in many flavors. There are deterministic simulations:

- Numerical solutions of differential equations - the same set of inputs always produces the same output.
- Stochastic simulations - the same set of simulation inputs may produce different output unless the random-number streams are carefully controlled.

Simulations that model a process that occurs over time can also be characterized as terminating or non-terminating, depending on the stopping conditions. In this annex it is assumed terminating simulations are used, because vignettes for joint operations and tests naturally fall in the terminating simulation category. The simulation stops after either a pre-specified amount of simulation time has elapsed or when a specific event or condition occurs. The run is the appropriate experimental unit for terminating simulations, but the output measure of effectiveness (MOE) for a single run may be an aggregate (average, extreme, or some other summary output measure) computed from detailed simulation results.

Mathematically, let $X_1; : : : ;X_k$ denote the k factors in an experiment, and let Y denote a response of interest. Sometimes graphical methods are the best way to gain insight about Y, but often there is interest in constructing response surface metamodels that approximate the relationships between the factors and the responses with statistical models (typically regression models).

First, suppose that the $X_i$'s are all quantitative, although they can be discrete or continuous. A main-effects model means this assumption:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon \tag{1}$$

where the ε's are independent random errors with mean zero. Ordinary least-squares regression assumes that the ε's in (1) are also identically distributed, but the regression coefficients are still unbiased estimators of the $\beta_i$ even if the underlying variance is not constant.

To explore any quadratic effects, terms like $X_1^2$ are used as potential explanatory variables for Y. Similarly, two-way interactions are terms like $X_1X_2$. A second-order model includes both quadratic effects and two-way interactions; although it is best for numerical stability to fit this equation after centering the quadratic and interaction terms, as in (2):

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{i,i} \left( X_i - \overline{X}_i \right)^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{i,j} \left( X_i - \overline{X}_i \right) \left( X_j - \overline{X}_j \right) + \varepsilon \tag{2}$$

Some statistical packages do this centering automatically.

It is worth noting that regression can also be used when some of the X's are qualitative, in fact, the analysis of variance (ANOVA) technique commonly used for experimental designs with qualitative X's is a special case of regression. Equations (1) and (2) should be modified, but this can be done by using binary *indicator variables* (also called *dummy variables*). If there is a qualitative factor X with *m* distinct levels $c_1; : : : ; c_m$, then X is replaced by m - 1 indicator variables $I_1; : : : ; I_{m-1}$. Here, $I_i$ is set to 1 if $X = c_i$ and 0 otherwise (i = 1, : : : , m-1); $X = c_m$ is a baseline, and its instances correspond to $I_i = : : : = I_{m-1} = 0$. Estimation of the regression equation proceeds as usual, and any statistics package can be used. Once the equation has been fit, the value of the estimated β coefficient of indicator variable $I_i$ is interpreted as the effect on Y of the categorical variable X's being at level $c_i$ rather than at its baseline level $c_m$. Terms that involve interactions between qualitative factors, or between qualitative and quantitative factors, are permissible. Some statistical packages automatically handle this recoding, so the model-specification process is transparent to the user.

A *design* is a matrix where every column corresponds to a factor, and the entries within the column are settings for this factor. The T&E community may refer to these as *test design matrices*. Each row represents a particular combination of factor levels, and is called a *design point*. If the row entries correspond to the actual settings used, these are called *natural levels*. Coding the levels in a standardized way is a convenient way to characterize a design. Different codes are possible, but for quantitative data the low and high levels are often coded as -1 and +1,

respectively, for arithmetic convenience. Table D-1 shows a simple experiment, in both natural and coded levels, that could be conducted on a two-factor example.

**Table D-1. Simple Experimental Design in Natural and Coded Levels**

| | Natural Levels | | Coded Levels | |
|---|---|---|---|---|
| Design Point | $X_1$ | $X_2$ | $X_1$ | $X_2$ |
| 1 | 16 | 20 | -1 | -1 |
| 2 | 18 | 20 | +1 | -1 |
| 3 | 16 | 22 | -1 | 0 |
| 4 | 18 | 22 | +1 | 0 |
| 5 | 16 | 24 | -1 | +1 |
| 6 | 18 | 24 | +1 | +1 |

Each repetition of the whole design matrix is called a *replication* and it is generally assumed the replications are independent. Let $n_d$ be the number of design points, and $n_r$ be the number of replications. The total number of experimental units is $n_{tot} = n_d n_r$.

### D.2.1.2 Pitfalls to Avoid

Two common types of simulation studies are ill-designed experiments. The first can occur if several people each suggest an “interesting" combination of factor settings, so a handful of design points end up being explored where many levels change simultaneously. Consider an agent-based simulation model of a ground force attempting to seize an objective, while an opposing ground force is trying to do the same thing. Suppose that only two design points are used, corresponding to different settings for speed (X1) and stealth (X2), with the results in figure D-1. (Here, instead of providing numerical values for the MOE, a blue circle is used to represent a “good" average outcome, while a red square represents a “bad" average outcome.) One subject matter expert (SME) might claim these results show that high stealth is of primary importance, another that speed is the key to success, and a third that they are equally important. There is *no way* to resolve these differences of opinion without collecting more data. In statistical terms, the effects of stealth and speed are *confounded*. In practice, simulation models easily have dozens or hundreds of potential factors. A handful of haphazardly chosen scenarios, or a trial-and-error approach, can use up a great deal of time without addressing the fundamental questions.

| Speed | Stealth | Success? |
|---|---|---|
| Low | Low | No |
| High | High | Yes |

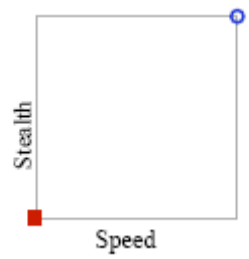



**Figure D-1. Confounded Factor Effects for Seizing the Objective**

The second type of study that can be problematic occurs when people start with a "baseline" scenario and vary one factor at a time. Revisiting the seize-the-objective example, suppose the baseline corresponds to low stealth and low speed. Varying each factor, in turn, to its high level yields the results of figure D-2. It appears that *neither* factor is important, so someone using the simulation results to decide how to train and equip would not know how (or if) to proceed.

| Speed | Stealth | Success? |
|---|---|---|
| Low | Low | No |
| High | Low | No |
| Low | High | No |

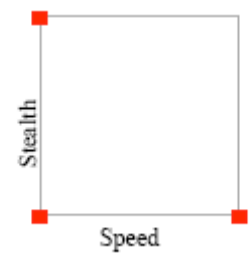



**Figure D-2. One-at-a-Time Sampling for Seizing the Objective**

However, if all four combinations of speed and stealth (low/low, low/high, high/low, and high/high) are sampled, it is clear that success requires both high speed and high stealth. This means that the factors interact and that one-at-a-time sampling will never uncover them.

The pitfalls of using a poor design seem obvious on this simple problem, but the same mistakes are made far too often in larger studies of more complex models. When only a few variations from a baseline are conducted, there may be many factors that change but a few that SMEs think are key. If they are mistaken, changes in performance from the baseline scenario may be attributed to the wrong factors. Similarly, many analysts change one factor at a time from their baseline scenario. In doing so, they fail to understand that this approach implicitly assumes that there are no interaction effects. This assumption may be unreasonable unless the region of exploration is very small.

### D.2.1.3 Choosing Factors

Potential factors in simulation experiments include the *input parameters* or *distributional parameters* of a simulation model. For example, a simulation model of a joint call for fire might have both quantitative factors (such as the number of weapons of different types, an unmanned aerial vehicle's [UAV] probability of detection or the mean time required to achieve airspace deconfliction) and qualitative factors (such as whether or not dynamic re-tasking is available, or the type of command and control structure).

Generating a list of the potential inputs to a simulation model is one way of creating an initial factor list. However, it is important to remember that factors need not correspond directly to simulation inputs. For example, suppose two inputs are the mean times $\mu_1$ and $\mu_2$ required for a specific agent to process messages from class 1 and class 2, respectively, where message class 1 is considered more complex than message class 2. Varying $\mu_1$ and $\mu_2$ independently may either result in unrealistic situations where $\mu_1 < \mu_2$ , or require the analyst to select very narrow factor ranges. Instead, one could use $\mu_1$ as one factor to represent the capabilities of the agent, and vary the ratio $\mu_1 / \mu_2$ over some range of interesting values (say, 0.4 to 0.9) to represent the relative difference in message complexity.

#### D.2.1.4 Sample-Size Issues

In test planning for live experiments, where data are extremely expensive, the total sample size is often very small. This affects the choice of an experimental design as well as the number of replications.

In simulation experiments, where a major portion of the effort often occurs in model development, the total sampling budget may not be so constrained. This increases the set of potential designs that can be used, and it may be possible to generate a great deal of information (even hundreds of thousands of runs) in a relatively short time. This is discussed further in section D.3.

#### D.2.1.5 Annotated Bibliography

Different types of simulation studies involve different types of *experimental* units. For a static Monte Carlo simulation, where no aspect of time is involved, the experimental unit is a single observation. For time-stepped or discrete-event stochastic simulation studies, it more often is a run or a batch, yielding an averaged or aggregated output value. When runs form the experimental units for non-terminating simulations, and steady-state performance measures are of interest, care must be taken to delete data during the simulation's warm-up period before performing the averaging or aggregation. Details may be found in any simulation textbook, such as Law (2007), Banks et al. (2005), or Zeigler et al. (2000).

### D.2.2 Potential Single-Stage Designs: Descriptions and Assumptions

Many designs are available in the literature. This annex focuses on a few basic types that are particularly useful for simulation experiments. Factorial or gridded designs are straightforward to construct and readily explainable, even to those without statistical backgrounds. Coarse grids ($2^k$ factorials) are most efficient if one can assume that the simulation response is well-fit by a model with only linear main effects and interactions, while fine grids (more than two levels for factors) provide greater detail about the response and greater flexibility for constructing metamodels of the responses. When the number of factors is large, then more efficient designs are required. Latin hypercubes are good general-purpose designs for exploring complex simulation models when little is known about the response surfaces. Designs called *resolution 5* $2^k$ fractional factorials (R5-FFs) allow the linear main effects and interactions of many factors to be investigated simultaneously; they are potential choices either when factors have only two qualitative settings, or when practical considerations dictate that only a few levels be used for quantitative input factors. Expanding these R5-FFs to central composite designs provides some information about nonlinear behavior in simulation response surfaces.

Factorials (or gridded designs) are perhaps the easiest to discuss. They examine all possible combinations of the factor levels for each of the $X_i$'s. A shorthand notation for the design is $m^k$, which means $k$ factors are investigated, at $m$ levels for each factor, in a total of $m^k$ design points. One can write designs where different sets of factors are investigated at different numbers of levels as, e.g., $m_1^{k1} \times m_2^{k2}$, where $k_1$ factors are evaluated at $m_1$ levels each, and another $k_2$ factors are evaluated at $m_2$ levels each. These are sometimes called *crossed* designs. For example, the design in Table D-1 is a $2^1 \times 3^1$ factorial experiment.

### D.2.2.1 $2^k$ Factorial Designs (Coarse Grids)

The simplest factorial design is a $2^k$ because it requires only two levels for each factor. These can be low and high, often denoted -1 and +1 (or - and +). $2^k$ designs are very easy to construct. Start by calculating the number of rows $N = 2^k$. The first column alternates -1 and +1, the second column alternates -1 and +1 in groups of 2, the third column alternates in groups of 4, and so forth by powers of 2. If you are using a spreadsheet, you can easily move from a design for $k$ factors to a design for $k + 1$ factors by copying the $2^k$ design, pasting it below to obtain a $2^k \times k$ matrix, and then adding a column for factor $k + 1$ with the first $2^k$ values set to -1 and the second set of $2^k$ values set to +1. Conceptually, $2^k$ factorial designs sample at the corners of a hypercube defined by the factors' low and high settings. Figure D-3 shows examples for $2^2$ and $2^3$ designs. Envisioning a $2^4$ or larger design is left to the hyper imaginative reader.

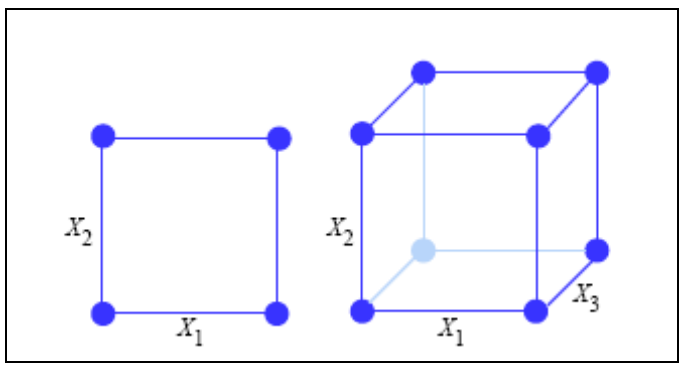



**Figure D-3. $2^2$ and $2^3$ Factorial Designs**

Factorial designs have several good properties. They allow the user to examine more than one factor at a time, so they can be used to identify important interaction effects. They are also orthogonal designs: the pair wise correlation between any two columns (factors) is equal to zero. This simplifies the analysis of the output (Y's) from running our experiment, because estimates of the factors' effects $\hat{\beta}_i$'s) and their contribution to the explanatory power ($R^2$) of the regression metamodel will not depend on what other explanatory terms are present in the regression metamodel.

From Table D-2, there are seven different terms (three main effects, two two-way interactions, and one three-way interaction) that one could consider estimating from a $2^3$ factorial experiment. Since we also want to estimate the intercept (overall mean), there are eight values we could try to estimate from eight data points. Statistically, only seven or less values can be estimated from the eight data points since the analyst always needs at least one degree of freedom (d.f.) for estimating error (and preferably, a few more).

**Table D-2. Terms for a $2^3$ Factorial Design**

| Design Point | Term | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1,2 | 1,3 | 2,3 | 1,2,3 |
| 1 | -1 | -1 | -1 | +1 | +1 | +1 | -1 |
| 2 | +1 | -1 | -1 | -1 | -1 | +1 | +1 |
| 3 | -1 | +1 | -1 | -1 | +1 | -1 | +1 |
| 4 | +1 | +1 | -1 | +1 | -1 | -1 | -1 |
| 5 | -1 | -1 | +1 | +1 | -1 | -1 | +1 |
| 6 | +1 | -1 | +1 | -1 | +1 | -1 | -1 |
| 7 | -1 | +1 | +1 | -1 | -1 | +1 | -1 |
| 8 | +1 | +1 | +1 | +1 | +1 | +1 | +1 |

A similar relationship holds as the number of factors *k* is increased. There will be *k* main effects, $\binom{k}{2} = \frac{k!}{2!(k-2)!}$ two-way interactions, $\binom{k}{3}$ three-way interactions, and so forth, up to a single k-way interaction. Adding all these yields $2^k$ - 1 terms plus the intercept. Once again, there will not be any d.f. left over for error estimation.

So, what do people do with a factorial design? One possibility is to *replicate* the design to get more d.f. for error. Estimating 8 effects from 8 observations (experimental units) is not possible, but estimating 8 effects from 16 observations is simple. Replication also makes it easier to detect smaller effects by reducing the underlying standard errors associated with the β's.

Another option is to make *simplifying assumptions*. The most common approach is to assume that some higher-order interactions do not exist. In the $2^3$ factorial of Table D-2, one d.f. would be available for estimating error if the three-way interaction could safely be ignored. One could then fit a second-order regression model to the results. Similarly, if there are data for a single replication of a $2^4$ factorial design and one can assume there is no four-way interaction, there is one d.f. for error; if one can assume there are no three-way or four-way interactions, there are five d.f. for error estimation.

Making simplifying assumptions sounds dangerous, but it is often a good approach. Over the years, statisticians conducting field experiments have found that often, if there are interactions present, the main effects also show up unless you “just happen" to set the low and high levels so the effects cancel. There is also a rule of thumb stating that the magnitudes of two-way interactions are at most about 1/3 the size of main effects, and the magnitudes of three-way interactions are at most about 1/3 the size of the two-way interactions, etc. Whether or not this holds for experiments on simulations of complex systems is not yet certain.

### D.2.2.2 $m^k$ Factorial Designs (Finer Grids)

Examining each of the factors at only two levels (the low and high values of interest) means we do not know how the simulation behaves for factor combinations in the interior of the experimental region. Finer grids can reveal complexities in the landscape. When each factor has three levels, the convention is to use -1, 0 and 1 (or -, 0, and +) for the coded levels. Consider the objective-seizing example once more. Figure D-4 shows the (notional) results of two experiments: a $2^2$ factorial (on the left) and a $11^2$ factorial (on the right). For the $2^2$ factorial, all that can be said is that when speed and stealth are both high, the agent is successful. Much more information is conveyed by the $11^2$ factorial: if the agent can achieve a minimal level of stealth, then speed is more important. In both sub graphs in Figure D-4, the blue circles, including the upper right-hand corner, represent good results, the tan triangles in the middle represent mixed results, and the red squares on the left-hand side and bottom represent poor results.

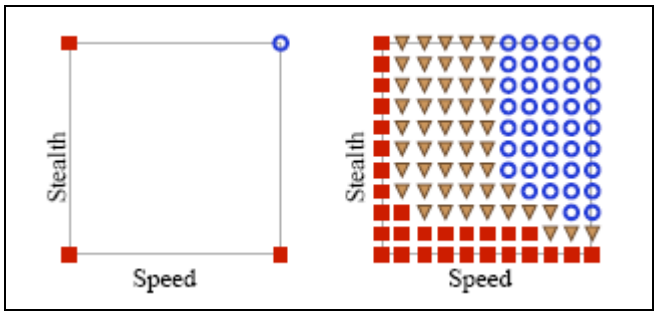



**Figure D-4. $2^2$ and $11^2$ Factorial Experiments for Seize-the-Objective**

The larger the value of $m$ for an $m^k$ factorial design, the better its space-filling properties. A scatter plot matrix of the design points shows projections of the full design onto each pair of factors. Consider the graph in Figure D-5 for a $5^4$ factorial. This graph contains cells of subplots of the design points for pairs of factors at a time. For instance, the third cell over in the top row plots the ($X_3$ , $X_1$) factor combinations, and the third cell down in the left column is just its transpose, plotting the pairs ($X_1$ , $X_3$), so it carries the same information. Each subplot has four points in the corners, three additional points along each edge, and nine points in the interior. The corresponding subplots for a $2^4$ factorial would each reveal only four points, one at each corner. The bad news is the finer grid requires 625 design points instead of 16.

Figure D-5 provides information about the design, not the output. Once the simulation has been run, the task of fitting a regression metamodel to the output is straightforward. Adding columns to the replicated design matrix will allow matching the inputs to the outputs. Start by fitting main-effects metamodels, then see if adding (centered) quadratic terms will improve the metamodel, or explore higher-order terms. Surface plots and contour plots of the average behavior can help view the results as a function of two factors at a time. These graphical methods allow the focus to be on interesting features of the response-surface landscape (such as thresholds, peaks, or at regions) without assuming a specific form for the regression model. Regression trees, interaction plots, contour plots, and parallel plots are also useful for exploring the data. Examples are found in Sanchez and Lucas (2002) or Kleijnen et al. (2005).

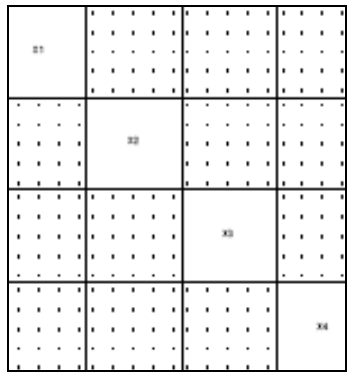


**Figure D-5. Scatter Plot Matrix for a $5^4$ Factorial Design**

Despite the greater detail provided, and the ease of interpreting the results, fine grids are not good experimental designs for a number of factors because of their massive data requirements. Even $2^k$ designs have this problem, as Table D-3 shows.

Considering the number of high-order interactions that could fit but may not be important (relative to main effects and two-way or possibly three-way interactions), this seems like a lot of wasted effort. It means smarter, more efficient types of experimental designs are needed to explore in many factors.

**Table D-3. Data Requirements for Factorial Designs**

| No. of Factors | $10^k$ factorial | $5^k$ factorial | $2^k$ factorial |
|---|---|---|---|
| 1 | 10 | 5 | 2 |
| 2 | $10^2 = 100$ | $5^2 = 25$ | $2^2 = 4$ |
| 3 | $10^3 = 1000$ | $5^3 = 125$ | $2^3 = 8$ |
| 5 | 100,000 | 3,125 | 32 |
| 10 | 10 billion | 9,765,625 | 1,024 |
| 20 | *Don't even* | $9.5 \times 10^9$ | 1,048,576 |
| 40 | *think of it* | $9.1 \times 10^{21}$ | $1.0 \times 10^9$ |

### D.2.2.3 $2^{k-p}$ Resolution 5 Fractional Factorial Designs

Sometimes many factors take on only a few levels. We might have a particular aspect, system, or capability that is either present or not present in a given SoS configuration (such as an unmanned aerial vehicle) or environmental conditions (such as the presence of civilians on the battlefield).

In these cases, one can consider variations of gridded designs. If willing to assume that some high-order interactions are not important, the number of runs required can be reduced (perhaps dramatically). This will be illustrated using a $2^k$ factorial, but the same ideas hold for other situations. Consider the $2^3$ design in Table D-2, and suppose that no interactions exist. One could call the $X_1X_2X_3$ column $X_4$, and investigate four factors in $2^3 = 8$ runs instead of four factors in 16 runs. This is called a $2^{4-1}$ fractional factorial.

As long as there are no interactions, a few more factors may be added to the study. Using Table D-2, which shows all the interaction patterns for a $2^3$ factorial, substitute in a new factor for each interaction term. The resulting design, Table D-4, is called a $2^{7-4}$ fractional factorial, because the base design varies seven factors in only $2^{7-4} = 8$ runs instead of $2^7 = 128$ runs. $X_4$ uses the column that would correspond to an $X_1X_2$ interaction, $X_5$ uses the column that would correspond to an $X_1X_3$ interaction, and similarly for $X_6$ and $X_7$. The design is said to be *saturated* since there is no room for any other factors. If the last column is ignored completely (i.e., do not have a factor $X_7$) then the six factors can be examined in only eight runs. If b = 2 replications, the seven factors can be examined in only 16 runs.

Graphically, fractional factorial designs sample at a carefully-chosen fraction of the corner points on the hypercube. Figure D-6 shows the sampling for a $2^{3-1}$ factorial design, i.e., investigating three factors, each at two levels, in only $2^{3-1} = 4$ runs. There are two points on each of the left and right faces of the cube. Each of these faces has one instance of $X_2$ at each level and one instance of $X_3$ at each level. One can isolate the effect for factor $X_1$. Averaging the results for the top and bottom faces allows the estimation of the effect for factor $X_2$, and averaging the results for the front and back faces allows us to estimate the effect for factor $X_3$.

**Table D-4. Terms for a $2^{7-4}$ Fractional Factorial Design**

| Design Point | $X_1$ | $X_2$ | $X_3$ | $X_4$ (1,2) | $X_5$ (1,3) | $X_6$ (2,3) | $X_7$ (1,2,3) |
|---|---|---|---|---|---|---|---|
| 1 | -1 | -1 | -1 | +1 | +1 | +1 | -1 |
| 2 | +1 | -1 | -1 | -1 | -1 | +1 | +1 |
| 3 | -1 | +1 | -1 | -1 | +1 | -1 | +1 |
| 4 | +1 | +1 | -1 | +1 | -1 | -1 | -1 |
| 5 | -1 | -1 | +1 | +1 | -1 | -1 | +1 |
| 6 | +1 | -1 | +1 | -1 | +1 | -1 | -1 |
| 7 | -1 | +1 | +1 | -1 | -1 | +1 | -1 |
| 8 | +1 | +1 | +1 | +1 | +1 | +1 | +1 |

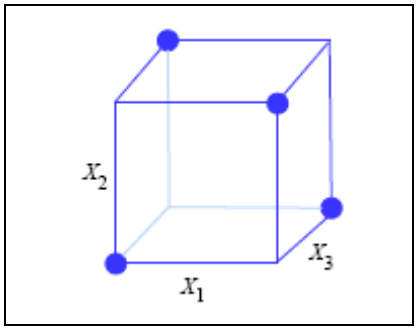



**Figure D-6. $2^{3-1}$ Fractional Factorial**

Saturated or nearly-saturated fractional factorials are often called *screening designs* because they can be useful for eliminating factors that are unimportant. They are very efficient (relative to full factorial designs) when there are many factors. For example, 64 runs could be used for a single replication of a design involving 63 factors, or two replications of a design involving 32 factors. Saturated or nearly saturated fractional factorials are also very easy to construct. However, these

designs will not do a good job of revealing the underlying structure of the response surface if there are strong interactions but they are ignored when setting up the experiment. A compromise is to use R5 fractional factorials. These allow two-way interactions to be explored but can require many fewer design points.

It is easy to create a $2^{k-1}$ factorial (called a *half fraction*) by setting up the first $2^{k-1}$ columns as if there are only k - 1 factors, and then constructing a column for the last factor by taking the interaction (product) of the first k - 1 columns. Except for the special cases when $k \leq 4$, one will also be able to estimate two-way interactions with the $2^{k-1}$ designs. Unfortunately, a half-fraction is still inefficient if *k* is large. Until recently it was difficult to find a very efficient R5 fractional factorial for more than about a dozen factors. The largest R5 fractional factorial in Montgomery (2000) is a $2^{10-3;}$ the largest in Box, Hunter, and Hunter (1978) and NIST/Sematech (2005) is a $2^{11-4}$. Sanchez and Sanchez (2005) recently developed a method, based on discrete-valued Walsh functions, for rapidly constructing very large R5 fractional factorials; a short Java program generates designs up to a $2^{120-105}$ in under a minute. These allow all main effects and two-way interactions to be fit, and may be more useful for simulation analysts than saturated or nearly-saturated designs. The Java program can be found on the Simulation Experiments & Efficient Designs at the Naval Postgraduate School website http://harvest.nps.edu/.

**D.2.2.4 Central Composite Designs**

Because $2^k$ factorials or fractional factorials sample each factor at only two levels, they are very efficient at identifying slopes for main effects or two-way interactions. Unfortunately, sampling at only two levels means the analyst has no idea about what happens to the simulation's response in the middle of the factor ranges. Going to a $3^k$ factorial would allow the estimation of quadratic effects, but it takes quite a bit more data, especially if *k* is large.

Another classic design that lets the analyst estimate all full second-order models (i.e., main effects, two-way interactions, and quadratic effects) is called a *central composite design* (CCD). Start with a $2^k$ factorial or R5 $2^{k-p}$ fractional factorial design. Then add a center point and two "star points" for each of the factors. In the coded designs, if -1 and +1 are the low and high levels, respectively, then the center point occurs at (0,0, …,0), the first pair of star points are (-c,0, …,0) and (c,0, …,0); the second pair of star points are (0,-c,0,…,0) and (0,+c,0, …,0), and so on. A graphical depiction of a CCD for three factors appears in Figure D-7. If c = 1 the star points will be on the face of the cube, but other values of c are possible.

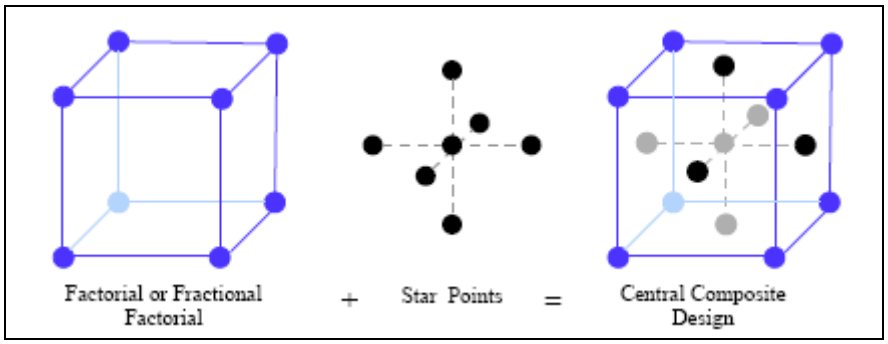



**Figure D-7. Construction of Central Composite Designs**

Although the CCD adds more star points when there are more factors, using a fractional factorial as the basic design means the CCD has dramatically fewer design points than a $3^k$ factorial design for the same number of factors. The additional requirements grow only linearly with *k*. Some examples are given in Table D-5, using the efficient R5 fractional factorials of Sanchez and Sanchez (2005) as the base designs for the CCDs. Once again, it is clear that a brute force approach is impossible when *k* is large, but efficient experimental designs allow the analyst to conduct an experiment.

**Table D-5. Data Requirements for 3-Level Designs**

| | | CCD | $3^k$ |
|---|---|---|---|
| **k** | **No. of Terms** | **No. of Design Pts** | **No. of Design Pts** |
| **2** | **5** | **10** | **9** |
| **5** | **20** | **28** | **243** |
| **10** | **65** | **152** | **59,049** |
| **30** | **495** | **2,110** | **$2.1 \times 10^{14}$** |
| **70** | **2,555** | **16,526** | **$2.5 \times 10^{33}$** |
| **120** | **7,380** | **33,010** | **$1.8 \times 10^{57}$** |

**D.2.2.5 Nearly Orthogonal Latin Hypercube Designs**

*Latin hypercube* (LH) designs provide a flexible way of constructing efficient designs for quantitative factors. They have some of the space filling properties of factorial designs with fine grids, but require orders of magnitude less sampling. Once again, let *k* denote the number of factors, and let $m \geq k$ denote the number of design points. The factor levels can be coded as *m* equally-spaced values $\left\{-1, -\frac{m-2}{m-1}, -\frac{m-3}{m-1}, \ldots, \frac{m-2}{m-1}, 1\right\}$. A *random LH design* means that each column of the design matrix is a random permutation of these *m* values, and can be constructed for any number of factors *k* provided that $m \geq k$, but co linearity problems often arise unless $m >> k$.

Figure D-8 lists a random LH with $k = 2$ and $m = 11$, and provides a picture of results that might arise by using this experimental design for our seize-the-objective simulation. Compare this design to those of figure D-4. Unlike the $2^2$ factorial design, the LH design provides some

information about what happens in the center of the experimental region, but requires far less effort than the $11^2$ factorial design.

| Speed | Stealth | Speed | Stealth |
|---|---|---|---|
| 1 | 11 | 10 | 1 |
| 3 | 5 | 4 | 2 |
| 7 | 7 | 11 | 8 |
| 2 | 3 | 8 | 9 |
| 5 | 10 | 9 | 6 |
| 6 | 4 | | |

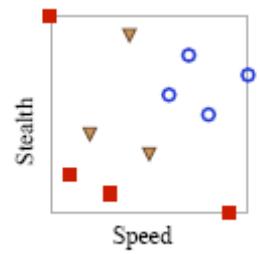



**Figure D-8. Random Latin Hypercube for Seize-the-Objective**

Cioppa and Lucas (2007) construct *nearly orthogonal Latin hypercube* (NOLH) designs that have good space filling and orthogonality properties for small or moderate *k*. Table D-6 lists the number of design points for NOLHs with $k \leq 29$. These designs are not square, but the number of design points is radically fewer than the numbers for the gridded designs in table D-5. For example, 20 factors can be explored in an NOLH with only 129 design points, as compared to over one million design points needed for a $2^2$ factorial.

**Table D-6. Data Requirements for Nearly Orthogonal Latin Hypercube Designs**

| No. of Factors | No. of Design Points |
|---|---|
| **2-7** | **17** |
| **8-11** | **33** |
| **12-16** | **65** |
| **17-22** | **129** |
| **23-29** | **257** |

Scatter plot matrices of four different designs are shown in figure D-9. These are a $2^4$ factorial design, a $4^4$ factorial design, an NOLH design with 17 design points, and an NOLH design with 257 design points. The two-dimensional space filling behavior of the NOLH compares favorably with that of the $4^4$ factorial for roughly 1/15 the computational effort, so experimenters concerned about the level of computational effort might prefer the latter. Alternatively, experimenters considering the use of the $4^4$ factorial (and thus willing to run 256 design points) might prefer the NOLH with 257 design points (just one more), and gain the ability to examine a much denser set of factor-level combinations, as well as explore up to 25 additional factors using the same design. The benefits of LH sampling are greatest for large *k*. Assuming that a single design point takes one second to run, each replication of a 29-factor experiment would take under five minutes using an NOLH design, but over 17 years using a $2^{29}$ factorial design.

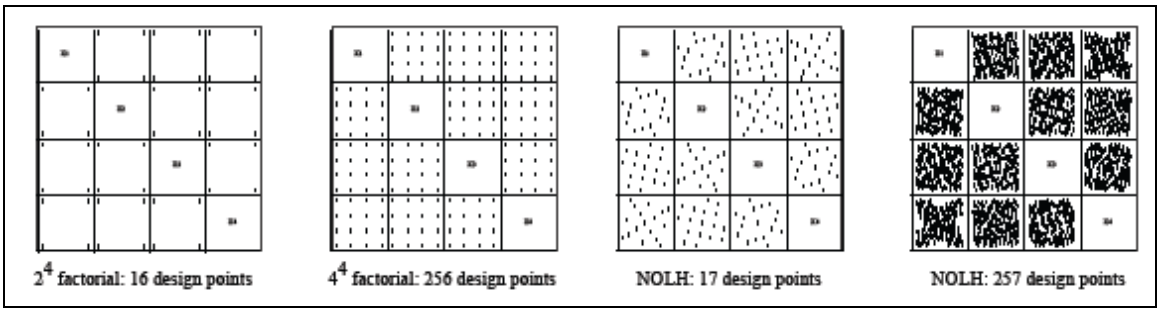



**Figure D-9. Scatter Plot Matrices for Selected Factorial and NOLH Designs**

Replicating the design allows us to determine whether or not a constant error variance is a reasonable characterization of the simulation's performance, and is highly recommended. If the time and budget is available for even more sampling, then two or more different Latin hypercubes or NOLHs can be *stacked* to obtain a larger design with better space filling properties. Stacking two designs means running both sets of design points; one way to obtain two different designs from the same NOLH matrix is to reassign the factors to different columns of the experimental design matrix.

### D.2.2.6 Robust Design Methods

Annex C of the Analyst's Handbook discusses the distinction between factors that are controllable during live tests, and factors that are not controllable during live tests. A similar distinction can be made between decision factors that can be controlled in actual operations, such as the specification of the systems that will comprise the SoS or the initial Blue force structure, and noise factors that are uncontrollable during actual operations, such as the threat force structure, threat capabilities, or variations in Blue force capabilities due to different levels of skill and training. An alternative to an exploratory analysis that seeks to understand how these noise factors affect the task measures of performance (Task MOP) and mission measures of effectiveness (Mission MOE) is a *robust design* approach, where the goal of the experiment(s) is to identify design points that yield good performance across the range of noise factor settings in other words, to identify *robust* SoS, rather than SoS that are effective only against specific threat and environmental conditions, particularly if these correspond to the most favorable settings threat and environmental factors. Taguchi (1987) pioneered the robust design philosophy for manufactured-product design, where it has been successfully used to achieve high-quality products while keeping costs in line; it also facilitates the evaluation of trade-offs between quality and cost. An important consideration for the T&E community is that the robust design philosophy explicitly requires analysts to consider variances, as well as means, in assessing system performance. *Taguchi orthogonal arrays* are discussed in several experimental design resources (e.g., Taguchi 1987) and included in some statistical software packages. These are a selected set of designs which have two-four levels and handle qualitative or quantitative factors, but are used to estimate main effects only. They are alternatives to fractional factorial or Plackett-Burman designs.

Key assumptions of Taguchi's designs are that only main effects are important, that *all* the uncertainty in the response can be ascribed to noise factors, and that all the noise factor settings can be controlled during the experiment even though they are uncontrollable in practice. The first two assumptions are problematic for experiments involving simulation models of joint SoS

test and operation (as well as for other complex, stochastic simulation models). Strong interactions and non-linear response behavior may be present, e.g., if two subsystems each must achieve a minimum level of performance in order for the overall joint SoS to achieve mission-level success. In addition, there is inherent variability in stochastic, time-driven systems that cannot be removed. For example, suppose call-for-re messages queue are routed to a single agent for processing. If the messages arrive at exactly one minute intervals, and take exactly one minute to process, then all messages can be processed with minimal (if any) delay. In contrast, if the message inter-arrival and processing times are modeled using exponential distributions (a common assumption in queuing models) with means of one minute each, then the expected queue length keeps rising (and approaches infinity as the simulated elapsed time increases). Applying robust design principles to simulation experiments is discussed in Sanchez (2000). A more detailed discussion and examples appear in Kleijnen et al. (2005), where *identifying robust systems and processes* is considered one of three primary goals of simulation experiments.

**D.2.2.7 Annotated Bibliography**

The "User's Guide to the Brave New World of Designing Simulation Experiments" by Kleijnen et al. (2005) has more detail on many aspects of performing simulation experiments. It also has an extremely detailed list of references, a detailed case study, and an online companion with additional information. This is geared toward readers with some knowledge of simulation and basic statistics who may not be familiar with experimental design, but it covers both basic and advanced designs, as well as single-stage and sequential approaches. Statisticians or other experimenters who are comfortable with planning experiments in live settings, and just beginning to perform experiments on simulation models, may find this a useful reference as well, because of detailed discussions of some aspects of simulation experiments that require the experimenter to have a different mindset.

For physical experiments where it is very costly or time-consuming to run design points, there are other single-stage designs that may require fewer design points that are often used. For example, Plackett-Burman designs and orthogonal arrays may be used instead of factorial or fractional factorial experiments. Box-Behnken designs have some advantages over central composite designs if the number of factors is small, although these advantages disappear for larger *k* (NIST/Sematech 2005). Details (including strengths and weakness) can be found in many experimental design texts; see, e.g., Ryan (2007); Box, Hunter, and Hunter (2005); Montgomery (2000); Santner, Williams, and Notz (2003); or NIST/Sematech (2005). In general, these designs are intended to handle only a small number of factors or else require the analyst to make restrictive assumptions regarding the response behavior. Flexible single-stage designs are preferred, such as NOLH designs, because they are very efficient and do not require as many assumptions. However, other than $2^k$ factorial or fractional factorial designs (which handle qualitative and quantitative factors equally well), the single-stage designs in this section are intended to be used for continuous, quantitative factors. If there are many continuous factors but also several discrete factors that take on only a limited number of potential levels, or qualitative factors with three or more levels, then special-purpose designs that involve *crossing* two or more designs may be reasonable options that are not difficult to construct. For example, if there are 3 potential command and control structures, 4 different rule-sets for airspace deconfliction, and 15 continuous factors representing various system-level factors, then we could use a crossed $2^1$ x $3^1$ x $NOLH_{65}$ design with 410 design points. Factorial designs, orthogonal arrays or Plackett-

Burman designs might be useful for specifying a design for those discrete factors with limited numbers of levels, and crossing with other designs. Hernandez, Carlyle, and Lucas (2008) have developed flexible NOLH designs that have fewer design points than the NOLH designs for the same maximum number of factors; these may be particularly useful for screening purposes or for noise factor designs.

Although the term *orthogonal arrays* often refer specifically to Taguchi's designs, factorials and fractional factorials are also orthogonal arrays, and other orthogonal arrays exist. Hedeyat, Sloane, and Stufken (1999) have a comprehensive book describing the combinatorics of orthogonal arrays.

Examples of the use of NOLH designs for agent-based simulations appear in Kleijnen et al. (2005), or Cioppa and Lucas (2007). The website of the Simulation Experiments & Efficient Design (SEED) Center for Data Farming (http://harvest.nps.edu) also has links to many papers, both methodological and application-oriented, as well as spreadsheets and software for NOLH and R5-FF designs, and is updated on a fairly regular basis.

### D.2.3 Potential Sequential Methods: Descriptions and Assumptions

#### D.2.3.1 Response-Surface Methodology

*Response-surface methodology* (RSM) is a sequential, interactive method for seeking combinations of the factors that lead to an "optimal" value (either a minimum or maximum) of the response. If there are many potential factors, RSM typically begins by using fractional factorial designs to fit first-order models as a screening procedure to eliminate unimportant factors. Once the factor space has been reduced, a series of central composite designs or other designs capable of estimating quadratic effects are used. The equations for these estimated models are used to estimate the location of the optima. The process is repeated until it appears to converge. RSM is a well-studied and powerful statistical approach, but it is not automated and convergence is not guaranteed. It is one of many used for simulation optimization, see Fu (2002) for other simulation optimization techniques.

*Response-surface modeling* refers to the process of fitting a response surface over a region of interest, one of the possible goals of exploratory analysis. It is a very general model fitting approach and can be used with any number of experimental designs. Models fit to simulation outputs are sometimes called *response-surface metamodels* because the simulation is itself a model of the process or system of interest.

#### D.2.3.2 Other Simulation Optimization Techniques

RSM is one way of seeking "optimal" solutions for stochastic simulations. Some other so called *simulation optimization* techniques are also well-known and have been used in many successful applications in practice. These include simulated annealing, tabu search, and genetic algorithms. One advantage of many of these techniques is that they are completely automated, which means that the experiments can run without constant human intervention. But a drawback of simulation optimization techniques is that they may leave the experimenter with a single "solution" (or a small set of solutions) without a corresponding understanding of why that solution is a good one.

These approaches can often be slow, and if they do not converge in a reasonable amount of time they may need to be restarted. Chang and Wan (2008) have developed a stochastic trust-region gradient (STRONG) method that has some of the benefits of optimization methods applied to deterministic systems (namely, automation and convergence guarantees). At the same time, its use of efficient experimental designs makes it applicable to the optimization of stochastic systems, and ideally it will also exhibit the more rapid convergence behavior associated with RSM.

Some of these approaches are easy to port to high-performance computing environments. Others currently work by running on a single machine.

**D.2.3.3 Modified Response-Surface Methodology**

Given that it can be difficult to understand the complex behavior of the simulation well enough to find good solutions and determine trade-offs among factors when RSM or other simulation optimization techniques are used, Schamburg and Brown (2004) describe a *modified response-surface methodology* (MRSM) approach. MRSM is a general framework intended to be applied to complex, multi-objective simulation studies. (It has sometimes been called *generalized response-surface methodology* (GRSM) or *advanced response-surface methodology* (ARSM).) The overall approach consists of the following general activities:

- Conduct an iterative sequence of experimental designs, and analyze the results.
- Begin with a large set of experiments and conduct search analysis iteration to gain a broad understanding of the entire decision space and to determine the analysis area(s) of interest.
- Use graphical and statistical tools to help identify and understand important factors, relationships, and key features of the response surfaces. Constrain the decision space based on the prioritization of the multiple responses.
- Conduct experiments and analysis in sequentially constrained decision space until near optimal solutions are found or until change in the response(s) is minimal.

This idea of beginning with a broad experiment and capturing those insights, as well as looking for regions in the factor space that yield desirable responses, is a natural one to follow. Other aspects of the response surface that may be of interest are regions where the responses are particularly bad (i.e., regions that should be avoided) and other key features of the response surfaces, like thresholds, valleys, and plateaus.

Schamburg (2004) describes MRSM as a systematic approach that does not require humans-in-the-loop in running the experimental designs, but does allow them to learn about the system through the analysis process. He prefers RSM to heuristic simulation optimization methods (like simulated annealing, genetic algorithms, tabu search, and others) because it provides the experimenter with:

- A better understanding of the relative importance of factors, interactions among factors, relationships with the responses, and the broader decision space;
- An empirical model that can be used for future analysis, rather than simply a single recommended “optimal” solution;

- The ability to reuse and adapt the empirical model to shifts in the issues that are driving the process; and
- Human learning about the decision space.

Schamburg and Brown (2004) mention that MSRM is intended to be a general framework. Specific tools and techniques can be incorporated based on the experimenter's preferences, the availability of data, etc. They make use of NOLH designs to provide the experimenter with more detailed views of the response surfaces, and make extensive use of classification and regression trees (CART) for both constructing metamodels and displaying the results. They also advocate restricted cubic splines (RCS) for regression modeling. (CARTs and restricted cubic splines will be described in section D.4.1, but the process of arriving at the constrained region is shown in table D-7) (adapted from Schamburg and Brown, 2004).

**Table D-7. Multiple Response Optimization Process through Use of CARTs**

| |
|---|
| **GOAL:** Based on prioritization of the t responses, determine constrained regions for important factors (i.e., near-optimal settings) using the following iterative sequence of analysis of the CARTs.<br><br>**STEPS**<br>(a) Use smaller pruned trees so that terminal nodes have large numbers of observations, and place the trees in decreasing order of priority.<br>(b) Set t = 1, i.e., start with the tree for the highest priority response.<br>(c) Develop constraints based on one or more “good" terminal nodes.<br>(d) Where conflicts exist, withhold judgment on appropriate constraints for that factor, or develop constraints based on the tree for the higher-priority response.<br>(e) Set t = t + 1 and go to step (c).<br>(f) If more information is needed about any factor(s), consider using larger trees and go back to step (b). |

Schamburg and Brown also discuss the use of a normalized “desirability function” (as suggested by Ben-Gal and Bukchin 2002) that weights the normalized responses and sums them to obtain an overall desirability measure. Mathematically, this is:

$$DF = \sum_{i=1}^{t} v_i z_i \tag{3}$$

where $v_i$ is the importance value of the *i*th normalized response, $z_i$. The $v_i$ are all nonnegative and sum to 1.

The $z_i$ in equation (3) does not necessarily use the same normalization. Schamburg and Brown describe different types of measures of performance (MOP). Their descriptions and normalization methods follow.

- $y_1$ = (Lethality): the number of enemy casualties, normalized as

$$z_1 = \frac{y_1 - y_{1\min}}{y_{1\max} - y_{1\min}}$$

- $y_2$ = (Seizing the Objective): zero if we do not seize the objective and one if we do, normalized as

$$z_2 = y_2.$$

- $y_3$ = (Sensor Survivability): zero if the sensor survives and one if it is destroyed, normalized as

$$z_3 = 1 - y_3.$$

This ensures that if any single MOP improves, then the overall desirability function will increase. Using the desirability function, in conjunction with more detailed information about important factors and factor constraints, can provide guidance as the experimenter seeks to summarize the findings from the series of experiments and make generalizations about what constitutes a "good" solution.

**D.2.3.4 Group Screening**

While MRSM allows the experimenter to explore across the broad input space and identify sub regions that yield desirable behavior, sometimes the cost of doing a detailed series of experiments can be prohibitive. An alternative is to use a sequential group screening approach. Group screening approaches were developed by Dorfman (1943) during World War II as a means of efficiently conducting blood tests to screen new Army recruits for syphilis. By taking blood samples from a large group of soldiers and pooling a portion of each sample, a single test could be used to clear a large group as long as the test came back "negative." If the test was positive, the soldiers' samples were tested in smaller groups until those with the disease could be isolated.

The same idea is used to identify factors with important effects from the larger list of potential factors. *Controlled sequential bifurcation* (CSB) is an extremely efficient technique that allows the experimenter to specify two thresholds (one above which all factor effects are deemed critical; the other below which all factors are unimportant). The CSB procedure begins with all samples in a single, large group, then recursively breaks the groups into roughly half until all factors have been identified as important or unimportant. CSB is not only efficient, but it also has the nice feature of guaranteed limits on Type I and Type II errors (false positives and false negatives).

A drawback to CSB is that it requires the experimenter to know the direction of all effects before experimentation begins. This is often not reasonable, especially for complex systems or future systems, where little data exist. Fractional Factorial (FFCSB) and FFCSB-X (where the "X" indicates that two-way interactions are controlled using a fold over fractional factorial design) are two-phase procedures that extend CSB by adding an initial phase where a fractional factorial design is conducted, splitting the factors into two groups depending on the sign of their estimated effect, and then employing CSB (Sanchez, Wan, and Lucas 2008). These procedures can be

extremely efficient. FFCSB assumes that interactions are either negligible or, if interactions exist, they can be identified by the main effects. FFCSB-X allows important main effects to be identified appropriately even if two-way interactions exist in the underlying simulation model.

#### D.2.3.5 Annotated Bibliography

There are many books on response-surface modeling, including Box and Draper (1987) and Draper and Smith (1998). Information about the nonparametric technique of multivariate adaptive regression splines (MARS) can be found in Friedman (1991).

### D.2.4 Design Comparisons

Figure D-10 provides some guidance about experimental designs for simulation experiments. This is not intended to be an exhaustive list of all possibilities, but hopefully it will help experimenters identify some suitable designs for particular contexts. The focus is on designs discussed in section D.2 that are relatively easy to generate, implement, and analyze, as well as to interpret and explain the results. However, if circumstances dictate that none of the designs are suitable, alternatives are often available from a statistician or a detailed look at some of the experimental design references. A version of this chart is maintained at the SEED Center web pages (http://harvest.nps.edu), and updated as new designs become available to fill some of the gaps. All acronyms are defined on the web site.

Selecting a design is an art, as well as a science. Clearly, the number of factors and the mix of different factor types (binary, qualitative or discrete with a limited number of levels, discrete with many levels, or continuous) play important roles. These are rarely cast in stone, particularly during exploratory analysis. The experimenter has control over how factors are grouped, how levels are determined, etc. Even if these are specified, different experimenters may prefer different designs. Figure D-11 summarizes the process that an experimenter goes through to come up with an initial experimental design.

| | $2^k$ factorial | $m^k$ factorials, 3<=m<=5 | $m^k$ factorials, 6<=m<=10 | R3FF, orthogonal arrays | Foldover designs | R4FF, Plackett-Burman designs | R5FF | Central composite with full factorial | Central composite with R5FF | Random LH with n>>k | Smallest possible NOLH (i.e., very few extra columns) | Larger NOLH | Crossed NOLHs | FFCSB (main effects) | FFCSB-X or Hybrid method |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **FACTOR CHARACTERISTICS** | | | | | | | | | | | | | | | |
| **Total number of factors: 2-6** | B*● | L*● | ● | ● | ● | ● | ● | ● | ● | ● | ● | C*■ | ◍[3] | ◍ | ◍ |
| **Total number of factors: 7-10** | ● | ○[1] | | ● | ● | ● | B*● | ● | ● | ● | ● | C*■ | ◍[3] | ◍ | ◍ |
| **Total number of factors: 11-29** | | | | ● | ● | ● | B*● | | ● | ● | ● | C*■ | ◍[3] | ◍ | ◍ |
| **Total number of factors: 30-99** | | | | ● | B*● | ● | ○[2] | | ○[2] | ●[4] | | | ● | ●[1] | ● |
| **Total number of factors: 100-300** | | | | ● | B*● | ● | ○[2] | | ○[2] | ●[4] | | | ○[3] | ●[1] | ● |
| **Total number of factors: 300-1000** | | | | ● | ○ | | | | | | | | | ●[1] | B*C*● |
| **Total number of factors: 1000-2000** | | | | ● | ○ | | | | | | | | | ●[1] | B*C*● |
| **Binary factors** | B*● | | | ● | ● | ● | ● | | | | | | | ● | |
| **Qualitative factors with 3 or more levels** | | L*● | ● | | | | | | | | | | | | |
| **Discrete or continuous factors treated as binary** | ◍[1] | | | ◍[1] | ◍[1] | ◍[1] | ◍[1] | | | | | | | ● | |
| **Discrete factors, 3-5 levels of interest** | ◍[1] | ● | | | | | | ● | L*● | | | | | | |
| **Continuous factors, or discrete with many** | | ● | ● | | | | | ● | ● | ● | ● | ● | ● | | ●[2] |
| **Decision factors (controllable in real world)** | ● | ● | | | | | ● | | | | ● | ● | | | |
| **Noise factors (uncontrollable in real world)** | | | | ● | ◍[4] | ◍[4] | ◍[4] | | | ◍[5] | ● | | | | |
| **RESPONSE CHARACTERISTICS** | | | | | | | | | | | | | | | |
| **Main effects only (initial screening)** | ◍[2] | | | ●[1] | ■ | ■ | ■ | | ■ | | ● | ◍[6] | | ●[1] | ■ |
| **Main effects (valid when 2-way interactions** | ◍[2] | | | | ● | | | ▒[2] | ■ | | | | | | ● |
| **Main effects and all 2-way interactions** | ◍[2] | ■ | ■ | | | | ● | ■ | ● | | | | ● | | |
| **Main effects and many interactions** | ● | ◍[2] | ◍[2] | | | | | ● | | ●[5] | ●[5] | ●[5] | ● | | |
| **Quadratic effects** | | ◍[2] | ◍[2] | | | | | ● | ● | ● | ● | ● | ● | | |
| **Thresholds / non-smooth effects** | | ◍[2] | ◍[2] | | | | | | | ● | ● | ● | ● | | |
| **Flexible modeling – not all pre-specified** | | ◍[2] | ◍[2] | | | | | | | ● | ● | ● | ● | | |
| **OTHER CONSIDERATIONS** | | | | | | | | | | | | | | | |
| **Batch mode unavailable – all runs through GUI** | | ◍[2] | | ●[3] | | | | | ●[3] | | ●[3] | | | | |

**LEGEND:**

■ Provides additional modeling flexibility or allows some assumptions to be assessed

**B*** Good design choice for binary factors

**L*** Good design choice for factors with a limited number of qualitative or discrete levels

**C*** Good design choice for continuous factors, discrete factors with many levels

● Works well

- ●[1] Assumes that interactions are negligible or that they'll show up with the main effects – must follow up with confirmation runs
- ●[2] For FFCSB-X, "many" means 2 or 3 levels
- ●[3] Smaller designs are the only ones feasible until this gets "fixed" – work with the developer
- ●[4] Design' correlation structure must be checked – stacking many designs may be an alternative
- ●[5] Degrees of freedom limit the number of terms that can be estimated simultaneously, so not all main effects and two-way interactions can be estimated simultaneously

○ Consider these designs if additional computing resources are available

- ○[1] These require many more runs than other designs unless k is small. Consider NOLH designs.
- ○[2] Start with 2 replications and see if you can eliminate any factors – each time you do, you effectively double the number of replications for factors that remain.
- ○[3] Same as above, but to avoid overly-large designs may want to consider saturated or nearly-saturated NOLH

▒ Potential designs that provide additional modeling flexibility or allow some assumptions to be assessed, but typically require more design points

◍ Potential designs, but better designs exist for this purpose

- ◍[1] Unless used for initial screening, it may be a good idea to explore at least 3 levels
- ◍[2] These require many more runs than other designs unless k is small. Consider R5FF (for binary) or NOLH designs
- ◍[3] Easier to use a larger NOLH (if all factors are quantitative) or cross a full factorial for factors with just a few levels with an NOLH
- ◍[4] Since you do not need to estimate interactions among noise factors, use a screening design like R3FF or a small NOLH
- ◍[5] In the spirit of keeping noise factor designs small, you might prefer an NOLH
- ◍[6] If you're interested in screening and want to keep the number of runs down, go for one of the smaller LH designs

**Figure D-10. Design Comparison Chart**

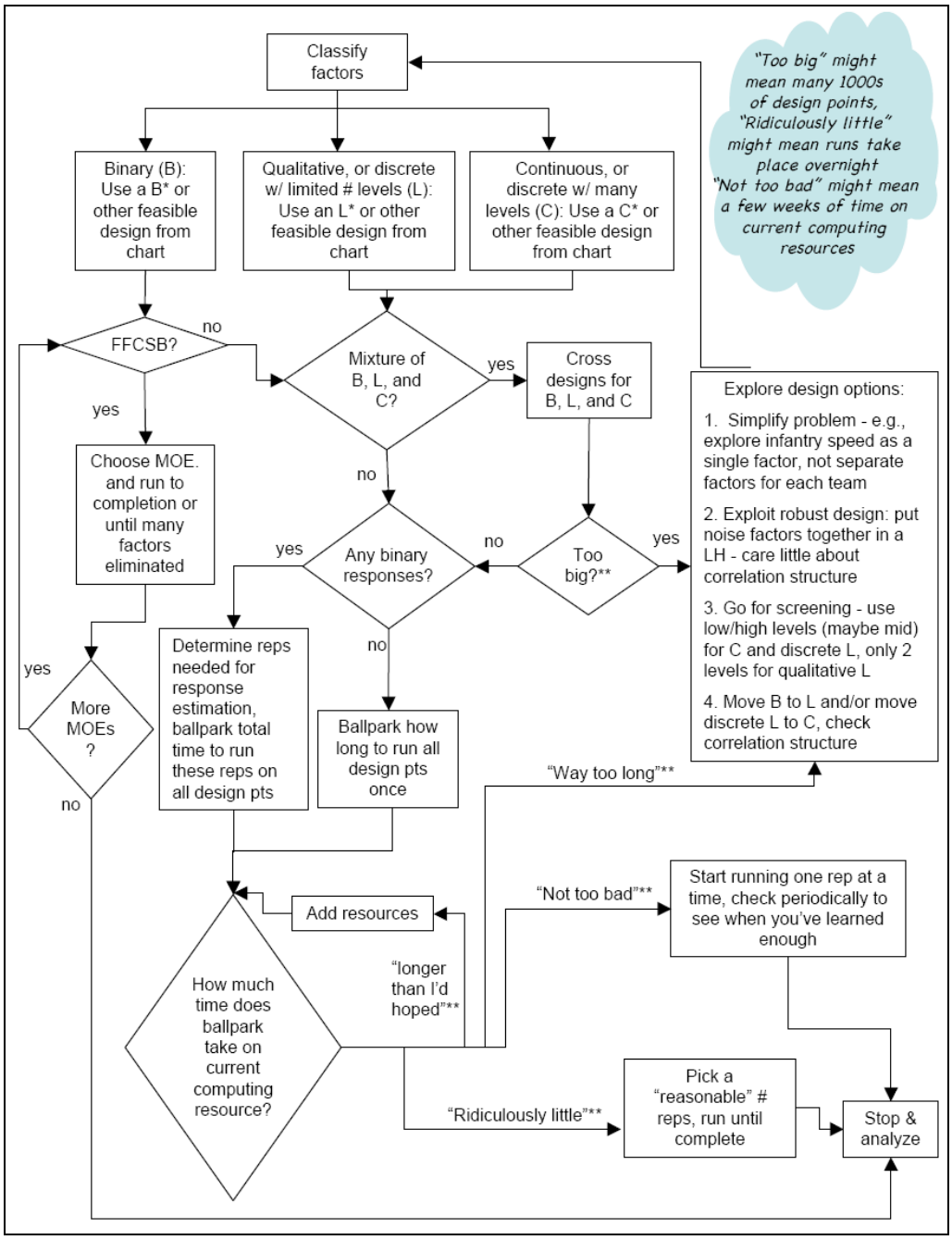



**Figure D-11. Flowchart of Initial Design Selection Process**

## D.3 Special Opportunities With Computer-Simulation Experiments

This section applies only to computer-simulation experiments, and is intended to highlight some specific opportunities in that environment to enhance the statistical efficiency even beyond what efficient experimental designs offer.

### D.3.1 Randomization

First, unlike physical experiments, there is no need to randomize treatments to "experimental units," or to randomize the order of experimentation. This is assuming that the random number generator used is of high quality:

- Long cycle length (longer than can be exhausted in years of nonstop computation).
- Demonstrated valid statistical properties (uniformity and independence).

### D.3.2 Variance Reduction

Since variance in simulation output represents the imprecision in the results (or looked at another way, statistical inefficiency), every opportunity should be taken to reduce output variance. Doing so will either provide more precise results for the same computational effort, or (equivalently) allow a desired level of precision with less computation. It has long been known that various *variance-reduction techniques* (VRT) can sometimes be applied, and sometimes to dramatic effect (though that is model- and input-parameter dependent). Since, in a DOE context, many different scenarios are compared, one particular VRT, *common random numbers* (CRN) is of special interest. This is a highly intuitive notion that can reduce the variance in differences between the output performance measures of different scenarios, sometimes substantially. The basic idea is that by re-using the same random numbers to simulate all scenarios in the design, that any output differences seen will be attributable not to variation in basic input random numbers (since they are the same), but rather to the fact that the scenarios are different, which is what we want to detect. Effective implementation of CRN, however, is seldom as simple as it appears and there are a number of considerations:

- For CRN to work as intended, it is not enough just to start with the same seed in a single random-number stream (sub segments of the underlying generator, effectively independent sub-generators) since in general the random-number use will get mixed up across the scenarios, diluting and usually destroying the effect of CRN. Rather, the experimenter must synchronize random-number use across scenarios, so the same random numbers are used for the same purpose across scenarios in the design.
- Effective synchronization always requires a detailed understanding of exactly how the simulations of all scenarios in the design work, and the ability to manipulate exactly which random numbers are used for which purposes.
- Two specific synchronization techniques are often of use. First, the designer of the simulation should use different streams for each source of randomness in the model. A random-number stream is just a sub segment of the complete random-number cycle, and should be easily addressable by stream number, rather than by seed-vector values, which users will hardly ever know. Modern random-number generators have streams built in, which do not overlap and are of astronomical length so that experimenters never have to worry about overlapping. Second, in some situations it might be best to pre-assign all random aspects of a simulated entity's experience, whether they will be needed or not along the way, and stored with the entity as an attribute to be used downstream in the logic. This might sometimes afford better synchronization than the first method, but has the downside of more memory use to store all attributes for all entities.

- If the simulation is replicated independently, as many are, then care must be taken to maintain synchronization for all replications. The best way to do this is to divide further each stream into many long sub streams, which are automatically incremented with each replication. In this way, synchronization will be maintained after the first replication, since all scenarios will start the second and each subsequent replication at the beginning of a "fresh" sub stream, even if different scenarios consume different numbers of random numbers from their streams during a particular replication.
- Due to the logical structure of some simulations, it may not be possible to implement perfect random-number synchronization across all scenarios in the design. In such cases, the experimenter should generate independently across the scenarios for these sources of randomness in the model. This is generally not automatic, however, and the experimenter thus needs to ensure that separate streams for such sources are explicitly specified for the sources that are to be generated independently. This could be done by developing a formula for the random-number stream that depends on both the scenario number and replication number; for instance, if there are 129 sources of randomness across the scenario to be synchronized, plus 3 more to be generated independently, and 100 replications are to be run, streams 1-129 could be used for the synchronized sources, then stream $200 + r$, $300 + r$, and $400 + r$ for the three independent streams, respectively, where $r$ is the replication number.
- Implicit in all of this is that the underlying random-number generator must be very good, and support a very large number of very long streams and sub streams. While such generators have been developed in recent years, they are not always implemented in simulation software, including some agent-based combat simulations. There is a real need to improve this situation to support VRTs in simulations, which often take a lot of computer time, rendering VRT of particular importance.

### 3.3 Effective Use of Computing Resources

Implementing DOE for simulation experiments can be a non-trivial task. A few considerations include:

- Setting up some type of automated run control is very important. There should be some type of automated program for converting the experimental design file into the necessary input values or file(s), executing the simulation, matching the output from each run back to the inputs, and the random-number streams used to generate that output, and consolidating the results into a single file for analysis. This has several benefits. It provides traceability; if some results appear anomalous, the corresponding runs can be repeated. It dramatically reduces the potential for error by removing the need for hand-entry of multiple items. It also cuts down on the overall time to run the experiment by eliminating the need for an analyst to physically launch each run.
- Because the run time can vary dramatically across design points for complex simulations, the experimenter should make a few pilot runs to get a rough idea of the time required before launching the full experiment. For example, in the objective-seizing scenario it might be that runs leading to "good" outcomes (i.e., the blue circles on the graphs in Figures D-4 and D-8 occur when the objective is seized quickly, so the simulation also finishes quickly; similarly, the "bad" outcomes (i.e., the red squares in these graphs) might occur when there is a protracted engagement, meaning that the corresponding runs will take longer to complete.

- The total run time for a particular experiment can be reduced by running the simulation on multiple processors. This could happen in a high-performance computing environment, or it could be that a small number of workstations (perhaps with multiple processors) are used. However, while multiple processors can reduce the total time linearly, the only way to study a large number of factors is to use an efficient experimental design.
- If a single replication of the entire experimental design can be run very quickly, the experimenter may simply specify a "reasonable" number of replications and wait until all runs are complete before beginning analysis. Some simulations automatically spawn runs and consolidate the data for multiple replications at a single design point; running all replications for design point 1, followed by all representations of design point 2, and so forth, is an easy process to automate. Another alternative is to step through the entire design one replication at a time. This has several advantages. First, it means that a preliminary look at the results can be taken once one or two replications of the entire design are complete. (Note that if runs are made on a cluster, jobs do not necessarily finish in the order in which they start). If obvious problems in the output are apparent, the experiment can be terminated early and restarted after the problems have been identified and fixed. Alternatively, at any point in time after the first replication is complete, the experimenter is able to do some preliminary output analysis. This means that the model can be run until the experimenter has learned enough; there is no need to do approximate sample-size calculations before beginning the experiment.
- DOE can also be used as part of the vignette-creation process. Conducting experiments on initial versions of the vignettes provides some insight into the simulation's performance. Focusing efforts on those systems, subsystems, or attributes that seem to drive performance may allow the simulation developers to add detail until they have modeled enough, rather than try to incorporate all the detail from the outset. Conducting a series of small experiments as new features are added to the model also assists in debugging efforts.

### D.3.4 Annotated Bibliography

Variance reduction is discussed in more detail in most simulation texts, such as Banks et al. (2005), Bratley et al. (1987), Law (2007), and Leemis and Park (2006).

For those interested in harnessing the power of computing clusters without purchasing new hardware, software such as Condor (freeware developed at the University of Wisconsin, see http://www.cs.wisc.edu/condor/) can be used to distributed jobs to idle machines across a network.

## D.4 Examples: Designs, Analyses, and Insight

Several small examples using different types of designs for notational test planning support simulation experiments are provided. Brief descriptions are provided of the notional scenarios, select appropriate experimental designs, and provide numerical and graphical summaries of the results. These are intended to demonstrate the types of insights that test planners might gain from computer-simulation experiments prior to live test events, but are provided for illustrative purposes only.

### D.4.1 Potential Analytic and Graphical Methods

Sequential screening approaches like FFCSB may provide the experimenter a list of important factors (hopefully much smaller than the list of potential factors). Sequential approaches like response surface methodology, evolutionary algorithms, or STRONG may finish with specific alternatives of interest, such as those associated with a particularly good (or bad) MOE. If the goal is to develop a broader understanding of the underlying system, there are several analytic and graphical methods that can be used during the exploratory analysis. Some of those are summarized here:

- Many descriptive statistics plots are valuable before jumping into a more detailed analysis. Histograms can be used to make sure that the simulation is not giving unreasonable response values. One-way plots of the response vs. a single factor, such as a series of box plots or a means-diamonds plot, can be used to see if there are any obvious patterns in the data. Scatter plot matrices that include not just the factors, but also the responses, give some indication of whether the relationships between individual factors and the responses are strong or weak, as well as whether there are any unusual patterns. Scatter plot matrices for the MOEs alone can indicate whether these are strongly correlated with one another, whether it is likely that trade-offs will need to be made, or whether a few alternatives clearly present themselves as superior.
- *Least-squares regression* was previously discussed as a method for fitting a statistical model to the simulation output. As we discuss in section D.2.1.1, it can be used to fit a variety of types of models, such as main-effects models or second-order models, but the model-building process is an art as well as a science. A model that does a good job of fitting the input/output relationships with a limited number of terms is desirable. In general, the number of terms in the final model will be far fewer than the number of potential terms. All terms that are included should be statistically significant with one exception, if a particular factor appears in a significant higher-order term (e.g., the $X_1X_2$ interaction or the quadratic effect for $X_1$ is statistically significant), then $X_1$ should also be included even if it does not have a statistically significant effect on its own. As discussed in section D.2.1.4, the analyst might decide not to include some statistically significant terms in the final model because they have little practical impact. Some graphs that are quite valuable with regression models include interaction plots, contour plots or surface plots, and prediction profilers. Residual plots may also be useful. The reader is cautioned that when the number of runs is large, these can look radically different from the sample residual plots seen in textbook examples.
- *Stepwise regression* is a variant of least-squares regression. The experimenter must specify two threshold significance values that determine whether terms meet statistical criteria to enter or remain in the model, but then the procedure runs automatically to populate the model. At each step, the term with the lowest p-value (i.e., with the most statistical significance) is added to the model (along with any of its lower-order effects), the significance is recalculated, and the p-values are checked to see if any existing model term should be dropped. Stepwise regression is useful when there are far more potential terms than the number of terms that can end up in the model. For example, a second-order model for 20 factors has 20 potential main effects, 20 potential quadratic effects, and 190 potential two-way interactions. It is impossible to estimate 230 terms simultaneously if an NOLH with 129 runs is used; instead, stepwise regression can be a quick method of culling out the much smaller number of terms that drive the performance. Stepwise regression can be

problematic when applied to observational data, which is another reason why it is a good idea to use well-designed experiments. Stepwise regression is sensitive to outliers in the dataset. Any graphs useful for regression are also useful for stepwise regression.

- *Logistic regression or multinomial logistic* models can be used when the response is binary or qualitative and there are one or more quantitative factors. The fitted model estimates the probabilities of falling into each category as a function of the input factors. For example, suppose one MOE for the objective-seizing scenario is simply whether we are successful (the MOE Y = 1) or unsuccessful (Y = 0) at seizing the objective by a pre-specified time. If p represents the (unknown) probability of success, then the underlying model is assumed to take the form in equation (4):

$$p = \frac{1}{1 + \exp\left[-f\left(X_1, \ldots, X_k\right)\right]} \tag{4}$$

The function $f(X_1,\ldots,X_k)$ can be a main-effects model, such as

$$f\left(X_1, \ldots, X_k\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i$$

or take on a more complicated form that includes interactions or polynomial terms. Regardless of the functional form of $f(\ )$, the fitted predictions will all fall between 0 and 1, inclusive, however that may not happen if linear regression is used to estimate the probabilities directly.

Graphs that show the logistic curves for each response category, or contour or surface plots that show how the predicted probabilities vary as a results of a few key factors, help provide insight into the simulation's behavior.

- *Partition trees* are recursive partitioning techniques that can be used to fit either qualitative or quantitative responses. They go by several other names, including decision trees and CART. Partition trees find a set of cuts or groupings of factor values that best predict the response of interest by exhaustively searching all possible cuts or groups at each level. This splitting process continues until the desired fit is reached. The result is a tree that can be easily traversed by following simple decision rules, and the “leaves” at the bottom of the tree correspond to the final groupings. This technique is quite useful in exploratory analysis because it handles large problems easily, it does not require the experimenter to have a good prior model, and the results are easy to interpret and explain. For example, interactions need not be specified ahead of time; if interactions are present, then the left and right branches will split along different factors. Partition trees are easy to display graphically if the number of leaves is not too large. Other useful items (in either graphical or numeric form) are summaries of the contributions of the factors to the overall fit, and the number of splits associated with each factor in the tree. This helps when the same factor appears in several different cuts within the tree.
- *Restricted cubic splines* fit piece-wise linear functions, rather than smooth curves, based on the chosen number of “knots" in the function (typically, between 2 and 5). These can be used

to fit a variety of patterns in the data while fitting linear relationships between the factor and the response in the tails.

### D.4.2 Sample Designs

Overviews are provided of the initial design selection process for a notional seize-the objective simulation where there are a total of 14 factors of interest. Each subsection heading indicates one type of design that could end up being chosen, while a context associated with that choice is provided in the opening paragraph. These are provided for illustrative purposes only.

#### D.4.2.1 Large $2^k$ Fractional Factorial or CCD Experiments

Suppose that nine of the 14 factors are qualitative, such as the presence or absence of particular subsystems, and there is also one discrete-valued factor with three levels that indicate the number of unmanned vehicles associated with the headquarters squadron (0, 1, or 2), two discrete-valued factors with five levels each, and three continuous factors that describe certain system-level attributes. The initial classification of factors into sets B, L, and C (for binary, limited-value, and continuous factors, respectively) has B= $\{X_1,\ldots,X_9\}$, L= $\{X_{10},X_{11}\}$, and C= $\{X_{12}, X_{13},X_{14}\}$. Crossing an R5FF for factors in B with a $3^1$ x $2^2$ design for factors in L and a 65-run NOLH design for the factors in C allows a detailed exploration of this factor space, but requires 128 x 75 x 65 = 624,000 design points. This is a very large number, so other designs will be explored. Items to consider include:

1. There are no factors to simplify.
2. May want the option to examine interactions among all 14 factors, so may prefer not to use a screening design for those that are uncontrollable in the real world. May consider crossing designs for the decision factors and noise factors that are each suitable for exploring models with two-way interactions. Suppose the decision factors are $\{X^1, X^2, X^{10}, X^{11}\}$. Some options are:

   (a) Crossing a R5FF (which happens to be a full factorial) for these four factors with a $2^{5-1}$_ R5FF for the ten noise factors. This yields 16 x 128 = 2,048 design points.

   (b) Adding star points and two center points for the non-binary factors to obtain CCDs for the decision factor and noise factor designs. This yields $2^2$ x ($2^2$ + 6) 40 design points for the decision factors and $2^{8-2}$ x ($2^2$ + 6) = 640 design points for the noise factors, for a total of 25,600 design points.
3. For screening purposes, if sample initially at just the low and high levels for each quantitative factor is chosen, may have possibilities that require even fewer runs. These include:

   (a) Use an R3FF with 16 design points, recognizing that interactions will be confounded with main effects.

   (b) Use a foldover design with 32 design points, which controls for two-way interactions but does not allow us to estimate them.

   (c) Use an R5FF with 256 design points.

   (d) Augment the R5FF with ten star points and two center points (to allow quadratic effect estimation for all the non-binary factors) for a total of 268 design points.

4. If treating the factors in L as continuous, it could be possible to cross an R5FF for the nine binary factors with a fairly large (say, 129 design point) NOLH for the rest, rounding the discrete-valued factor levels appropriately. The maximum pair wise correlation in this NOLH is 0.087. The total number of design points is 128 x 129 = 16,512 design points.

The decision may be made that the designs in 3(c) and 3(d) above are not too big in terms of the design space. If the simulation model has no binary measures and pilot runs show it takes only about 25 seconds per run, a single replication takes under two hours to complete; could run 30 replications of this experiment in under 2.5 days on a single machine. If able to let this run over a weekend, do not need to be concerned with run order or the ability to obtain preliminary information along the way.

**D.4.2.2 NOLH Design**

If 8 of the 14 factors are continuous, and 6 factors are discrete-valued with 4 levels, the initial classification of factors into sets has C= $\{X_1, X_2, \ldots, X_8\}$ and L= $\{X_9, \ldots, X_{14}\}$. Crossing a $4^6$ for the factors in **L** with an NOLH with 33 design points for the factors in C yields 4096 x 33 or 135,168 design points. This, again, is a large number, and other designs should be explored. Some options are:

1. If the six discrete-valued factors represent the number of subsystems in each of six locations, could consider these to be a single factor (held constant across all locations during a simulation run). If so, this reduces the factors to one 4-level discrete factor and eight continuous factors. These could be explored by crossing a 41 factorial and an NOLH with 33 design points, yielding 4 x 33 = 132 total design points.
2. If all of the factors are considered decision factors, you cannot generate design alternatives by differentiating between decision and noise factors.
3. If all factors are treated as binary, could construct a full factorial with 214 = 16,384 design points or an R5FF with 256 design points.
4. If all factors are treated as continuous, several different NOLHs are possibilities:

   (a) An NOLH with 65 design points. Assigning the four discrete-valued factors to the first four columns and rounding appropriately yields a maximum pair wise correlation magnitude of 0.107.

   (b) An NOLH with 129 design points. Assigning the four discrete-valued factors to the first four columns and rounding appropriately yields a maximum pair wise correlation of 0.101.

   c) An NOLH with 257 design points. Assigning the four discrete-valued factors to the first four columns and rounding appropriately yields a maximum pair wise correlation of 0.064.

Based on this information, the NOLH with 257 design points might be preferred. It has essentially the same number of design points as the R5FF, yet provides more modeling flexibility. (Stacking two NOLHs might reduce the maximum pair wise correlation even further.)

Perhaps the runs take a moderately long time, even on a computing cluster, so the estimate is that a single replication of this NOLH may require 10 hours of computing time. If the system is deterministic and we need run it only once, then the run order is of no concern. If the system is stochastic and will require an unknown number of replications to provide good insight, this run time may be considered "not too bad," but starting all runs for a single replication before moving to the next replication may be advantageous. It is possible to run about 5 replications over a weekend, so the results can be examined after a few days and get a better idea of how many replications are necessary.

**D.4.2.3 Robust Design**

Suppose that 8 of the 14 factors are continuous, and 4 factors are discrete-valued with many levels, and 2 factors are qualitative with three levels each. The initial classification of factors into sets has C= $\{X_1, X_2, \ldots, X_{12}\}$ and L= $\{X_{13}, \ldots, X_{14}\}$. Crossing a $3^2$ for the factors in L with an NOLH with 257 design points for the factors in C yields 9 x 257 = 2,313 design points. Smaller NOLHs are possible, but figure D-10 indicates that larger ones are preferred because they provide very detailed exploration of the response behavior corresponding to the quantitative factors. This is a manageable number of design points. If the division of factors into L and C also splits them into controllable and uncontrollable factors, the design is particularly easy to interpret. For example, if the only decision factors are the two qualitative factors, then we are running (and replicating) the same 257 test scenarios for each of nine possible decision-factor configurations. Of course, the decision factors do not necessarily break down along the same lines as the factor characteristics. Suppose that the decision factors are the first four factors in C and the two qualitative factors. Then options for a crossed design decision-factor design x noise-factor design include:

- An NOLH with 17 design points crossed with a 32 (for the qualitative factors) for the decision-factor design, yielding 153 design points. One possible design for the noise factors could be obtained by treating all eight noise factors as continuous and using an NOLH with 33 design points. The total number of design points would then be 153 x 33 = 5,049 design points.

**D.4.2.4 Screening Design**

The last example supposes that there are a much larger number of factors (for example, 100) and all these factors are binary. Potential designs from Figure 10 include R3FF designs, fold over designs, R5FF designs, or the sequential screening methods FFCSB and FFCSB-X. An R3FF design for 200 factors has 128 design points, a fold over design has 256 design points, and an R5FF design has 32,768 design points. If the simulation runs quickly we might consider running the R5FF to explicitly allow estimation of two-way interactions once the runs are complete. Although this appears to be a large number of design points, we may need very few replications if many of the factors have little impact on the response.

Alternatively, it may be possible to winnow the potential factors down to a much smaller list by using FFCSB-X. This eliminates the need for the experimenter to assess the results periodically to see whether more replications are needed. Depending on the error structure and the underlying pattern of factor effects, this may be accomplished in a very small number of runs. For example, studies of the performance of FFCSB-X in identifying important factors show that

there are cases where FFCSB-X successfully classifies all critical and unimportant factors from a list of 127 potential factors in under 3,000 total runs, and it limits the probabilities of false negative and false positives (Sanchez, Wan, and Lucas 2008). This option may be preferred to the single-stage screening designs listed above, particularly if the individual simulation runs take a long time or all runs must be completed on a single machine.

### D.4.3 Gaining Insight

Kleijnen et al. (2005) describe three basic goals of simulation experiments:

- Develop a basic understanding of a particular simulation model or system;
- Finding robust decisions or policies; and
- Comparing the merits of various decisions or policies.

Experimental design approaches, coupled with graphical and analytic methods such as those described in section D.4.1, can be useful for all these goals. They help the experimenter develop a better understanding of the system. This process typically follows an iterative cycle of model-test-model, where insights gained from simulation experiments can be used in many ways. Results can be used to evaluate or improve the simulation model. By identifying important factors, interactions, and nonlinear effects, the experimenter can improve their understanding, find robust solutions, or raise questions to be explored in subsequent experiments. Thresholds, plateaus, or other interesting features of the response surfaces might provide guidance about situations that are particularly good (or particularly bad).

The results of the exploratory analysis may also be useful in selecting live test conditions. Some possibilities include:

- Selecting design points that correspond to unusually or exceptionally high or low mean measurement values (outliers).
- Selecting design points that correspond to desirable mean measurement values that also have high variances. This might confirm whether certain system configurations are very fragile, i.e., whether the results vary widely depending on the uncontrollable environmental conditions.
- Selecting design points that correspond to desirable mean measurement values that also have low variances. This might confirm whether certain system configurations are, indeed, robust to changes in uncontrollable conditions.
- Selecting design points that allow investigation into any counterintuitive or surprising results discovered during the exploratory analysis. Such findings might consist of significant factors with unexpected signs, surprising values where the form of the response changes, such as places where “knees in the curve” or jumps in performance occur; unexpectedly strong interactions between factors; or factors that were anticipated to be important but have little impact in the simulation model.
- Using the results of the data analysis to identify threat characteristics that are sufficiently challenging, to supply the conditions for the enemy in the live test.

In all cases, data obtained from live tests can be useful for a variety of purposes. Such data could be used to determine more realistic input distributions or low and high factor settings. Results from live tests can be compared to those from simulation experiments to determine whether enhancements or changes to the model are required. Over time, leveraging results obtained from simulation experiments may improve the test community's ability to field better systems in a timely manner with a small but effective set of live test events.

### D.4.4 Annotated Bibliography

Further information about many of the tools described in section D.4.1, as well as other data mining techniques, can be found in introductory texts such as Hand, Mannila, and Smyth (2001) or Tan, Steinbach, and Kumar (2005).

Examples for simulation experiments appear in Sanchez and Lucas (2002), Kleijnen et al. (2005), and Lucas et al. (2007). Many other applications can be found via the SEED Center web pages at http://harvest.nps.edu.

## D.5 Conclusions

This annex is intended to support the T&E community in planning an efficient and informative set of tests by focusing on several aspects of DOE. We can use DOE directly to plan the testing event itself, as well as plan a set of simulation experiments prior to live testing, intended to reduce the number of important factors and thereby making for a far more efficient test event. The general ideas of DOE have been explained, along with traditional terminology such as factors, responses, effects, and interactions. A table and flowchart attempt to provide aid in selecting specific designs under a variety of circumstances concerning the number and nature of the factors, the kind of information sought, and the difficulty (or time) involved in running a test. While most of these ideas apply both to live tests and to simulation tests, some of the special opportunities when conducting simulation based tests have been pointed out, including variance reduction, which can be particularly important when dealing with large time-consuming models typical of combat simulations. The aspiration is that the T&E community will be better able to exploit modern DOE technology to increase greatly the efficiency of tests, and perhaps to render practical what might otherwise be a hopelessly time-consuming course of experimentation, be it live or simulation-based.

## Acknowledgments

Thanks to David Kelton for his careful reading of this annex, and his many helpful suggestions.

## References


Banks, J., J. S. Carson, B. L. Nelson, D. M. Nicol. 2005. Discrete-event Simulation, 4th ed. Prentice-Hall, Upper Saddle River, NJ.

Ben-Gal, I., and J. Bukchin. 2002. The ergonomic design of workstations using virtual manufacturing and response surface methodology. IIE Transactions 34: 375-391.

Bettonvil, B., J. P. C. Kleijnen. 1997. Searching for important factors in simulation models with many factors: sequential bifurcation. European Journal of Operational Research 96 180-194.

Box, G. E. P., and N. R. Draper. 1987. Empirical Model-Building and Response Surfaces. New York: Wiley.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 2005. Statistics for experimenters: An introduction to design, data analysis and model building. 2nd ed. New York: Wiley.

Bratley, P., B.L. Fox, and L.E. Schrage. 1987. A Guide to Simulation, 2nd. ed. Springer-Verlag, New York.

Chang, K.-H., and H. Wan. 2007. Stochastic trust region gradient-free method (STRONG): A new response-surface-based algorithm in simulation optimization. In Proceedings of the 2007 Winter Simulation Conference, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 346-354. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Cioppa, T. M., and T. W. Lucas. 2007. Efficient nearly orthogonal and space-filling Latin hypercubes. Technometrics 49(1): 45-55.

CONDOR 7.1.1. Available via http://www.cs.wisc.edu/condor/ [last accessed June 17, 2008].

Dorfman, R. (1943). The detection of defective numbers of large populations. Annals of Mathematical Statistics 14: 436-440.

Fu, M. 2002. Optimization for simulation: Theory vs. practice. INFORMS Journal on Computing 14 (3): 192-215.

Draper, N. R., and H. Smith. 1980. Applied regression analysis. New York: Wiley.

Friedman, J. H. 1991. Multivariate adaptive regression splines (with discussion). Annals of Statistics 19(1): 1-67.

Hand, D., H. Mannila, and P. Smyth. 2001. Principles of data mining. Cambridge, Massachusetts: MIT Press.

Hernandez, A. S., M. Carlyle, and T. W. Lucas. 2008. Using integer programming to generate flexible nearly orthogonal Latin hypercubes. Working paper, Operations Research Department, Naval Postgraduate School, Monterey, California.

Kelton, W. D., R. P. Sadowski, and D. P. Sturrock. 2007. Simulation with Arena. 4th ed. New York: McGraw-Hill.

Khuri, A. I. 1996. Multiresponse surface methodology. S. Ghosh, C. R. Rao, eds. Handbook of Statistics, Volume 13. Elsevier, Amsterdam, The Netherlands.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. A user's guide to the brave new world of simulation experiments. INFORMS Journal on Computing 17 (3): 263{289 (with online companion).

Law, A. M. 2007. Simulation modeling and analysis. 4th ed. New York: McGraw-Hill.

Leemis, L.M. and S.K. Park. 2006. Discrete-Event Simulation: A First Course. Pearson Prentice Hall, Upper Saddle River, NJ.

Lucas, T. W., S. M. Sanchez, F. Martinez, L. R. Sickinger, and J. W. Roginski. 2007. Defense and homeland security applications of multi-agent simulations. In Proceedings of the 2007 Winter Simulation Conference, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 138-149. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Montgomery, D. C. 2000. Design and analysis of experiments. 5th ed. New York: Wiley.

Myers, R. H., D. C. Montgomery. 2002. Response Surface Methodology: Process and Product Optimization using Designed Experiments. 2nd ed. John Wiley & Sons, New York.

NIST/Sematech. 2005. e-Handbook of statistical methods. Available via <http://www.itl.nist.gov/div898/handbook/> [accessed July 1, 2006].

Osborne, D. M., R. L. Armacost, and J. Pet-Edwards. 1997. State of the art in multiple response surface methodology. 1997 IEEE International Conference on Systems, Man, and Cybernetics, 3833-3838.

Park, K. S., and K.-J. Kim. 2005. Optimizing multi-response surface problems: How to use multi-objective optimization techniques.

Rees, L. P., E. R. Clayton, and B. W. Taylor. 1985. Solving multiple response simulation models using modified response surface methodology within a lexicographic goal programming framework. IIE Transactions 17(1): 47-57.

Ryan, T. P. 2007. Modern experimental design. Hoboken, New Jersey: Wiley.

Sacks, J., W. J. Welch, T. J. Mitchell, H. P. Wynn. 1989. Design and analysis of computer experiments (includes Comments and Rejoinder). Statistical Science 4: 409-435.

Sanchez, S. M. 2000. Robust design: Seeking the best of all possible worlds. In Proceedings of the 2000 Winter Simulation Conference, eds. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 69-76. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Sanchez, S. M and T. W. Lucas. 2002. Exploring the world of agent-based simulations: Simple models, complex analyses. In Proceedings of the 2002 Winter Simulation Conference, eds. E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. Charnes, 116-126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Sanchez, S. M. and P. J. Sanchez. 2005. Very large fractional factorials and central composite designs. ACM Transactions on Modeling and Computer Simulation 15(4): 362-377.

Sanchez, S. M., H. Wan, and T. W. Lucas. 2008. Two-phase screening procedures for simulation experiments. ACM Transactions on Modeling and Computer Simulation, forth-coming.

Santner, T. J., B. J. Williams, W. I. Notz. 2003. The Design and Analysis of Computer Experiments. Springer-Verlag, New York.

Schamburg, J. 2004. A modified response surface methodology for knowledge discovery. Ph.D. dissertation, University of Virginia, Charlottesville, Virginia.

Schamburg, J. and D. Brown. 2004. A generalized multiple response surface methodology for complex computer simulation applications. In Proceedings of the 2004 Winter Simulation Conference, eds. R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 958-966. Available via <http://www.informs-sim.org/wsc04papers> [accessed July 15, 2008].

SEED Center for Data Farming. 2008. Simulation experiments & efficient designs [online]. Available via <http://harvest.nps.edu> [accessed May 15, 2008].

Taguchi, G. 1987. System of experimental design, vols. 1 and 2. White Plains, New York: UNIPUB/Krauss International.

Tan, P.-N., M. Steinbach, and V. Kumar. 2005. Introduction to data mining. Boston, Massachusetts: Addison-Wesley.

Zeigler B. P., K. Praehofer, T. G. Kim. 2000. Theory of Modeling and Simulation. 2nd ed. Academic Press, San Diego, CA.

# ANNEX E – CAPABILITY CROSSWALK

## E.1 Purpose

This annex discusses the process for developing a capability crosswalk that maps critical capability issues (CCI) and critical operational issues (COI) to mission desired effects, tasks, and system/system of systems (SoS) elements to set the stage for developing capability test designs.

## E.2 Overview

The evaluation thread is focused on developing an evaluation strategy for test and evaluation (T&E) of a joint capability. Within Capability Test Methodology (CTM) 1, the capability is described in terms of an SoS and individual systems. CCIs and COIs are identified which include descriptions of the mission, tasks, and key attributes of the SoS and system. A capabilities crosswalk matrix maps these key components into a logical fashion that can be used in the evaluation process. A complete crosswalk will consist of CCIs and COIs.

### E.2.1 Critical Capability Issue (CCI)

CCIs capture the ability to achieve a desired effect under specified standards and conditions through combinations of means and ways to perform a set of tasks. A CCI with joint impact is used to assess performance pertaining to capabilities which support joint missions. The essential elements of a CCI include a capability's essential tasks, mission desired effects, blue SoS across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF), and conditions involving threat and environmental factors. These essential elements are contained in a capability crosswalk. See Figure E-1 for a depiction of the relationships between the CCI and the SoS configuration, tasks, and desired effects.

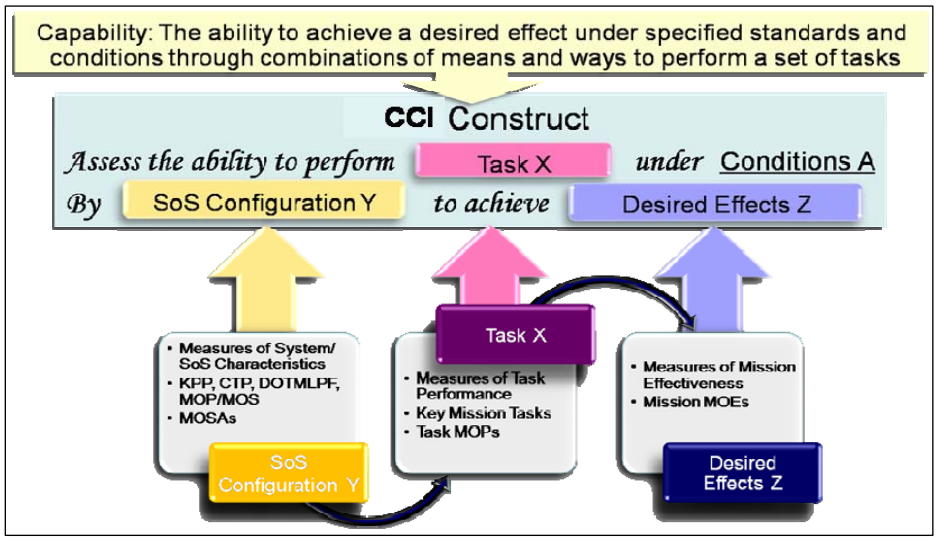



**Figure E-1. CCI to Configuration, Task, and Effects Relationship**

It is important to state how the test issue contributes to achieving the mission end state outcomes in terms of mission desired effects. The CCI should address the SoS capability to perform joint operational tasks and/or the SoS, system, or Service attribute performance. CCIs are of primary importance to the decision authority in reaching a decision to allow the SoS to advance into the next phase of development.

**Example CCI format:**
***Assess the ability to perform Task X by SoS Configuration Y under Condition Set A to achieve Mission Desired Effect Z.***

The condition set is composed of threat and environmental factors that may impact mission effectiveness. A capabilities crosswalk matrix, see Figure E-2, depicts the key elements of the CCI.

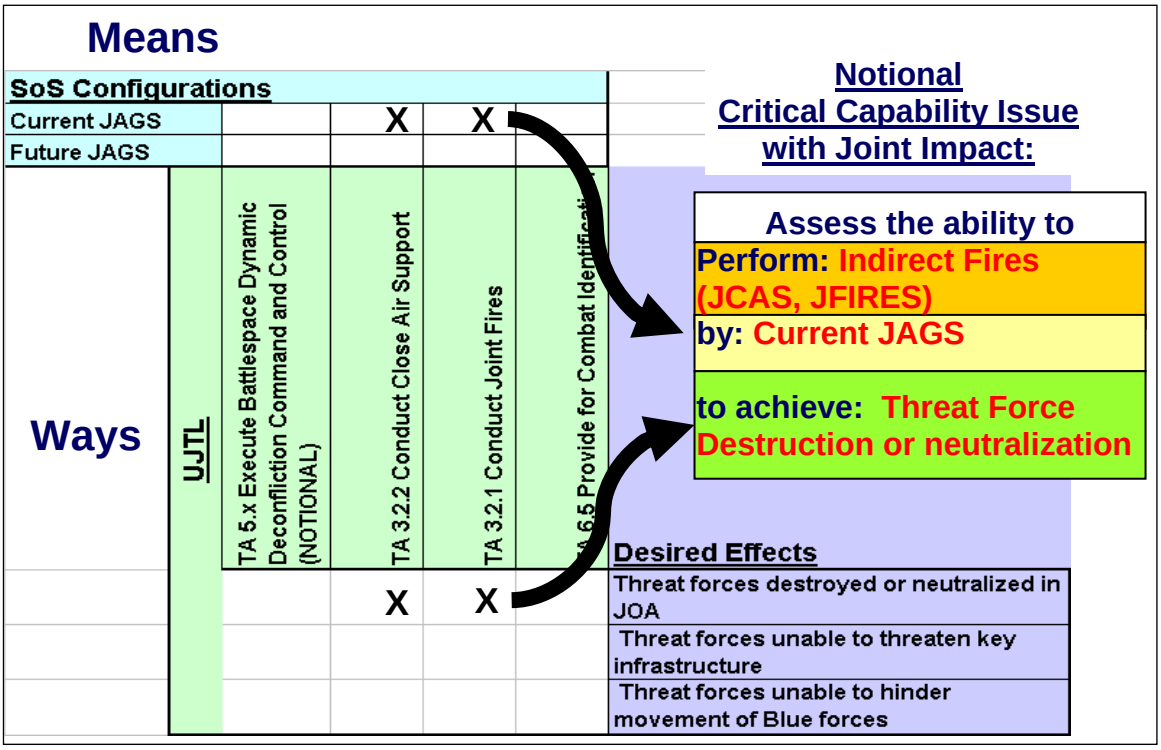



**Figure E-2. Capability Crosswalk Matrix Example**

## E.2.2 Critical Operational Issue (COI)

COIs are the operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined in T&E to evaluate and/or assess the system's capability to perform its mission. These test issues/questions should also identify those tasks (operational and tactical) identified by the ICD and/or CDD associated with the operational tasks that will be performed to accomplish the mission and ensure the intent of test is met. COIs include a description of the operational task threads and task interactions.

Although each Service has its own preferences, a COI is typically phrased as a question that must be answered in order to properly evaluate operational effectiveness (e.g., "Will the system detect the threat in a combat environment at adequate range to allow successful engagement?") and operational suitability (e.g., "Will the system be safe to operate in a combat environment?").[1]

[1] The reader may also consult the following service-level evaluation and analyst guides for more information about COI: AFOTEC (2007), pp 2-8 through 2-10; OPTEVFOR (200 ), p 1-2.

**Example COI:**
***How capable is the X System of making in-flight trajectory corrections?***

Even though not explicit, the COI has certain key components to it. A COI typically identifies a system, a Service task, and implies certain attributes for evaluation.

## E.3 Key Components

The key components of the capability crosswalk are described in the following sections.

### E.3.1 SoS Configurations

An SoS can be made in different configurations. An SoS is a set or arrangement of interdependent systems that are related or connected to provide a given capability. The loss of any part of the system will significantly degrade the performance or capabilities of the whole. Thus to fully describe the capability, the SoS must be decomposed into testable systems that can be integrated into the joint mission environment. SoS attribute areas may also be identified from capability gaps and key characteristics. It is possible to map these attributes to the mission areas across the materiel and non-materiel DOTMLPF elements. Since the SoS attribute areas are required to have associated performance measures, this mapping system ties the individual systems to SoS mission areas and to measurable and testable SoS attributes.

To describe different SoS configurations requires identifying materiel and non-materiel factors that can be varied as a part of the SoS description. Figure E-3 shows an example of ten different SoS configurations that are identified by a set of ten factors. Each factor has two levels identified as level one or level two. With ten factors and two levels each, it is possible to have $2^{10} = 1024$ combinations of factors. Exploratory analysis or other analytical techniques would be required to narrow the focus of the test to a smaller set, such as the ten shown in Figure E-3. These factors make up the independent variables for the decision-maker.

### E.3.2 Conditional Factors

Additional independent variables can consist of conditional factors. Threat or environmental conditions affect achievement of mission desired effects or affect performance of essential tasks. Threat conditions include those threat (e.g., threat actions, threat order of battle, threat command and control structure, threat systems, threat force lay down) condition variables of an operational environment or situation in which a unit, system, or individual is expected to operate and which may affect achievement of mission desired effects or affect performance of essential tasks. Environmental conditions include those physical environment (e.g., land, sea, air, and space) or civil environment (e.g., political, cultural, and economic) condition variables of an operational

| SoS Factors / SoS Configuration | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 | Factor 9 | Factor 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| SoS 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 |
| SoS 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 |
| SoS 3 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 |
| SoS 4 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| SoS 5 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| SoS 6 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| SoS 7 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| SoS 8 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| SoS 9 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 |
| SoS 10 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |

**Figure E-3. SoS Configuration – Factor Mapping**

environment or situation in which a unit, system, or individual is expected to operate and which may affect achievement of mission desired effects or affect performance of essential tasks. Figure E-4 shows an example of ten condition sets identified by five threat factors and six environment factors, each set at a particular level. Similar to identifying SoS configurations, the condition sets must be narrowed to a small subset through exploratory analysis or other analytical techniques.

| Threat Factors / Condition Sets | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Environment Factors / Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 |
| C 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| C 3 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 |
| C 4 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| C 5 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| C 6 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 |
| C 7 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| C 8 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |
| C 9 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 |
| C 10 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |

**Figure E-4. Conditional – Factor Mapping**

### E.3.3 Mission Decomposition

Mission decomposition is described in Chapter 1 of the *Analyst's Handbook for Testing in a Joint Environment* (Analyst's Handbook). Mission components include an overall mission statement, mission objectives and end states, mission desired effects, and mission measures of effectiveness. Key to the evaluation strategy is the later two components. A mission desired effect is the physical or behavioral state of a system that results from an action, a set of actions, or another effect. The desired effect for the SoS may be different than the individual systems that make up the system of systems. However, the desired mission effect should be the result

outcome, or consequence, that the commander desires to achieve by performing a mission, which could be a change to a condition, a behavior, or a degree of freedom change for either blue or threat forces. Mission desired effects are identified and derived from an authoritative source (i.e., Analytic Agenda, etc). Once the mission desired effects are defined as a condition for achieving an associated strategic or operational objective, the Universal Joint Task List (UJTL) is used to determine the necessary tasks to accomplish the mission. The mission measure of effectiveness (Mission MOE) focus is on mission accomplishment and mission utility for the SoS. The MOE should serve as the higher-level measures for the SoS or SUT. Sections 4b and 4d of the ICD should describe the capabilities in terms of desired outcomes, how they complement the integrated joint warfighting force, and indicate the minimum value at which the SUT/SoS capability will no longer be effective. When detailing the Mission MOE for the SoS it is important to trace each specific MOE back to the appropriate desired effect as shown in Figure E-5. When detailing the Mission MOE for the SUT, the measures should be designed to correspond to the accomplishment of the systems mission, desired end state and the systems desired effects.

**Mission: Joint Forcible Entry**

**Mission Statement: On order, Blue forces conduct joint forcible entry operations to expand lodgment and control key infrastructure in order to facilitate rapid force buildup in the Joint Operating Area (JOA).**

| Objectives | End State | Desired Effects | Threat forces destroyed or neutralized in JOA | Threat forces unable to threaten key infrastructure | Threat forces unable to hinder movement of Blue forces | |
|---|---|---|---|---|---|---|
| Expand lodgment and control key infrastructure to facilitate rapid force build-up | Expand lodgment and control key infrastructure | | P | S | | |
| Preserve key infrastructure and prevent environmental contamination | Key infrastructure preserved without environmental contamination | | P | P | | |
| Secure LOCs to facilitate flow of forces needed to conduct regime change operations | LOCs secured | | P | | P | |
| | | | | | | Mission Measures of Effectiveness |
| | | | P | S | S | Threat Systems Combat Ineffectiveness |
| | | | P | S | S | Time to Threat Systems rendered ineffective |
| Legend | | | P | S | S | Cumulative effectiveness time Threat Systems in JOA |
| P: Primary relationships to Desired Effects | | | S | P | P | Blue control of JOA |
| S:Secondary relationships to Desired Effects | | | S | P | S | Key infrastructure vulnerability |
| | | | S | S | P | Blue force vulnerability |

**Figure E-5. Mission Decomposition**

## E.3.4 Task Decomposition

A task is an action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability or a directive statement used to assign a discrete action or set of actions to an organization that enables a mission or function to be accomplished. A single task may incorporate multiple individual actions. Joint tasks describe what tasks a system will be expected to perform to accomplish the mission. These performance tasks enable test personnel to evaluate the proposed systems based on specific requirements. The joint tasks are general statements of task performance requirements. Where possible, the tasks are described in quantitative terms. These tasks become the standards for comparing alternative systems.

Task identification can be logically broken down into several sequential steps:

1. Identification of key events;
2. Primary technical tasks and their entry and exit criteria;
3. Needed support tasks; and
4. Management tasks for accomplishment of primary and support tasks.

Some examples of mission tasks include:

- Conduct deep interdiction precision air strikes against heavily defended targets; or
- Conduct air strike operations in a stealth mode.

The primary authoritative source for joint tasks is the UJTL (CJCSM 3500.04D). This document should be the primary source for determining the key joint tasks required to achieve the mission desired effects, and the standards associated with the tasks. In some cases, particularly for a unique capability, there may be one or more required tasks that cannot be found in the UJTL. In those cases, the analysis should identify those tasks and document their source.

A task measure of performance (Task MOP) is a criterion used to assess friendly actions tied to measuring task accomplishment. Critical parameters selected from the MOP will be key indicators of overall system or SoS performance. If a Task MOP is not met, it may prevent the associated MOE from being satisfied and thus put the program/project at cost, schedule, or performance risk. A Task MOP should be written as a criterion used to assess changes in system behavior, capability, or operational environment. It is a directly measurable parameter that can be specifically tested and measured during the test or evaluation. The MOP is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect.

The initial source of information about a Task MOP and associated criteria is the UJTL. For each joint task, the UJTL identifies measures for describing varying levels of joint task performance and discusses the process for developing criteria that define successful levels of performance. The UJTL considers standards, the combination of measures and criteria, and describes the process for developing these standards as follows:

The standard for a joint task is set within the framework of the JFC's mission and in the context of the conditions, either most likely or worst case, that are linked to those missions. The standard(s) for a joint task can only be set when (1) the mission analysis is complete, (2) the conditions affecting the task have been identified and described, and (3) measures and criteria have been selected that reflect the task contribution to mission accomplishment. This means that standards are tied to missions. If a joint task has a particular standard on one mission that does not mean that the same standard will necessarily apply to other missions. A task standard could be the same across missions, but it could also be different for each mission. (CJCSM 3500.04D CH 1, B-B-1)

While the analysis team can use the measures defined in the UJTL, they must still look at the measures in the context of the mission by considering the mission desired effects and conditions before assigning the criteria or thresholds for assessing task performance. The UJTL includes

additional, valuable discussions about selecting measures and criteria in its Appendix B, Enclosure B. There are a few techniques for determining feasible ranges and levels for Task MOP. One technique is to use the current baseline SoS task performance as a threshold against which to compare the performance of the new capabilities. However, in many cases, these values may be unknown or unobtainable, particularly for unique measures developed specifically for a particular set of capability test events. Another technique is to query key operational stakeholders to determine the expected and desirable ranges. This has the advantage of garnering user acceptance for the measures and thresholds. In some cases, it may be more appropriate not to assign an initial threshold for the measure range, but to use the measure responses to compare different SoS configurations under differing conditions. The measure would assume a comparative role instead of the role as a success/fail criterion for the SoS or capability.

Figure E-6 shows a mapping of UJTLs to Task MOPs. The analyst must document these to use in the capability crosswalk matrix.

| | UJTL | Task Measures of Performance |
|---|---|---|
| ***TA 5.x*** | ***Exercise Battlespace Dynamic Deconfliction C2*** | |
| | | Time to C2 indirect fires missions |
| | | Time to C2 indirect fires missions by task thread |
| | | Time to C2 indirect fire missions when airspace is controlled by requesting Service |
| | | Time to C2 indirect fire missions when airspace is controlled by another Service |
| | | Number of indirect fire missions denied |
| ***TA 3.2.2*** | ***Conduct Joint Close Air Support*** | |
| | | Time to get ordnance on target after initiating air-to-surface attack request |
| | | Percent of requested JCAS missions executed |
| ***TA 3.2.1*** | ***Conduct Joint Fires*** | |
| | | Percent of targets attacked IAW requests for fires |
| | | Time to get ordnance on target after initiating task (Immediate Call For Fire) |

**Figure E-6. Task Decomposition**

### E.3.5 SoS Attributes

SoS performance attributes are significant attributes that must be verified by testing and evaluation/analysis. The CDD and CPD state the operational and support-related performance attributes of a system(s) that provide(s) the capabilities required by the warfighter. Each attribute will be supported by an operationally oriented analysis to determine threshold and objective values. Below the threshold value, the military utility of the system(s) becomes questionable. In an evolutionary acquisition, it is expected that threshold values will generally improve between increments. Different attributes may come into play as follow-on increments deliver additional capability. An attribute may apply to more than one increment. The threshold and objective values of an attribute may differ in each increment. DoD components will, at a minimum, budget to achieve all stated thresholds.

Measures of SoS Attributes (MOSA) are system and SoS attributes, which measure SUT and SoS performance, and include critical test parameters (CTP) and key performance parameters (KPP). These are used to assess SoS and SUT performance and limitations based on specified system requirements and additional characteristics such as reliability, maintainability, human factors, and susceptibility characteristics.

KPPs are those attributes or characteristics of a system that are critical or essential to the development of an effective military capability. There are other types of compatibility and interoperability attributes (e.g., databases, fuel, transportability, ammunition) that might need to be identified for a capability. These statements will guide the acquisition community in making tradeoff decisions between the threshold and objective values of the stated attributes. Operational testing will assess the ability of the system(s) to meet the production threshold values as defined by the KPP and other performance attributes.

Figure E-7 shows a mapping of SoS characteristics to SoS configurations. A mapping of system attributes and attribute measures is not included.

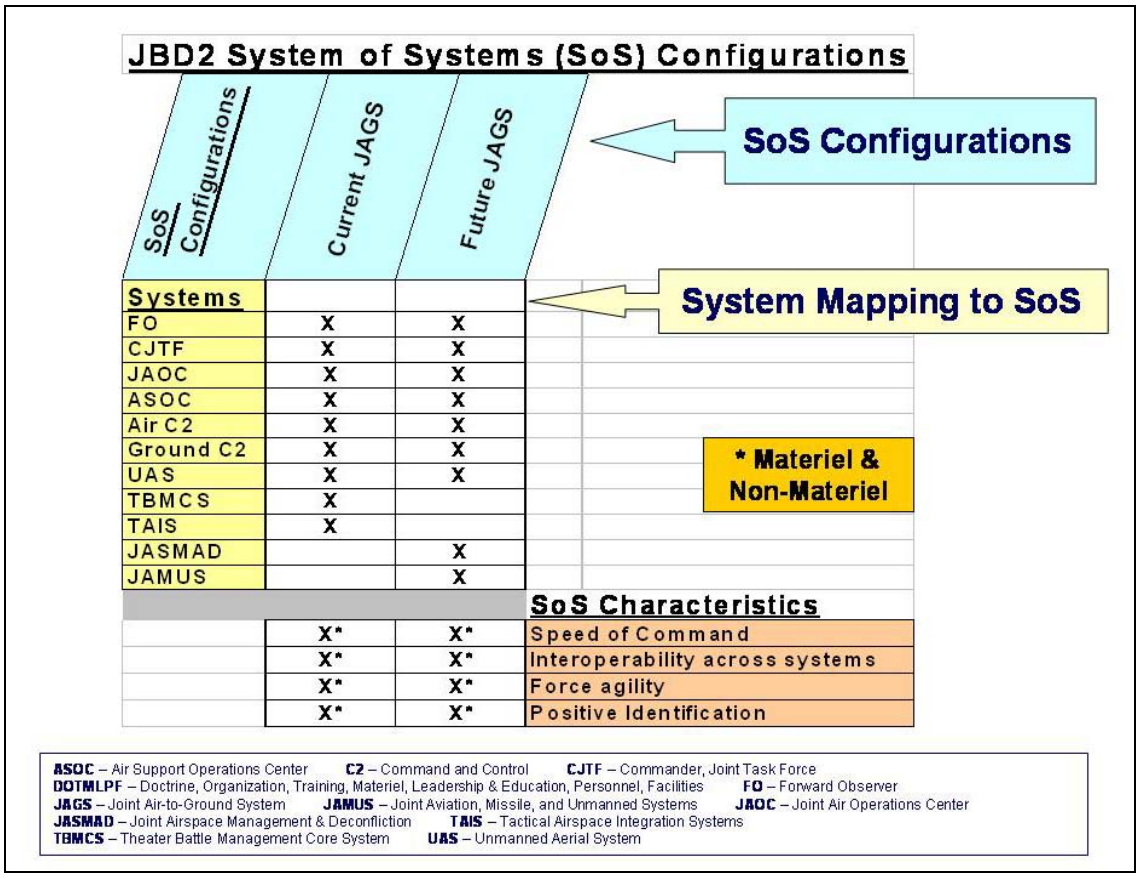



**Figure E-7. SoS Attributes Decomposition**

## E.4 Capability Crosswalk

Once the key components of the capability crosswalk have been developed, the crosswalk can be completed. Figure E-8 provides an example of a capability crosswalk for the CCIs. Each CCI is listed with the applicable components of each. This figure displays the task, the conditions, the SoS configuration and the mission desired effect. The capability factors are included for the system under test, the threat and the environment. The SoS factors and conditional factors are described in the capability crosswalk matrix with their set levels.

**CCI Construct:**

**Assess the ability to perform** **Task X** **under Conditions**

**by** **SoS Configuration Y** **to achieve** **Mission Desired Effect Z**

| | Desired Effects | | | Tasks | | | | | SoS Factors | | | | | | Threat Factors | | | | Environmental Factors | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Critical Capability Issue | Destroyed or neutralized Threat Forces (in-Zone, moving into Zone, indirect fires, ADA) | Blue unhindered in maneuver within the JOA | Blue force survivability in the JOA | TA 5.x Execute Immediate Battlespace Deconfliction C2 | TA 3.2.2 Conduct Close Air Support | TA 3.2.1 Conduct Joint Fires | TA 6.5 Provide for Combat ID | SoS Configuration | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Condition Set | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 |
| CCI-1 | X | | | X | | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-2 | X | | | | X | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-3 | X | | | | | X | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-4 | | X | | X | | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-5 | | X | | | X | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-6 | | X | | | | X | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-7 | | | X | X | | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-8 | | | X | | X | | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-9 | | | X | | | X | | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-10 | | | X | | | | X | SoS 1 | 1 | 1 | 1 | 2 | 2 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-11 | X | | | X | | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-12 | X | | | | X | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-13 | X | | | | | X | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-14 | | X | | X | | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-15 | | X | | | X | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-16 | | X | | | | X | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-17 | | | X | X | | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-18 | | | X | | X | | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| CCI-19 | | | X | | | X | | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| CCI-20 | | | X | | | | X | SoS 2 | 2 | 2 | 2 | 1 | 1 | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |

**Figure E-8. CCI Crosswalk Matrix**

A second COI crosswalk matrix can be created from system, task, and condition factor components. Figure E-9 shows an example of a COI crosswalk matrix.

| | Service Tasks | | | | Systems under test | | | | | | Threat Factors | | | | Environmental Factors | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Critical Operational Issue | Service Task 1 | Service Task 2 | Service Task 3 | Service Task 4 | System 1 | System 2 | System 3 | System 4 | System 5 | Condition Set | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 |
| COI-1 | X | | X | | X | | | | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-2 | | X | | X | X | | | | | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| COI-3 | | | X | | X | | | | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-4 | X | | X | | | X | | | | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| COI-5 | | X | | X | | X | | | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-6 | | | | X | | X | | | | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| COI-7 | X | | | | | | X | | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-8 | | | | X | | | X | | | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| COI-9 | | X | X | | | | | X | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-10 | | | X | | | | | X | | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| COI-11 | | | | X | | | | X | | C-1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| COI-12 | | | | X | | | | | X | C-2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |

**Figure E-9. COI Crosswalk Matrix**

## E.5 Measures Crosswalk

The two crosswalk matrices provide the mapping of mission, tasks, SoS, systems, and conditional factors to test CCIs and COIs. Once a test event has identified its goal, test

objectives, and test issues, a measures crosswalk provides an easy method for identifying applicable test measures. Figure E-10 provides an example of a measures crosswalk to mission, task, SoS, and system components. Figure E-11 provides a crosswalk list of measures to CCIs and COIs.

| | | Desired Effects | | | Joint Tasks | | | | SoS Configurations | | Service Tasks | | | | Systems under test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Measures | Destroyed or neutralized Threat Forces (in-Zone, moving into Zone, indirect fires, ADA) | Blue unhindered in maneuver within the JOA | Blue force survivability in the JOA | TA 5.x Execute Immediate Battlespace Deconfliction C2 | TA 3.2.2 Conduct Close Air Support | TA 3.2.1 Conduct Joint Fires | TA 6.5 Provide for Combat ID | SoS-1 | SoS-2 | Service task 1 | Service task 2 | Service task 3 | Service task 4 | System 1 | System 2 | System 3 | System 4 | System 5 |
| Mission Measures of Effectiveness (MMOE) | Threat Systems Combat Ineffectiveness | X | | | | | | | | | | | | | | | | | |
| | Velocity which Threat Systems rendered ineffective | X | | | | | | | | | | | | | | | | | |
| | Cumulative effectiveness time Threat Systems in JOA | X | X | | | | | | | | | | | | | | | | |
| | Military control of JOA | | X | X | | | | | | | | | | | | | | | |
| | Threat force Situational Awareness | | X | X | | | | | | | | | | | | | | | |
| | Blue force vulnerability | | | X | | | | | | | | | | | | | | | |
| Task Measures of Performance (TMOP) | Time to C2 Indirect Fires | | | | X | X | X | X | | | | | | | | | | | |
| | Time to C2 Indirect Fires by task thread | | | | X | X | X | | | | | | | | | | | | |
| | Time to C2 Indirect Fires when airspace is controlled by requesting Service | | | | X | X | X | | | | | | | | | | | | |
| | Time to C2 Indirect Fireswhen airspace is controlled by another Service | | | | X | X | X | | | | | | | | | | | | |
| | Number of Indirect Fires denied | | | | | X | X | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating air-to-surface attack request | | | | | X | | | | | | | | | | | | | |
| | Percent of requested missions executed | | | | | X | X | | | | | | | | | | | | |
| | Number targets attacked IAW requests for fires | | | | | X | X | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating task | | | | X | X | X | | | | | | | | | | | | |
| Measures of SoS Attributes (MOSA) | Speed of Decision (C2 time minus the Deconfliction time) | | | | | | | | X | X | | | | | | | | | |
| | Speed of Deconfliction (Airspace cleared or assigned box) | | | | | | | | X | X | | | | | | | | | |
| | Force Simultaneity (Simultaneous C2) | | | | | | | | X | X | | | | | | | | | |
| Measures of Effectiveness (MOE) | Measure of Effectiveness 1 | | | | | | | | | | X | | | X | | | | | |
| | Measure of Effectiveness 2 | | | | | | | | | | X | X | | | | | | | |
| | Measure of Effectiveness 3 | | | | | | | | | | | X | X | | | | | | |
| | Measure of Effectiveness 4 | | | | | | | | | | | X | | X | | | | | |
| Measures of Suitability (MOS) | Measure of Suitability 1 | | | | | | | | | | | | | | | X | | | X |
| | Measure of Suitability 2 | | | | | | | | | | | | | | X | | X | X | |
| | Measure of Suitability 3 | | | | | | | | | | | | | | | X | X | | X |
| | Measure of Suitability 4 | | | | | | | | | | | | | | X | | | X | X |

**Figure E-10. Measures Crosswalk Matrix**

| | | Critical Capability Issues | | | | | | | | | | | | | | | | | | | | Critical Operational Issues | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Measures | CCI-1 | CCI-2 | CCI-3 | CCI-4 | CCI-5 | CCI-6 | CCI-7 | CCI-8 | CCI-9 | CCI-10 | CCI-11 | CCI-12 | CCI-13 | CCI-14 | CCI-15 | CCI-16 | CCI-17 | CCI-18 | CCI-19 | CCI-20 | COI-1 | COI-2 | COI-3 | COI-4 | COI-5 | COI-6 | COI-7 | COI-8 | COI-9 | COI-10 | COI-11 | COI-12 |
| Mission Measures of Effectiveness (MMOE) | Threat Systems Combat Ineffectiveness | X | X | X | | | | | | | | X | X | X | | | | | | | | | | | | | | | | | | | |
| | Velocity which Threat Systems rendered ineffective | X | X | X | | | | | | | | X | X | X | | | | | | | | | | | | | | | | | | | |
| | Cumulative effectiveness time Threat Systems in JOA | X | X | X | X | X | X | | | | | X | X | X | X | X | X | | | | | | | | | | | | | | | | |
| | Military control of JOA | | | | X | X | X | X | X | X | X | | | | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Threat force Situational Awareness | | | | X | X | X | X | X | X | X | | | | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Blue force vulnerability | | | | | | | X | X | X | X | | | | | | | X | X | X | X | | | | | | | | | | | | |
| Task Measures of Performance (TMOP) | Time to C2 Indirect Fires | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Time to C2 Indirect Fires by task thread | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Time to C2 Indirect Fires when airspace is controlled by requesting Service | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Time to C2 Indirect Fireswhen airspace is controlled by another Service | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Number of Indirect Fires denied | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating air-to-surface attack request | | X | | | X | | | X | | | | X | | | X | | | X | | | | | | | | | | | | | | |
| | Percent of requested missions executed | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Number targets attacked IAW requests for fires | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating task | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| Measures of SoS Attributes (MOSA) | Speed of Decision (C2 time minus the Deconfliction time) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Speed of Deconfliction (Airspace cleared or assigned box) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Force Simultaneity (Simultaneous C2) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| Measures of Effectiveness (MOE) | Measure of Effectiveness 1 | | | | | | | | | | | | | | | | | | | | | X | X | | X | X | X | X | X | | | X | X |
| | Measure of Effectiveness 2 | | | | | | | | | | | | | | | | | | | | | X | X | | X | X | | X | | X | | | |
| | Measure of Effectiveness 3 | | | | | | | | | | | | | | | | | | | | | X | X | X | X | X | | | | X | X | | |
| | Measure of Effectiveness 4 | | | | | | | | | | | | | | | | | | | | | | X | | | X | X | | X | X | | X | X |
| Measures of Suitability (MOS) | Measure of Suitability 1 | | | | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | | | X |
| | Measure of Suitability 2 | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | X | X | X | X | X | |
| | Measure of Suitability 3 | | | | | | | | | | | | | | | | | | | | | | | | X | X | X | X | X | | | | X |
| | Measure of Suitability 4 | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | | | X | X | X | X |

**Figure E-11. Measures Crosswalk List**

*This page intentionally left blank.*

## ANNEX F - INTEGRATED DATA REQUIREMENTS LIST (IDRL)

### F.1. Introduction

An Integrated Data Requirements List (IDRL) is developed as part of a dendritic analysis process to decompose measures to data elements and data collection requirements. This annex will focus on the development of an IDRL and its use in a joint capability evaluation.

### F.2. Discussion

The dendritic analysis process is a methodology to decompose critical test issues to the point where actual data requirements and test measurements can be identified. The dendritic analysis process is the standard test planning technique used to develop the test concept and identify the test resources required to execute the proposed test project. The development of test issues and measures is one of the most important and difficult tasks that the test team must perform during test planning activity. This requires the test team to have a detailed knowledge of the related system of systems (SoS), processes, joint tactics, techniques, and procedures (TTP), doctrine and Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) architectures. In this process, test issues are successively broken down into mission measures of effectiveness (Mission MOE), task measures of performance (Task MOP), measures of SoS attributes (MOSA), and system measures of effectiveness and suitability (MOE and MOS). Measures may be mapped to the issues as part of the capability crosswalk development as described in Annex E. Figure F-1 shows an example of a mapping of issues to measures.

| | | Critical Capability Issues | | | | | | | | | | | | | | | | | | | | Critical Operational Issues | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Measures | CCI-1 | CCI-2 | CCI-3 | CCI-4 | CCI-5 | CCI-6 | CCI-7 | CCI-8 | CCI-9 | CCI-10 | CCI-11 | CCI-12 | CCI-13 | CCI-14 | CCI-15 | CCI-16 | CCI-17 | CCI-18 | CCI-19 | CCI-20 | COI-1 | COI-2 | COI-3 | COI-4 | COI-5 | COI-6 | COI-7 | COI-8 | COI-9 | COI-10 | COI-11 | COI-12 |
| Mission Measures of Effectiveness (MMOE) | Threat Systems Combat Ineffectiveness | X | X | X | | | | | | | | X | X | X | | | | | | | | | | | | | | | | | | | |
| | Velocity which Threat Systems rendered ineffective | X | X | X | | | | | | | | X | X | X | | | | | | | | | | | | | | | | | | | |
| | Cumulative effectiveness time Threat Systems in JOA | X | X | X | X | X | X | | | | | X | X | X | X | X | X | | | | | | | | | | | | | | | | |
| | Military control of JOA | | | | X | X | X | X | X | X | X | | | | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Threat force Situational Awareness | | | | X | X | X | X | X | X | X | | | | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Blue force vulnerability | | | | | | | X | X | X | X | | | | | | | X | X | X | X | | | | | | | | | | | | |
| Task Measures of Performance (TMOP) | Time to C2 Indirect Fires | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Time to C2 Indirect Fires by task thread | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Time to C2 Indirect Fires when airspace is controlled by requesting Service | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Time to C2 Indirect Fireswhen airspace is controlled by another Service | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| | Number of Indirect Fires denied | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating air-to-surface attack request | | X | | | X | | | X | | | | X | | | X | | | X | | | | | | | | | | | | | | |
| | Percent of requested missions executed | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Number targets attacked IAW requests for fires | | X | X | | X | X | | X | X | | | X | X | | X | X | | X | X | | | | | | | | | | | | | |
| | Time to get ordnance on target after initiating task | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | | |
| Measures of SoS Attributes (MOSA) | Speed of Decision (C2 time minus the Deconfliction time) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Speed of Deconfliction (Airspace cleared or assigned box) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| | Force Simultaneity (Simultaneous C2) | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | | | | | | | | | |
| Measures of Effectiveness (MOE) | Measure of Effectiveness 1 | | | | | | | | | | | | | | | | | | | | | X | X | | X | X | X | X | X | | | X | X |
| | Measure of Effectiveness 2 | | | | | | | | | | | | | | | | | | | | | X | X | | X | X | | X | | X | | | |
| | Measure of Effectiveness 3 | | | | | | | | | | | | | | | | | | | | | X | X | X | X | X | | | | X | X | | |
| | Measure of Effectiveness 4 | | | | | | | | | | | | | | | | | | | | | | X | | | X | X | | X | X | | X | X |
| Measures of Suitability (MOS) | Measure of Suitability 1 | | | | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | | | X |
| | Measure of Suitability 2 | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | X | X | X | X | X | |
| | Measure of Suitability 3 | | | | | | | | | | | | | | | | | | | | | | | | X | X | X | X | X | | | | X |
| | Measure of Suitability 4 | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | | | X | X | X | X |

**Figure F-1. Issue – Measure Crosswalk Matrix**

The IDRL is a product of the dendritic analysis process and provides the means to identify and display the required data elements to be collected during execution of test events. It also ties the data elements to the issues, test measures, any pre-test acceptance criteria, data collection instrumentation required, sample sizes required for statistical rigor, data structures and formats,

source of the data, data collection media, scenarios, method of collection (e.g., automated, manual form, etc.), and other information that supports data collection requirements. There should be an IDRL to support collection of data for all events; additionally the IDRL will assist in the final costing of the test event. Factors that impact final cost include test instrumentation requirements, sample sizes that dictate number of test trials, e.g., aircraft sorties, range time, etc., that must be accomplished, locations of each data collection site, number of data collectors, etc. Since the IDRL is also connected to the evaluation strategy, any changes to either the evaluation strategy or IDRL must result in changes to the other.

The IDRL provides the issues, measures, acceptance criteria, data requirements (elements), test conditions, sample sizes required for statistical rigor, data structures and formats, source of the data, data collection media, scenarios, method of collection, and required test resources such as data collection instrumentation. Most importantly, it facilitates traceability of data and data collection requirements to the measures, critical capability issues and critical operational issues, and provides a summarized format for logically planning and scoping the test events. Any potential impact on the test due to test limitations such as scheduling, technical constraints, or cost constraints becomes more readily apparent. The IDRL guides development of an executable test concept that provides information that answers the what, who, where, and how aspects of a test plan execution. The IDRL also drives the when component of test planning since it identifies items that may have long lead-time coordination requirements.

## F.3. Process for Developing the IDRL

The process to develop an IDRL is not complex given a robust evaluation strategy and crosswalk matrix has already been constructed. The data collection and IDRL must support the test goal and objectives. Therefore the relevant critical capability issues (CCI) and critical operational issues (COI) in the evaluation strategy and crosswalk matrix must be identified for the test event.

**(1) Inputs.** The following inputs are required for the analysis team to begin the IDRL development:

- **Evaluation strategy.** The evaluation strategy must specify the CCI and COIs decomposed into measures and data elements.

- **Crosswalk matrix.** The crosswalk matrix must show the mapping of measures to issues and data elements.

- **JOC-T.** The analysis team must have a thorough understanding of the joint operational context for test (JOC-T) to include relevant scenarios and vignettes, trace of operational mission threads, and threat and environmental conditions. This will lead to the development of the test concept that will be the basis for conducting test planning and executing the test events.

- **Evaluation views.** An evaluation view for each measure will have identified data elements needed to evaluate the measure. If not already constructed, the analysis team may need to draft evaluation views in order to develop the IDRL. Example evaluation views are provided in annex J to this handbook.

**(2) <u>Actions</u>.** The following actions are required for the analysis team to develop the IDRL:

- **Construct the IDRL data reduction sheet.** The analyst should gather all the required information to construct the IDRL data reduction sheet. See Figure F-2 for an example.

  - **Identify relevant issues.** The analyst shall determine what CCIs and COIs will be investigated during the test. This will form the basis for understanding relevant dependent and independent variables for the test; measures as dependent variables and test factors as independent variables.

  - **Identify relevant measures.** The analyst shall gather from the crosswalk matrix mapped issues to measures. If not complete, the analyst shall have to create the issues to measures mapping for the test event.

  - **Identify measures criteria.** The analyst shall specify the evaluation test criteria. Criteria are usually quantitative values or statements of the required performance or effectiveness of the SoS or concept being tested and evaluated. Criteria are normally expressed as thresholds, standards, or numerical ranges and should be unambiguous. Normally the criteria must be established by the key stakeholders. If criteria do not exist, the analyst must work with the stakeholders to develop quantitative and qualitative measurable values.

  - **Identify data elements.** The analyst must identify data elements from the evaluation views. If identified data elements do not exist, then the analyst must develop evaluation views that include required data elements. The specification of the data elements derives directly from the variables that compose the measure mathematical equations, and for that reason, is the simplest of the dendritic analysis processes.

  - **Identify data fields.** The analyst must identify additional data fields that would support the collection of data elements. Data fields are meta-data or supplemental data that may be needed to associate data elements. For example, if you are trying to collect the start and stop time of a mission thread, you would need a data field that identifies that mission thread. Therefore, three data fields are required in the example: start time, end time, and mission thread identification number. The analyst must work with test planners and systems engineers to determine what data fields are needed and can be collected.

- **Construct the IDRL data requirements sheet.** The analyst should develop all required information to construct the IDRL data requirements sheet. See Figure F-3 for an example.

  - **Identify data field descriptions.** The analyst shall describe each data field in as much detail as possible in order to eliminate ambiguity and data collection errors. Data field descriptions should include expected values, required units of measure, required data format, and any other amplifying information that may be of importance.

- **Identify test resources.** The analyst shall work with test planners and system engineers to begin identification of test resources. This shall include data collection instrumentation, type of military systems and personnel required to conduct the test event, test ranges, modeling and simulation facilities, operational scenario, Service and joint battle laboratories, test and training ranges, and communications links. Information that is captured in the IDRL data requirements sheet includes the collection rate, data sources, data media, identified instrumentation, and output file structure. Additional information may be captured as needed.

**(3) Products.** The following products should be derived from the IDRL development:

- **IDRL data reduction sheet.** This is the basic IDRL that contains the measures decomposition to data elements and identified data fields. The IDRL data reduction sheet contains the Mission MOE, Task MOP, MOSA, MOE, MOP and MOS necessary for properly evaluating the system or systems of systems under test. For ease of display the information is captured in a spreadsheet. Items of information which support the analyst and the data collector in obtaining measures are listed as column headings, and the measures are listed in the left hand column of the IDRL spreadsheet. Most IDRL spreadsheets contain all the measure levels (Mission MOE, Task MOP, MOSA, etc.) in a single document. Some additional data fields in an IDRL might include; the measure, the reference source for measures such as the UJTL, the form the data element is expected to be in upon collection and expected value ranges. There many ways to build an IDRL, however these elements are generally common IDRL features. These spreadsheets serve as a basis for the analyst to display the logical method of decomposing the problem to be solved. Additionally, the IDRL spreadsheet serves as a guide for the data collector to obtain the information from the test which is needed to address the IDRL measure requirements. See Figure F-2 for an example IDRL data reduction sheet.

| Measure | CCI COI | Scale Units | Measure | Parent Relation | Criteria | Data Element | Data Fields |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| MMOE-1 | CCI-1 | Percent | Threat Systems Combat Ineffectiveness | Mission Desired Effect of neutralized forces | 80% | | |
| | | Number | | | | Count threat units in JOA at test run start time | For each test run: (Test Trial #, Test run #, Threat Unit Count at Start) |
| | | Number | | | | Count threat units ineffective in JOA at test run end time | For each test run: (Test Trial #, Test run #, Count Threat Units Ineffective at end time) |
| | | | | | | | |
| TMOP-1 | CCI-1 | Time | To C2 Indirect Fires (JCAS. JFIRES) missions | UJTL - TA 5.x Execute Immediate Deconfliction | 10 min | | |
| | | Time | | | | Start time when Indirect Fires (IF) request enters C2 (Decide & Deconflict - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time) |
| | | Time | | | | End time when Indirect Fires (IF) request exits C2 (Decide & Deconflcit - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time) |
| | | | | | | | |
| MOSA-1 | CCI-2 | Time | Speed of Decision (C2 time minus the Deconfliction time) | Indirect Fires C2 SoS | 5 min | | |
| | | Time | | | | Start time when IF request accepted (Accepted C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time) |
| | | Time | | | | End time when IF request exits request evaluated (Evaluate C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide end time) |
| | | | | | | | |
| MOE | COI-1 | Percent | Detection at max range | NLOS-LS | 90% | | |
| | | Number | | | | Count of detection attempts | For each test run: (Test Trial #, Test run #, Weapon ID #, Detection #) |
| | | Number | | | | Count of successful detections | For each test run: (Test Trial #, Test run #, Weapon ID #, Detection #, Detection status) |
| | | | | | | | |
| MOS | COI-1 | Percent | Weapon reliability | NLOS-LS | 99% | | |
| | | Number | | | | Count of weapon fires | For each test run: (Test Trial #, Test run #, Weapon ID #, Fires #, Fires Status) |
| | | Number | | | | Count of weapon mis-fires | For each test run: (Test Trial #, Test run #, Weapon ID #, Fires #, Fires Status) |

**Figure F-2. Example IDRL Data Reduction Sheet**

- **IDRL data requirements sheet.** This spreadsheet is used to describe in greater detail the data fields and test resources required to collect the data. Figure F-3 shows an example IDRL data requirements sheet. The "Data Fields" column provides information on information tracking or meta-data. "Possible Values" column provides the range of possible values for that measure. "Amplifying Information" as its name suggests provides additional information which can be very useful for those data collectors who have little time to become acquainted with the overall test methodology prior to the start of the event. The "Data Structure" and "Data Format" columns provide the format the information needs to be in for further processing. The "Test Run" column shows which run or runs that data element needs to be collected by the data collector. "Data Source" column would provide a list of sources for that data element such as observation, survey, interview, Digital Collection, Analysis, and Review System (DCARS), Observation Tool, etc. The "Data Media" column would list such things as audio tape, written survey, electronic file, data collection form, etc. "Instrumentation" column lists the various devices used to collect the data. Finally, "Data File Structure" lists the format that the data needs to be in for further analytical processing.

| Data Element | Data Fields | Possible Values | Amplifying information | Data Structure | Data Format | Collection Rate | Test Run | Data Source | Data Media | Instru-mentation | Output File Structure |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Test Trial | | | | | | | | | | | |
| | Test Trial ID # | 1 to 4 | Identifies the Test Trial | Integer | 1,2,3, ... | Before test event | All | Controller | Manual | None | |
| | | | | | | | | | | | |
| Test Run | | | | | | | | | | | |
| | Test Run ID # | 1 to 4 | Identifies the Test | Integer | 1,2,3, ... | Each Test Run | All | Controller | Manual | None | |
| | Test Run Start time | 0 | Test run start at time 1 | minutes | xxx.xx | Each Test run | All | RTTC | Electronic | DCO/IWS Trigger | |
| | Threat Count at Start Time | | Count of threat units in the JOA at test run start time | Integer | 1,2,3, ... | Each Test run | All | RTTC | Electronic | Prestart File | |
| | | | | | | | | | | | |
| Indirect Fires Request | | | | | | | | | | | |
| | ID Number | 1,2,3, ... | Sequential ID number for each indirect fires (JCAS/JFIRES request | Integer | 1,2,3, ... | At creation of request | All | RPWS | Electronic message | VMF K02.05 | |
| | Type | JCAS, JFIRES, CCA | Type of IF | Text | JC, JF, CC | At creation of request | All | RPWS | Electronic message | VMF K02.06 | |
| | Threat Detection Start Time | | Time threat is detected and | minutes | xxx.xx | At start of detection box | All | RPWS | Electronic message | VMF K02.07 | |

**Figure F-3. Example IDRL Data Requirements Sheet**

# ANNEX G - ANALYSIS TECHNIQUES

## G.1 Purpose

This annex discusses the tools and techniques for evaluating the capability in Capability Test Methodology (CTM) Step 6. It focuses on the steps to analyze the data with respect to the test objectives; and to evaluate system/SoS performance, task performance, and contributions to joint mission effectiveness.

## G.2 Overview

The last process in the CTM is the evaluation of the capability. Figure G-1 shows the flow diagram for its sub-processes. Although the sub-processes are shown in parallel in the figure, there is a natural serial order (that follows the process numbering) that normally occurs. However, the steps can be, and often are, executed in parallel. Subsets of data can be analyzed as other data are still being processed. Insights can be drawn from subsets of data that have been analyzed, while other data are still being processed and analyzed. Furthermore, analyst discoveries during data analysis may identify additional data processing requirements to format data differently or to process other data that may have been overlooked. The same is true of the evaluation step. In trying to evaluate a system, the team may identify additional analyses that may be required to complete the evaluation. It is expected that all of these processes will occur in parallel.

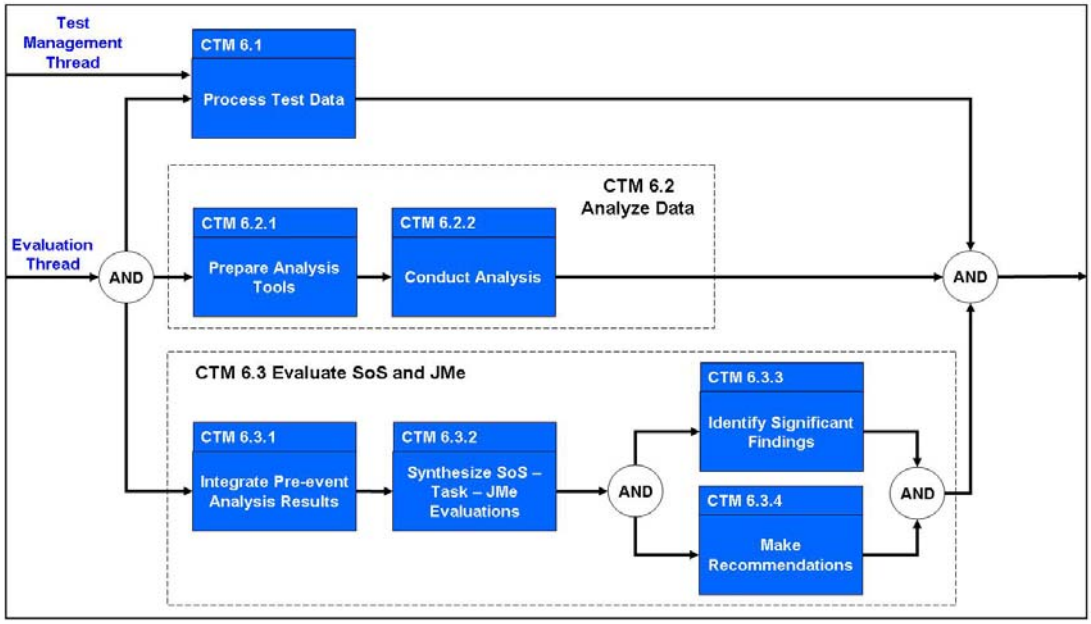



**Figure G-1. CTM 6, Evaluate Capability**

The ultimate output of this process is an objective evaluation of the SoS or SUT in terms of its contribution to delivering the intended capability and contributions to joint mission effectiveness. All the previous processes lead to this ultimate goal. It is this evaluation which is used by decision-makers to determine the future of those systems: to field them as is, to return them to product managers (PM) and developers for further work, or to cancel those programs in favor of

others that are more promising. The PM should select systems that not only meet the KPPs and KSAs as outlined in JCIDS documents (CDD and CPD), but also demonstrate a beneficial and cost-effective impact on mission effectiveness in the SoS's operational context. This can be a very politically-charged evolution, so the analyst is cautioned to exercise objectivity and be aware of ethical implications.

Returning to the Capability Evaluation Metamodel (CEM) shown in figure G-2, the ***CTM 6, Evaluate Capability***, axis produces the Joint Capability Evaluation (JCE). The evaluation uses the response data from the test event to provide SoS recommendations based on the analysis design produced by the capability test design. It is through the connection of achieving the mission desired effect that the ultimate utility of any SoS is evaluated.

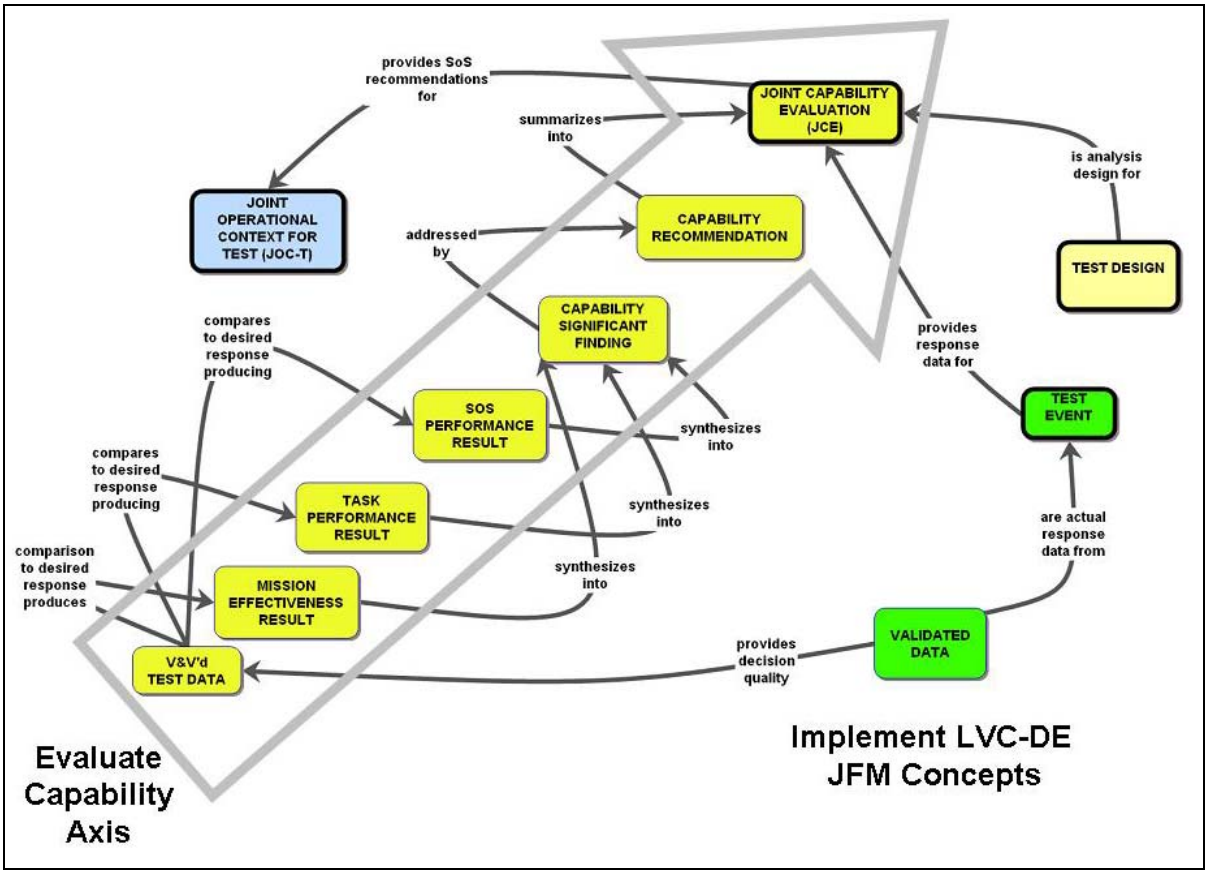



**Figure G-2. CEM Representation of CTM 6, Evaluate Capability**

***CTM 6.1, Process Test Data***, includes all data handling from the source to input into the analysis activity. Data reduction is the process by which the analysis team manipulates the data set to prepare it for subsequent analyses. The data available to the team will likely have been collected from disparate sources, such as multiple models and simulations, as well as questionnaires, interviews, expert panels, etc. The data will be of varying quality and in different forms and may require further reduction and reconstruction to be useful for analysis. Thus, the goal of this process is to get the data into a consistent and usable form.

***CTM 6.2, Analyze Data***, includes preparing analysis tools and conducting analysis. Analysis is the mechanical manipulation (hand or computer) and examination of actual data. Qualitative and quantitative data are analyzed to determine SoS and system under test (SUT) attribute performance measures, SoS task performance measures and JMe measures under various test trials.

In ***CTM 6.3, Evaluate System of Systems and Joint Mission Effectiveness***, the joint mission evaluator uses test results to evaluate the overall JMe and the contribution an individual system under test makes to the accomplishment of the joint mission. Processes include integrate exploratory analysis results; synthesize SoS, task, and mission effectiveness evaluations; identify significant findings; and make recommendations.

## G.3 Analyze Data

This step provides an initial analysis of the test results. It involves using tools such as statistical software, database software, and other related tools required to reduce and analyze the data. More importantly, it requires the analysts to draw general conclusions on SoS performance, task performance, and JMe.

### G.3.1 Tool Selection

First, the analysis team must identify and select the analysis tools at their disposal, based upon their ability to support the evaluation with regard to assessing the test data and drawing conclusions concerning mission effectiveness. An equally important problem is the availability of knowledgeable personnel. The problem is similar to the importance of a non-materiel component where the tools are the weapons and the analysts are the soldiers. The application, however, is slightly different; the tools that might have been appropriate for the exploratory analysis may not be ideal for the post-test analysis. For instance, in exploratory analysis, much of the data are developed via modeling and simulation (M&S); whereas, in this process, the data are coming from an entirely different venue. Different decision makers receiving the results of the post-test analysis may have varied preferences for the types of analytical products they receive. Thus the team must reassess the analytical tools they will need during this process. Because of the similarity of this step to previously-discussed processes, much of this discussion will refer the reader back to the exploratory analysis annex in the handbook for more detail. We will explicitly point out again the analysis team should go through a deliberate process of selecting appropriate tools instead of defaulting to what they have on hand or normally use. This will ensure that the analysts have the tools they need for the task, instead of forcing their analysis to fit the capabilities of the tools they have.

Possible tools and techniques for analysis include:

- Alternative identification, evaluation, and comparison techniques.
- Statistical analysis tools.
- Database tools.
- Statistical techniques.
  - Box plots.
  - Hypothesis testing.
  - Confidence intervals.

  - Regression.
  - ANOVA.
  - Correlation and cross-correlation matrices.
  - Classification and regression trees.
  - Contour plots.
  - Response-surface methods.
  - Advanced response-surface methodology.
  - Nonparametric statistics.
- Qualitative measure relationship techniques.
- Matrix Failure Modes and Effects Analysis (MFMEA).
- Decision analysis techniques, such as multi-attribute decision making (MADM).

#### G.3.1.1 Data Analysis Tool Requirements

Analysis tool requirements will be identified as early as possible. Many of the requirements may have been captured in ***CTM 1*** as part of the exploratory analysis, since similar analytical methods will apply to both processes; however, as mentioned previously, there are likely to be unique requirements during this phase of the testing lifecycle that should be considered. The requirements will depend upon the questions that must be answered in the analysis and the types of data collected in the actual test environment. This step should consider not only the typical statistical-analysis requirements, but also the need to analyze qualitative and observational data elements, if collected. Although the test data may have already been reduced prior to this step, the analysis will likely require additional reduction and organization of the data. Such requirements should also be identified here.

The development of requirements should begin with an identification of the analytical techniques to be used. It is unlikely that the team will be able to identify all required analytical techniques up front, since particular analysis results may suggest subsequent techniques to be used. However, the team should do their best to forecast their analytical needs based upon the questions to be answered. Once the analytical techniques are identified, these begin to define the requirements for the tools that will be used. As before, it is important to note that the use of the term "analytical tools" here denotes statistical and database software, as well as other more-specialized tools.

#### G.3.1.2 Data Analysis Tool Identification

Based upon the requirements identified in the previous step, potential tools and analytical techniques should be identified. Tools refer to the actual means for accomplishing the analyses, and might include particular statistical software packages and databases, as well as other software with more specialized purposes, such as decision-analysis packages. Techniques deal with the methods that will be used in the analyses, such as traditional statistics, data mining, and decision analysis. The techniques will have already been partially considered in the development of analysis tool requirements, but, in this step, the techniques will be further developed in the context of the tools that have been identified. A complete list of common general statistics packages can be found on-line.

#### G.3.1.3 Data Analysis Tool Assessment

The assessment process provides the analyst measurable terms by which to evaluate the suite of potential tools. The capabilities of the potential tools and techniques are compared against the requirements to assess their ability to fulfill the analysis needs. Both strengths and weaknesses are identified and captured, and the tools and techniques are compared. The assessment and evaluation should consider not only their analytical capabilities but also considerations such as availability, user familiarity and experience, and cost. Table G-1 shows a simple example tool-comparison matrix that ranks the importance of each of the criteria by number of checkmarks, and scores each of the tools against each criterion using "+" signs, with more being better for both the checkmarks and the "+" signs. The total score for each tool is calculated by multiplying the number of checkmarks for a particular criterion by the number of "+" signs, and then summing across all criteria. This is a very simple example of a multi-attribute decision-making technique. More complex techniques exist; however, for the purposes of this comparison, simple techniques are recommended.

**Table G-1. Example Tool Comparison Matrix**

| Criteria | Weight | Tool #1 | Tool #2 | Tool #3 |
|---|---|---|---|---|
| Factor representation | ✓✓✓ | +++ | + | + |
| Achievable measures | ✓✓✓ | ++ | +++ | ++ |
| Data output | ✓✓ | + | ++ | +++ |
| Ease of use | ✓ | +++ | + | ++ |
| Cost | ✓ | + | ++ | + |
| Availability | ✓ | ++ | + | +++ |
| **Total Score** | | **23** | **20** | **21** |

#### G.3.1.4 Data Analysis Tool Selection

The analyst should use the tool and technique assessments developed in the previous step to compare the potential tools. As previously mentioned, it is important for the analyst to note that some tools may be included even if they score poorly in a comparison. For example, consider a team examining a complex command and control (C2) capability. They may consider a tool that models the C2 network well, but does not represent other key factors and scores poorly for the other criteria. However, if that model is the only one of those under consideration capable of representing the C2 network, the team may choose to include that model anyway, even if it scores poorly overall. Thus, the analyst must also ensure coverage of all the factors and measures by the suite of tools, not the capabilities of each tool in isolation. A table like that shown above is good for getting an initial cut on the relative ranking of the models, but additional evaluation is required to select the most appropriate set of tools and techniques to be used. The next step will capture that final mapping of factors and measures to the tools being used.

### G.3.2 Conducting Analysis

This process is a key step in the final evaluation. It is here that the team examines and analyzes the data in order to draw the insights necessary to achieve the primary goal of the test data analysis – to evaluate the individual test measures responses and their relationships.

In this step, the team will reduce and organize the data, as well as evaluate the individual measure responses and their relationships.

### G.3.2.1 Data Reduction

The identification of data-collection requirements earlier in the CTM should have affected how the data were collected during the actual test event. Nonetheless, the form of the data actually delivered to the team may not be sufficient. The data, then, may require further reduction and reconstruction to be useful for analysis. In any case, the reduced data will have to be organized into appropriate databases and statistical software packages to facilitate the follow-on analysis.

Processed data will be organized, reduced, prepared for analysis, and organized into tables capable of supporting analysis. Data should be able to be combined, selected, sorted, summarized, counted, and binned relative to any test condition. This will enable easy reconstruction of test conditions that generated any SUT or mission data for selective evaluation of any group of parameters. Groupings could be for an individual SUT and mission data, or combinations of systems and mission data within a test event or across test events. Data will be grouped, sorted, and summarized by a multitude of detailed search parameters to determine under what test conditions the situation most frequently occurred. This approach provides maximum flexibility and allows for multiple levels of analysis, as required. The output from this step is reconstructed mission level data and system level data.

### G.3.2.2 Test Measure Response Analysis

In these steps, the analysis team analyzes the individual responses for each measure at each of the three levels (mission measure of effectiveness [Mission MOE], task measure of performance [Task MOP], and Measure of System/SoS Attributes [MOSA]). For each of these measures, the analyst will explore the response region to identify key findings. Such an analysis will include an examination of the impact of the factor levels, and their interactions, on the measures. This assumes, of course, that a valid experimental design was used during the test, which may not always be possible due to resource constraints or unexpected events during testing. In cases during which such a design was not possible, statistical results and subsequent insights can still be drawn, but will likely not be as complete or thorough as they would using experimental design techniques. Additionally, in those cases, it is critically important to put the results of the test into context of the exploratory analysis results. In all cases, the team should identify areas that are significantly different from those examined during the exploratory analysis and seek to identify reasons for discrepancies.

There is a wide range of available statistical analysis techniques that can be used for such analyses, including traditional statistics and regression, classification and regression trees (CART), contour and interaction plots, and decision analysis techniques (particularly for observational data). Some of these analyses can be quite complex and required sophisticated techniques. However, these data analysis and data mining techniques may be critical given the large amount of data likely to be generated. The analyses conducted in this step provide input for the final joint mission effectiveness evaluation.

#### G.3.2.2.1 Analyze System of Systems Performance

System/SoS level data will be used to produce quantitative and qualitative analysis results, using statistical and operational measures developed as part of the evaluation framework, including key performance parameters (KPP), system/SoS measures of effectiveness (MOE), system/SoS measures of suitability (MOS), and system/SoS measures of performance (MOP). These measures will provide results at the lowest level of the joint measure framework. The analysis should provide the statistical rigor and operational context to produce results that analysts and SMEs will use for evaluation of SUT(s) and SoS at the system/SoS level. Upon completion of this step the analysts and SMEs should have system of systems performance information.

The analyst should try to gain an understanding of how the different factors tested affected system/SoS performance. Thus, the analyst should be able to show the decision maker how different experimental factors, and combinations of factors, affected system/SoS performance and whether that performance met the required thresholds across the factor ranges of interest. If systems are being compared, the analyst should be able to show the performance of one system relative to the others via standard hypothesis testing techniques. This is not as simple as it appears at first glance. A key challenge is the fact that there are likely to be numerous MOSA against which the systems are being evaluated. Identifying systems that did not meet threshold values is fairly straightforward, but evaluating a system, whether individually or compared to other systems, based upon multiple measures is more challenging. The analysis cannot look at each measure in isolation, but must look across the measures to evaluate the system. The analyst should look at the correlation between the measures to identify trends or relationships between measures which may point to general strengths or weaknesses of the systems. Additionally, assuming that the systems met the threshold values set forth for the program, performance in different MOSA often represent trade-offs that must be made. The analysis must be conducted in a way that will support decision making concerning these trade-offs. That often requires an understanding of the value of one MOSA versus another and may require decision analysis techniques.

#### G.3.2.2.2 Analyze Task Performance

Similar to the previous step, the analysis team will produce quantitative and qualitative mission effectiveness analysis results at the task level, using statistical and operational measures reflecting mission-level task outcomes. These measures evaluate mission task accomplishment by the SoS in the joint mission environment. This level of measures is one of the two levels unique to joint test and evaluation, and is therefore essential to the evaluation of joint mission effectiveness (JMe).

The idea of examining the performance of systems/SoS at the task level is similar conceptually to that for system/SoS-level measures. The fact that task performance is likely to be affected by system/SoS-level performance is addressed in the next set of processes. Here, the analyst will want to show the decision maker how different experimental factors, and combinations of factors, affected task performance and whether that performance met the required thresholds across the factor ranges of

interest. One of the key factors will be the system involved, if more than one system is being compared. The analyst must have an understanding of the joint tasks and how the measures support them, because many of the Task MOP will not evaluate the performance of the joint task as a whole, but will evaluate a particular aspect of it. Thus, the team must know which measures correspond to the same task and analyze those together. Also, the analyst should look across tasks to gain insights about task performance as a whole. As before, there may be trade-offs to be made for performance at the task level, and these should be examined as well.

#### G.3.2.2.3 Analyze Mission Effectiveness

As with the previous two steps, the analysis team will produce quantitative and qualitative mission effectiveness analysis results at the mission level, using statistical and operational measures reflecting mission-level outcomes. These measures are used to assess changes in system behavior, capability, or operational environment tied to measuring the attainment of a mission end state, achievement of a mission objective, or creation of a mission effect. This level of measures is the other of the two levels unique to joint test and evaluation, and is therefore essential to the evaluation of joint mission effectiveness (JMe). To eliminate confusion, we must point out that the analyst here is focused on evaluating mission effectiveness based upon the Mission MOE. *Joint mission effectiveness (JMe)* is a term used to consider the whole picture (performance within and across Mission MOE, Task MOP, and MOSA), and is therefore the larger evaluation, of which this step is a necessary element.

As with the previous two steps, the idea of examining the performance of systems/SoS at the mission level is similar conceptually to that for system/SoS-level and task-level measures and those discussions apply here as well. One additional point to make, though, is that the higher up in the measures framework that the analyst gets, the more important will be the scenario within which the test was conducted. Such a consideration is important for all three measure levels, but at the mission level, the scenario takes on special importance. Variance between system/SoS performance in different scenarios is certainly likely, but expected or required performance at this level may be somewhat consistent. For example, many of a system's KPP will have threshold and objective values that are not scenario-dependent. At the task level, the standards for task performance may well be different across scenarios, but, in many cases, may not be. For example, the processing time for a call for fire may be the same regardless of whether the unit is conducting an attack or defending a forward operating base. But at the mission level, the standards for Mission MOE are likely to be wholly dependent upon the scenario. For example, the standards for survivability are likely to be higher for a defend mission than they are for an attack. Thus, the analyst must understand that the performance at the mission level must be understood in the context of the scenario, and that the joint SoS may perform entirely differently for another scenario. That, of course, is why multiple tests across varied scenarios are likely to be conducted.

#### G.3.2.3 Test Measure Relationship Analysis

Examination of multiple measures and their relationships across measure levels adds a new level of complexity and is the particular aspect that makes the post-test analyses for a joint test event unique. This step is important because it identifies the cross-measure level relationships which drive the final evaluation; it goes beyond traditional hierarchical relationships and seeks to bring order to what would otherwise be a chaotic set of information. It purposely seeks to examine the cross-level relationships between measures rather than relationships at the same level.

These techniques model and analyze interactions and correlations between multiple levels of performance measures to produce test results that measure a SUT(s) contribution to a mission outcome. This analysis can involve multiple evaluate – analyze – evaluate iterations within a test. As a capability evaluation is developed, the evaluation results from selected areas of the test's parameter space may improve the overall mission and system-to-mission analysis models. Thus, an evaluation of one area may improve the analyst's understanding of another area, leading to iterative refinements. Between-test analysis iterations also occur as techniques/parameters are further refined throughout the acquisition lifecycle as SUT/SoS fidelity increases. Combined, these analyses will help to produce the joint mission effectiveness information for the capability being tested.

In this step, the analyst starts with the existing measure structure model to evaluate the relationships of the measure responses during the test. The team may discover that the test results contradict the original model. Thus, they must be careful to reexamine all of the relationships during this step; however, the original model provides the team a starting point. Additionally, such information is essential for improving the exploratory analysis later if it is revisited for subsequent test events.

The analysis of relationships here takes on special importance beyond that conducted during the exploratory analysis. In that latter, the analysis of relationships was conducted to refine the evaluation framework and to identify factors of interest for the test event. Here, the analysis will provide the underpinning of the team's evaluation of the JMe, and will thus directly impact the future of acquisition programs. For the systems/SoS under test, the analysis techniques used and the subsequent results may depend on the location of the systems/SoS in their acquisition lifecycle. Thus, the questions to be answered, which will drive the analysis, will likely vary accordingly. Nonetheless, some analyses should be common across all tests. At the conclusion of this analysis, the team should understand how system-level performance (as evaluated by the MOSA) affected individual joint task performance (as evaluated by the Task MOP) and mission-level performance (as evaluated by the Mission MOE). Similarly, they should understand how task-level performance affected mission-level performance, and how performance at all three levels affected JMe as a whole. Thus, the analyst should be able to trace performance at each level through the other levels in the measure structure.

Techniques for doing this include analysis of correlation, regression analyses between levels, and other qualitative techniques, all of which are discussed in later in this annex. The Matrix of Failure Mode Effects Analysis (MFMEA) follows a systematic inductive

approach to map the relationships between basic failures at a low level and their impacts at the higher levels. Another technique might include the use of systems dynamics simulations, which provide insights into second- and third-order effects by modeling the system structure and feedback loops. In this technique, the analyst would develop the system of systems structure in the model and populate it with data from the test.

## G.4 SoS and Joint Mission Effectiveness Evaluation

In this process, the team consolidates all of the findings from the analysis of the data and draws key insights that will allow the team to report on overall joint mission effectiveness and make appropriate recommendations. With the synthesis of the evaluations, the team begins to identify significant findings and to make recommendations. It is expected, though, that the team will begin to formulate and refine recommendations as they begin to consolidate their findings. For instance, the team may look at the questions that they are required to answer in order to identify and group significant findings. Therefore, they may take a piecewise approach and address each question separately by grouping the findings and making recommendations relative the question. In the end, however, the team must ensure that they reexamine all of their findings and recommendations holistically before finalizing their evaluation.

Once the test data have been analyzed, the joint mission evaluator uses test results to evaluate the overall JMe and the contribution an individual system under test makes to the accomplishment of the joint mission. Using statistical and other analytical results regarding the system or systems-under-test across test trials and types of measures, conclusions are reached as to the system/SoS contributions to overall mission performance in the joint mission environment. Due to the complex nature of SoS interactions while testing in a joint environment parameter space, analysts need to explicitly state conclusions in terms of assumptions that were made throughout the CTM process. The assumptions should be related to test factor variations and controls during the test. When rigorous statistical significance is not found concerning test hypotheses, analysis can provide trend-line indications of system/SoS contributions to mission effectiveness, which can be further explored in follow-on test events. Evaluation takes the results of the analysis and applies additional reasoning (mental manipulation) to arrive at conclusions and recommendations.

### G.4.1 Integration of Exploratory Results

The exploratory analyses conducted in ***CTM 1*** refined the overall JMe test factor space into a subset of that space, from which test trial sets were taken. A goal of the test is to develop a summary performance evaluation and joint capabilities evaluation (JCE) based upon the data collected; however, only a limited subset of the possible conditions and scenarios will have actually been tested. If that limited subset was chosen via a good DOE in the exploratory analysis, then the team can still say quite a lot about how inputs affect outputs. Thus, the test should be examined in context of the exploratory analysis. Integration of those preliminary results will provide insights into how the test results fit into the larger context of the test factor space and will help determine the feasibility of extrapolating the test results to the broader factor space for a joint capabilities evaluation. For instance, if the trends and results observed during the test event seem to exhibit the same trends and relationships as that observed during the exploratory analysis, the evaluators may conclude that the factor subset chosen for test was appropriate and representative of the most important factors. They may

also be able to make recommendations about future testing based upon the exploratory analysis results that were not examined during this test event.

Additionally, the test results should be compared to the exploratory analysis results to validate (or not) the assumptions made during the exploratory analysis. If major discrepancies exist between the exploratory analysis results and the test results, a recommendation might be to revisit the exploratory analysis using the data obtained from the actual test event (as part of a model-test-model analytic framework).

### G.4.2 Evaluation Synthesis

The key results may be in different places or in different forms at this point, as expected due to the myriad of sources of raw data and tools used to support analysis. Additionally, not all of the results may help shed light on the questions to be answered. A simple collation of results, then, is not sufficient for a capability evaluation. The synthesis that should occur here is a matter of identifying, grouping, and relating key results. First, the analysis team will identify and select key results from the analysis that will directly support the questions to be answered and the recommendations to be made. They should then group those results into logical categories, possibly by the primary analysis questions. Finally, they will relate the results within and between the groupings and develop insights into what the analysis is conveying about the joint capability.

As part of this process, joint capability evaluation results are synthesized into a summary performance evaluation and the joint capabilities evaluation (JCE) can begin to be drafted. The JCE contains test findings and recommendations as well as indicates impacts to test products and recommends possible transitions to the war fighter community. These findings are prioritized based on factors such as contribution to mission desired effects, fulfillment of test goals, and considerations for risk, to produce a secondary listing of significant findings and critical test product issues to support SoS and system program acquisition decisions. Here, the team can begin to develop the JCE by drafting the results and insights portions of the document. The JCE will then be developed further in the next two processes.

### G.4.3 Identification of Significant Findings

From all of the analysis conducted thus far and the results synthesis, findings that have significant implications concerning the system under test (SUT) or the joint capability are captured and organized. These findings will be the foundation for the recommendations made during the next step. They must be captured and put into context given the assumptions made throughout the test preparation and execution. The relationships between measures and their impact on mission effect or end state achievement should be reviewed and captured. Additionally, the final quantification of how the SoS or non-materiel changes impact the creation of a mission effect should be presented. Also, during this step, the evaluators will continue to refine the JCE.

In this process, the analysis team must tell the story that will lead directly to the recommendations. Thus, they should examine the evidence and present it in such as way as to suggest the recommendations even before they are made. This is not to say that insights and findings that may caveat or contradict the recommendations should be omitted or

ignored. On the contrary, all key evidence uncovered by the analysis team that has a significant impact on the conclusions must be presented. To do otherwise is ethically wrong. However, the team must still ensure that the information presented is arranged in a way that leads to the recommendations. They must not expect the decision maker to sift through the results and findings to connect the dots.

### G.4.4 Recommendations

The final recommendations should be based on the evidence and supported by the above analysis. As previously mentioned, if critical thinking was applied throughout the process and the results and findings are presented well, the results should be self-evident. Even so, players in the PPBE and Acquisition systems rely on joint test results to support key decisions. At every milestone or control gate, a decision to move forward to the next phase of deployment is made. That decision can be one of three options: move forward unconditionally, return to early phases for more development work, or discontinue all investments and cancel the project. The recommendations made by the analysis should be very clear in this regard, supported with objective evidence and analysis involved in all the preceding steps. With the conclusion of this step, the joint capability evaluation should be in final draft and ready for review

## G.5 Techniques for Understanding Measure Relationships

### G.5.1 Overview

A critical step in the development of the evaluation framework is capturing the relationships between the multiple measure levels. In the exploratory analysis of ***CTM 1***, understanding the measure relationships will aid the analysis team in considering the responses across multiple measure levels in order to identify factors of interest as they refine the factor test space. In the capability evaluations of ***CTM 6***, an understanding of the measure-level relationships will assist the team in understanding the implications of the measure responses to the system/SoS.

There are several techniques discussed in this section. The first describes how the team can use the DoDAF products created as part of the capability lifecycle and during the JOC-T development to understand the inter-relationships between systems/SoS and the operational context (missions and tasks). The second proposes the use of an analytical technique called Matrix Failure Modes and Effects Analysis (MFMEA), to facilitate an understanding of the measure relationships and to trace measure responses between measure levels. The third technique discusses qualitative methods for assessing multiple measures within and across levels.

### G.5.2 Using DoDAF Products

Test analysts can use DoD Architecture Framework products to support joint mission effectiveness (JMe) evaluation strategy development. The DoDAF products not only describe the relationship between system elements but also include a mapping of operational tasks to system functions. They reveal the allocation of system functions to hardware and software elements and how those functions support the tasks. That is, they can help address the CCI:

***Assess the ability to perform Task X by SoS Configuration Y under Condition Set A to achieve Mission Desired Effect Z.***

Architecture products are the natural artifacts of good systems engineering: translating a customer's needs into a set of technical requirements and then developing and fielding a system that meets the requirements and satisfies the needs [Miller]. Starting with that need means examining who the end user is (Operational Node), what he intends to accomplish in his domain (Operational Activities) and what information is exchanged between those activities. The operational activities and information exchanges are enabled by systems and their interfaces. FigureG-3 (inspired by Gunn, et al [2008]) is a simplified visualization that describes the underlying relationships.

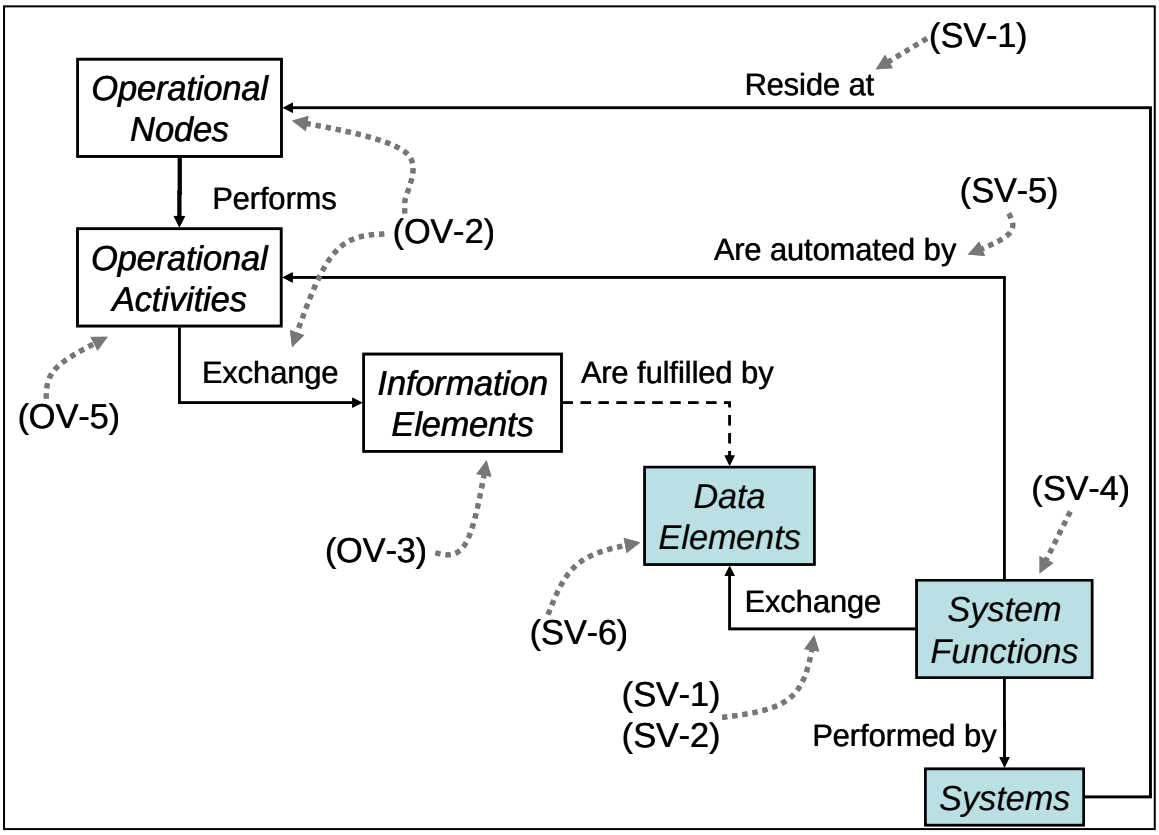



**Figure G-3. Simplified Representation of DoDAF Product Relationships**

Beginning with the doctrinally-supported (UJTL, Service-level Mission Essential Tasks, etc.) set of operational tasks captured in the OV-5 Operational Activity Model, one understands the mission area and joint capability to be supported by the SoS under test. Most activities require the exchange of information between operational nodes to support decision-making and proper execution, which is partially captured in the OV-2 Operational Node Connectivity Description and detailed in the OV-3 Operational Information Exchange Matrix. To this point, the architecture focuses on the operations domain. Objectives are achieved by the accomplishment of activities and tasks which work together because they produce and consume information [Lee, et al]. The exchange of information at the human level is

supported, automated and enabled by hardware and software systems. The functionality of the systems is described in the SV-4 Systems Functionality Description which is mapped to the operational activities they support by the SV-5 Operational Activity to Systems Function Traceability Matrix. The proper functioning of those systems requires the exchange of data, which is captured by SV-1 Systems Interface Description and SV-2 Systems Communication Description, and further detailed in the SV-6 Systems Data Exchange Matrix. The SV products depict the tools utilized to facilitate the conduct of the operations. They capture the communications and hardware-centric requirements of the SoS in support of the mission [Lee, et al]. Further, assignment of systems (with their interface) to operational nodes (with their information needlines) is depicted in SV-1 Systems Interface Description. By identifying the facilities which host system elements, we see which operational organizations operate which system elements. (Note: the above discussion is a simplification of much more complex concepts; it does not differentiate the network-centric extensions of DoDAF by calling out SV-4 and SV-4b or SV-5a, SV-5b and SV-5c separately.)

For products that are created within this integrated schema, the connections between system- and SoS-level measures (KPP, KSA, etc), Task MOPs and joint Mission MOEs are explicitly known. Further, the relationships between the activities and the relationships between the system nodes can help identify the measure relationships that exist between different hierarchical levels of MOEs and MOPs. They provide insight into SoS emergent behavior in which lateral influences dominate hierarchical relationships and causal interactions are not obvious or direct. Figure G-4 is a more comprehensive illustration of how one product integrates with others.

The integrated nature of a complete set of the products is apparent. No product stands alone. It is that very nature of interconnectedness that illuminates the SoS and operational relationships to support a solid composition of joint operational context. One possible scenario involves the application of artillery fires. Tasks like Conduct Fire Support Coordination, Select Target to Attack and Request Attack as decompositions of the higher level task Conduct Joint Fires are captured in the OV-5 along with related tasks like Collect Target Information and Provide Battle Damage Assessment. Measures for these tasks are available in the UJTL: a measure at the tactical level for Conduct Joint Fires is the percentage of targets attacked in accordance with requests for fires. Other measures are available in Service task lists. The operational nodes involved may include an infantry battalion fire support coordination center (FSCC), an artillery battalion fire direction center (FDC), an artillery battery FDC and a forward observer. The information they exchange includes Fire Request, Fire Order, Firing Data, Adjust Fires, and Battle Damage Assessment. Those nodes and the information exchanges are displayed graphically in the OV-2. The details of each information exchange (content, information producer and consumer, periodicity, timeliness, security, etc) are in the OV-3. Automated systems supporting the tasks include AFATDS, EPLRS and SINCGARS and are shown on the SV-1 with their communication connections on the SV-2. The details of each data exchange element (format, media, data standard, etc) are in the SV-6. The functions those systems perform include receive fire request, identify ordnance, deconflict airspace, verify trajectories and generate fire order and are described in the SV-4. A direct mapping of the system functions to the operational tasks they support is provided in the SV-5 matrix.

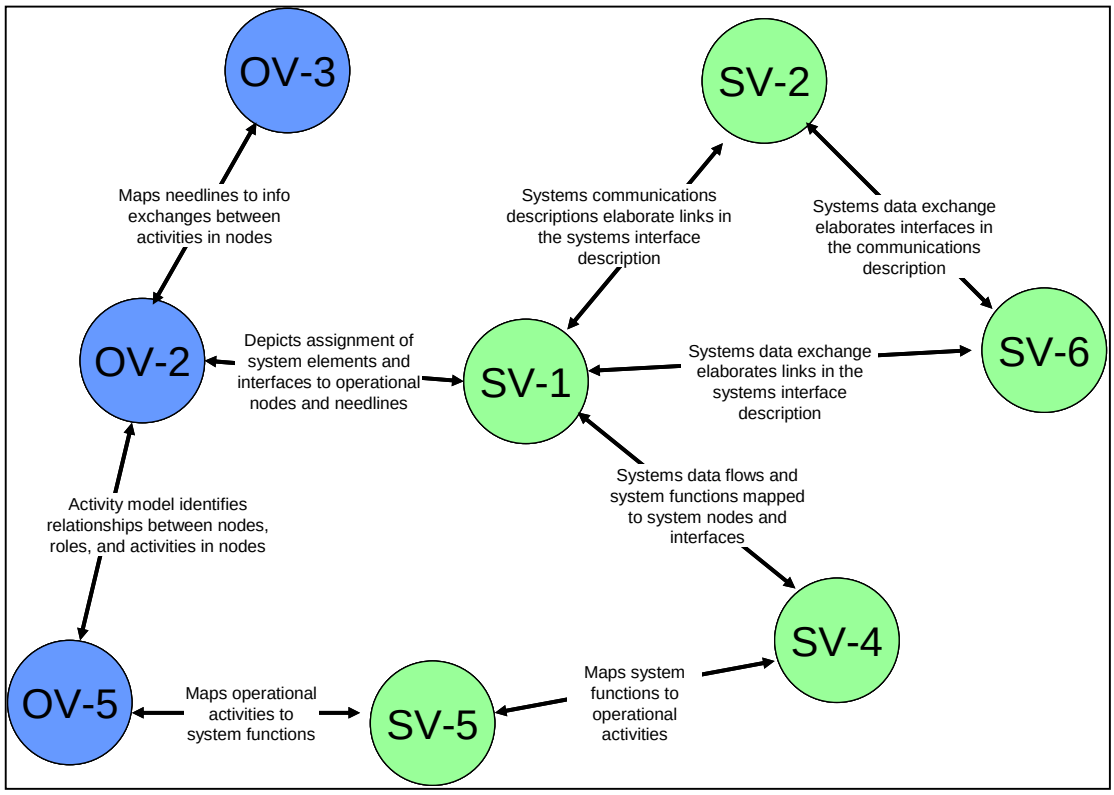



**Figure G-4. Selected DoDAF Product Relationships**

Ultimately, we should also consider where the DoDAF products fit in the larger scheme of CEM. FigureG-5 is a mapping of how DoDAF products can serve as a source of information in JOC-T. Some caution should be exercised, because not all the information required in JOC-T is captured in the DoDAF products and not all the information in the products is applicable in JOC-T.

Finally, when evaluating a SoS composed of elements from different acquisition program offices, each system may have its own set of DoDAF products. While they will be integrated within a set, they may not be completely consistent with other programs' products. The SoS under test may include System X fielded by PEO ABC and System Y fielded by PEO DEF, and both systems are required to achieve the task or desired mission effect. However, because the DoDAF products for those systems were developed, vetted and approved by different organizations, they may exhibit some differences. The opportunity of a critical examination of the products allows all stakeholders to identify such issues early in the process of developing an evaluation strategy and to take appropriate action.

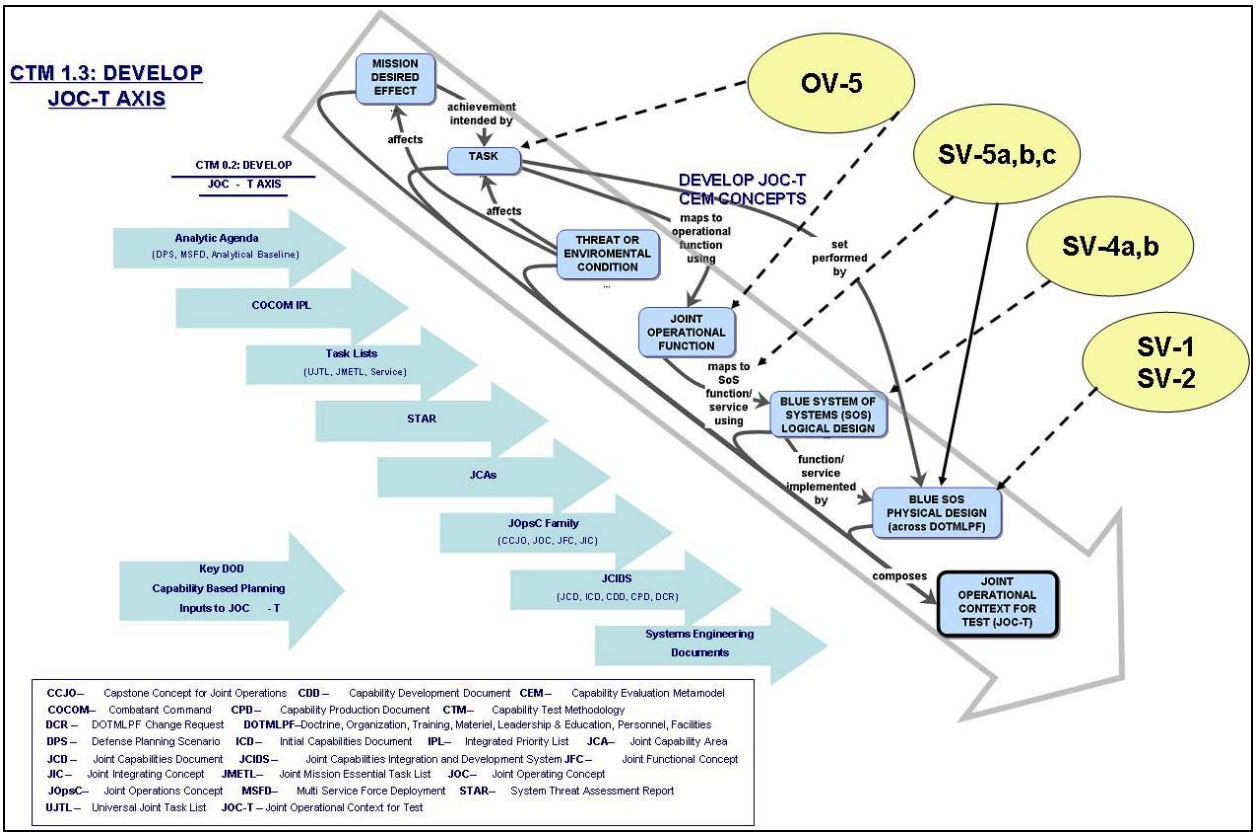



**Figure G-5. The JOC-T CEM and DoDAF Products**

### G.5.3 Matrix Failure Modes and Effects Analysis (MFMEA)

The power of Failure Modes and Effects Analysis (FMEA) as an effective design analysis and failure diagnostic tool is widely recognized (Mil-STD-1629A). It was initially introduced as a tool to systematically analyze all contributing component failure modes and identify the resulting effects on the system and mission. With minor modifications, the FMEA is now also being applied to analyze industrial processes and enterprise performance.

Without going into details, the FMEA follows a systematic inductive approach to map the relationships between basic failures at a low level and their impacts at the higher levels. The main disadvantage of the FMEA is that it considers only one failure at a time and not multiple or common cause failures. Its advantages are that it is simple to apply and provides an orderly examination of the hazard conditions at multiple hierarchical levels from component, to subsystem, to system, to mission.

The traditional FMEA is developed and reported in a tabular form. G. Barbour (1977) developed the MFMEA to provide a systematic and automated method for tracing the effects of a piece-part failure on the overall system and mission. A comparison of the MFMEA and the CTM measurement framework reveals a complementarily between them. The classical MFMEA traces failures from the circuit level to the system and mission: circuit failures (inputs, outputs, parts) → unit-level effects (performance) → system-level effects (mission impact). The CTM measurement framework maps SoS Performance → Task Performance → Mission Desired Effects. Based on this complementarily, we identify the MFMEA top

indenture level with the specific JME and the lower indenture levels with the tasks and systems under tests. FigureG-6 illustrates the development of the MFMEA for testing in a JME. Each row of the lowest matrix represents a system or system element and the columns show how those systems support or enable mission tasks. The MFMEA clearly traces failures from the system level to their effects on the tasks level to their effects on the mission.

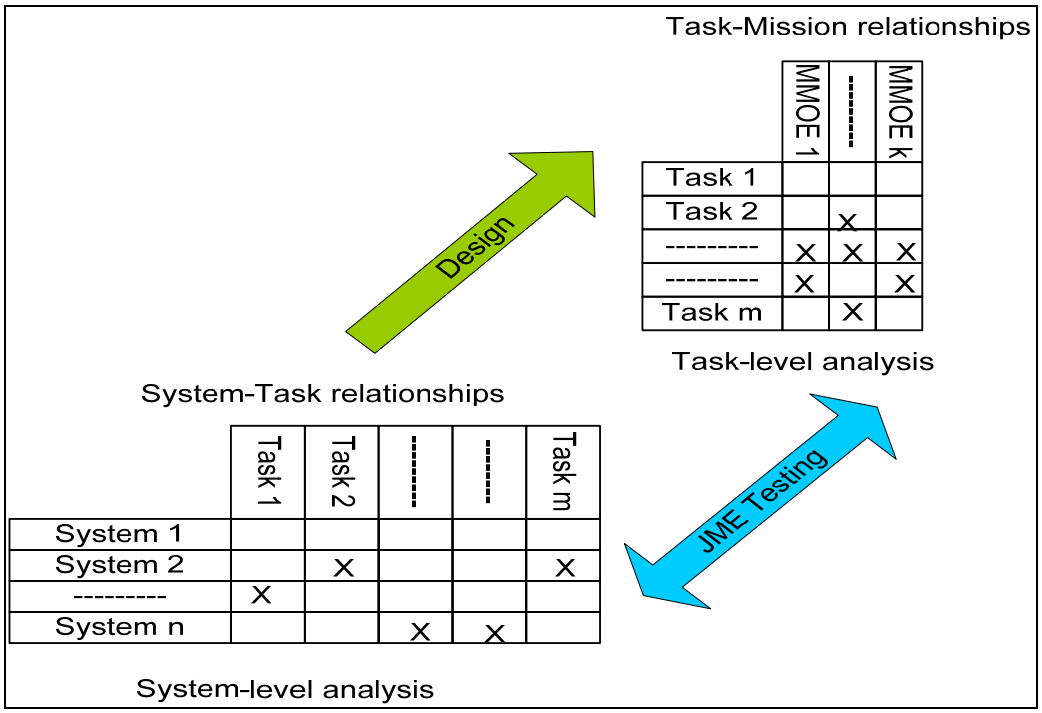



**Figure G-6. Illustrative MFMEA Applied to Evaluation Framework**

Today's SoS are often highly integrated and exhibit complex interactions across systems that dominate the SoS behavior and the Mission MOE (Calvano and John, 2004). The associated MFMEA is therefore complex and includes numerous off-diagonal elements. Test analysts must then consider effects of all the systems and tasks on mission performance. The MFMEA helps manage this complexity and ensure that the appropriateness of the analysis. For completeness, we note that the MFMEA parallels the DoDAF SV-5 which maps system functions to operational activities.

### G.5.4 Automated MFMEA Data Analysis

It is apparent from Figure G-6 that the proposed MFMEA organizes system failures, task substandard performance, and inadequate Mission MOE into structures cause and effect relationships. It provides test analysts with a tool that they can use to diagnose the root causes of poor Mission MOE performance through a combination of inductive and deductive reasoning.

The key steps in the JME analysis are as follows:

1) Evaluate the final states or outcomes for the desired mission states.
2) Compare the resulting Mission MOE against the required or desired Mission MOE.

3) For each below standard Mission MOE use deductive reasoning to identify the tasks that impact it.
4) For each below standard Task MOP identify the associated essential systems and personnel.
5) For each task identified in step 4, identify the below standard Task MOP and MOSA.
6) For each below standard Task MOP identify the material and non-material gaps.
7) For each below standard MOSA identify the hardware, software, and human system integration elements that are responsible for not achieving the standard MOSA.

The information derived by the proposed analysis assures that the weak or abnormal systems are identified for further analysis and corrections. Furthermore, by using inductive and deductive reasoning, the MFMEA is well suited to identify systems that meet all their KPP and KSA, but their very presence within the SoS and complex cross-system interactions cause tasks to be performed below standard resulting in poor mission performance. Note that different systems and tasks are applicable to different scenarios or operations. Different relationship matrices will therefore need to be developed for the different planned experiments. For complex systems of systems, the proposed matrix relationships can be entered into Excel spreadsheets and the links between given Mission MOE effects automatically rolled back to tasks and systems. The Rome Air Development Center has developed an automated MFMEA tool for use with electronic equipment (Dussault, 1984) and the National Aeronautics and Space Administration (NASA) has developed an automated monitoring and diagnostics system based on the FMEA (Herrin et al, 2001). We think that the development of MFMEA software based on Excel could greatly enhance the analysis of JME tests.

### G.5.5 Test Measure Relationship Analysis

While the previous analyses of individual measure responses are not simple tasks, they are fairly straightforward because they involve only one measure at a time. Examination of multiple measures and their relationships across measure levels adds a new level of complexity and is the particular aspect that makes the exploratory and post-test analyses for a joint test event unique. Unfortunately, there are no standard techniques for doing such analyses, but there are a variety of techniques from which to choose, each having its own strengths and weaknesses.

This step is critical because the analysis team must narrow the JMe factor test space using the entire measure structure. Unfortunately, this means that the analyst cannot just focus on the response characteristics of a single measure, but must consider all measures deemed important enough to be included in the evaluation framework. As previously discussed during the evaluation framework development process, the team should have chosen only those measures important to the evaluation of JMe. Thus, inclusion in the evaluation framework implies importance, and no measure should be ignored without justification.

There are two main aspects to the analysis of measure relationships. The first is an examination of the interactions and correlations among measures at the same level and between measures from different levels. The second is the development of techniques to consider the multiple measure responses from all three levels systematically to refine the JMe

factor test space. The first necessarily precedes the second, since the relationships between the levels must be understood before techniques can be developed to refine the JMe factor test space.

#### G.5.5.1 Measure Relationships at the Same Level

One technique the analysis team can use to explore the relationships between measures at the same level is to look at all of the measure responses by DOE design point and record, for each measure, whether that measure is above or below the mean value of the responses, calculated using all model runs and DOE design points. This requires that the analyst set upper and lower bounds on the mean; only if the measure response is outside of those upper and lower bounds will it will be designated as falling above or below mean. Otherwise, without bounds, all measure responses will fall either above or below the mean, even if the deviation is insignificant. The analyst may set the bounds by using standard deviations or another method, such as those used in statistical quality control and control charts (such as X-bar and R charts). In the latter technique, if an observation is within X-bar plus or minus R the process is deemed "in control," and otherwise "out of control." The resulting table may look something like table 2.

**Table G-2. Example Qualitative Examination of Multiple Measures at the Same Measure Level**

| DP | M1 | M2 | M3 | M4 | M1/2 | M1/3 | M1/4 | M2/3 | M2/4 | M3/4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | + | + | 0 | + | ++ | +0 | ++ | +0 | ++ | 0+ |
| 2 | 0 | + | + | + | 0+ | 0+ | 0+ | ++ | ++ | ++ |
| 3 | + | 0 | - | - | +0 | +- | +- | 0- | 0- | -- |
| 4 | + | 0 | 0 | + | +0 | +0 | ++ | 00 | 0+ | 0+ |
| 5 | - | 0 | 0 | - | -0 | -0 | -- | 00 | 0- | 0- |
| … | | | | | | | | | | |

Each row in the table represents a separate design point (DP). Each column in the left half of the table represents a measure (e.g., measure #1 is designated M1). A "+" in the column indicates that the measure response fell above the upper bound; a "-" indicates that the measure response fell below the lower bound; and a "0" indicates that the response fell between the bounds. The columns on the right half of the table are all of the possible pairs of measures (e.g., the combination of measures #1 and #2 is represented by M1/2). For each of the combinations, the results from the left half of the table are recorded together (e.g., both M1 and M2 fell above the upper bound for DP 1, so M1/2 has a "++" in the column). Using this technique, there are nine possible combinations of "+", "-", and "0". The team can then count, for each pair of measures across all design points, the number of times each combination appears. If the counts appear evenly distributed across the possible combinations, then the team may conclude that there are no interactive relationships between the measures. If a majority of the design points fell into the "++" and "--" categories then the team may conclude that there appears to be a positive relationship between the two measures. If a majority of the design points fell into the "+-" and "-+" categories, then the team may conclude that there appears to be a negative relationship between the two measures. This technique can be extended beyond pairs of measures to triplets, etc. Currently, this technique only provides a *qualitative* understanding of the potential relationships, but may be helpful for identifying areas of

further exploration. This is a relatively new idea; therefore, there are no statistical underpinnings to the technique, although there appears to be potential for that.

There are other, statistically-based techniques that consider interactions and correlation between multiple measures. One technique is to calculate the correlation between the measures using standard equations. Correlation will give the team an understanding of the degree of *linear* relationship between the measure responses, but will not capture non-linear relationships. Another technique that considers multiple responses and is appropriate for DOE is the Advanced Response Surface Methodology (ARSM), which is discussed further in this annex. Other potential techniques and methods to examine the results are discussed there as well.

#### G.5.5.2 Measure Relationships at Different Levels

When exploring the relationships between multiple measure levels, the analysis team should draw upon the measure structure developed in the evaluation strategy. With the measure structure as a starting point, there are multiple techniques that may be used to examine the relationships.[1] One qualitative technique is an extension of the one discussed above in conjunction with Table G-2. If the previously-discussed technique has already been used for each of the measure levels, the team can easily extend those results as shown in Table G-3. In this table, each of the measure responses is recorded in the same way for the lower-level measures (represented in this case by T1-T3 for Task MOP), as well as the results for the higher-level measure of interest (represented in this case by M1 for Mission MOE). For each of the possible pairings between a lower-level measure (T1-T3) and the higher-level measure (M1), the results are recorded together in the right half. Using this technique, there are again nine possible combinations of "+", "-", and "0" and the team can conduct a similar analysis as that discussed previously. The technique can be extended to combinations between a higher-level measure and combinatorial groupings of lower-level measures. Again, this technique only provides a *qualitative* understanding of the potential relationships, but may be helpful for identifying areas of further exploration. This is a relatively new idea; therefore, there are no statistical underpinnings to the technique, although there appears to be potential for that.

**Table G-3. Example Qualitative Examination of Multiple Measures at Different Measure Levels**

| DP | T1 | T2 | T3 | M1 | T1/M1 | T2/M1 | T3/M1 |
|---|---|---|---|---|---|---|---|
| 1 | + | + | 0 | + | ++ | ++ | 0+ |
| 2 | 0 | + | + | + | 0+ | ++ | ++ |
| 3 | + | 0 | - | - | +0 | 0- | -- |
| 4 | + | 0 | 0 | + | +0 | 0+ | 0+ |
| 5 | - | 0 | 0 | - | -0 | 0- | 0- |
| … | | | | | | | |

Another technique that can be used is to apply regression to fit a model for a particular higher-level measure response (e.g., a Mission MOE). For instance, if the analysis team

[1] The reader may also consult AEC (2008), pp 7-9 through 7-11 for a discussion of measure relationships and techniques to aggregate measures.

treats the Mission MOE response as the response to be modeled and treats the lower-level measures (e.g., Task MOP) as the regressors, they can fit a model that will indicate the impact of each of the Task MOP on the Mission MOE response. Using only a regression, though, will not be sufficient for understanding the relationships since it does not account for possible correlation between the Task MOP (multicollinearity), which may exist in many cases. Thus, the analysis team should use standard techniques to deal with multicollinearity by developing a cross-correlation matrix between the Task MOP and removing those regressors that are heavily correlated with others. One particular drawback to this technique is that the team must make the case that the Task MOP will occur before the Mission MOE, and thus this is a natural division of dependent (Mission MOE) and independent (Task MOP) variables. Thus the team must be careful in their selection of measures to examine in this way by choosing those tasks and mission effects that meet that assumption. For instance, in a call-for-fire scenario, the component tasks of the call-for-fire mission will occur before the mission effects of destroying (or not) the target. On the other hand, call-for-fire tasks would not necessarily occur before mission effects such as survivability of the firing unit. These relationships should be evident in the measure structure developed earlier, which should be consulted before executing this technique.

If the analysis team is able to directly manipulate task performance in M&S and measure the Mission MOE as output, they may have a third technique available. Instead of varying the larger set of factors in the JMe factor test space, they can choose to vary the task performance (and thus the Task MOP) directly by treating the Task MOP as the factors. Thus, the team can apply a DOE to the analysis and directly measure the impact of Task MOP on Mission MOE. The same is true between the other possible levels as well. In fact, since the other possible level-to-level analyses will involve the Measures of System/SoS Attributes (MOSA) as the lower level measures, this technique may be available in even more cases. As discussed previously, MOSA are often the factors that are varied in a simulation model anyway, instead of being the measures that are evaluated. Thus, the team may have already captured many of the MOSA as factors to be varied in the model, and so team may be able to explore the relationships between the levels through the original DOE.

Another caveat to the use of correlation is that it does not measure causality, which is ultimately what the team would like to know. It can, however, assist the team in identifying causality by pointing out situations where it might exist, and where a causal mechanistic link can be sought.

#### G.5.5.3 Techniques for Considering Multiple Measures

Once the analysis team has explored and gained insights into the relationships between measures on the same level and between levels, they must decide how they will consider the multiple responses for the purpose of refining the JMe factor test space.

One set of techniques can be drawn from the decision-analysis field, specifically multi-attribute decision making (MADM) and related techniques. In these methods, weights are assigned to each of the measures according to their importance. The measure

responses are then converted to a common scale. A total score is then developed by multiplying the weight of each criterion by the scaled value of the responses and the products are summed. The particular methods used to assign weights and convert responses vary according to the specific MADM technique used. The result is a single value that represents a combined measure of JMe. MADM techniques require significant expert and decision-maker input, as the subjective assignment of weights, and the methods used to convert individual measure responses to a common scale, have a significant impact on the resulting total values.

The advantage of these techniques is that the team can now refine the JMe test factor space based upon the responses of the single response measure. There are some significant disadvantages however. First is the subjectivity of the methods. Even with significant stakeholder input and extensive sensitivity analyses, the subjectivity may make the results less defensible. Second is that many decision makers are uncomfortable with combining critical measures into a single value that has no intrinsic natural meaning. Third, multiple levels of measures complicate the model. Many MADM techniques are capable of dealing with different levels of measures; however, the relationships between the levels must be well-understood to construct the hierarchy, which is seldom the case. Additionally, there are assumptions about preferential independence that must be met when using an additive model as described above.

Another technique is to identify the number of measures for which each factor and factor interaction was significant. Table G-4 shows an example that tracks the number of measures (out of five) for which 16 factor main effects were significant. This technique allows the analysis team to prioritize the factors according to the number of measures for which the factors were significant. The team can then choose to eliminate the factors that fall into the lowest-priority categories.

**Table G-4. Example Tracking Matrix for Factor Significance by Number of Measures**

| All 5 Measures | 4 Measures | 3 Measures | 2 Measures | 1 Measure | None |
|---|---|---|---|---|---|
| Factor 3<br>Factor 4<br>Factor 10<br>Factor 13 | Factor 1<br>Factor 11<br>Factor 16 | Factor 2<br>Factor 5<br>Factor 12<br>Factor 15 | Factor 9<br>Factor 14 | Factor 8 | Factor 6<br>Factor 7 |

Of course, this method assumes that all measures are of equal importance. Therefore, it is likely that the team will want to develop a similar matrix for each of the levels of measures and then develop rules for deciding which factors will move forward to the next iteration of the exploratory analysis. As part of this, the team should also capture the specific factor-level range that showed the most significance, through methods such as CART. This technique should not replace the need for the analysis team to explore the measure responses for regions of interest, but may be an acceptable technique for deciding which factors are most important.

Regardless of the technique chosen, the analyst should consider the controllability of the factor, particularly before deciding to include the factor in the recommended test trial set.

Since such factors cannot be controlled during experimentation, they cannot be factors in the test DOE. However, if the team concludes that an uncontrollable factor has a significant impact on JMe, that fact should be noted and carried forward to the test planning stage. For instance, if the training level of the participants appears to be a significant factor, but cannot be controlled in the experiment, the test planning and analysis teams will want to be aware of that insight when designing the actual test event or analyzing the post-test results, because it may help put the results in context, particularly if the training level of the participants can be assessed prior to the test event.

## G.6 Techniques for Analysis

The analysis of data may require the use of different tools and techniques. Not intended to be all inclusive, some of the tools and techniques are described in this section. Enclosure 2 to this annex provides a decision aid to help identify which techniques are most appropriate for different data sets.

**G.6.1 Box plots.** Box plots (Chambers 1983) are an excellent tool for conveying location and variation information in data sets, particularly for detecting and illustrating location and variation changes between different groups of data.

A **box plot** (also known as a **box-and-whisker diagram** or **plot**) is a convenient way of graphically depicting groups of numerical data through their five-number summaries (the smallest observation, lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation). A box plot may also indicate which observations, if any, might be considered outliers. The box plot was invented in 1977 by the British statistician N A Sheldon. See Figure G-7.

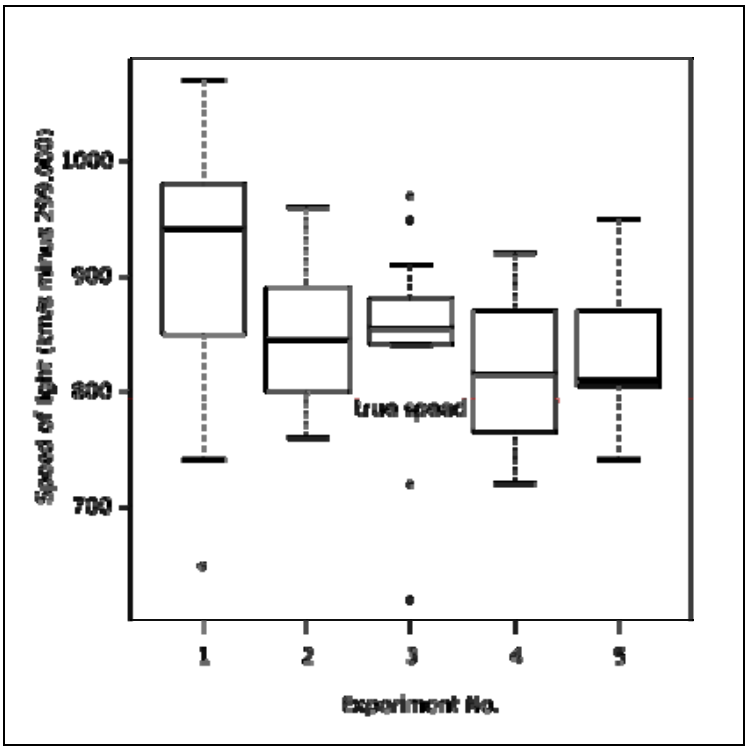



**Figure G-7. Box Plot of Data from the Michelson-Morley Experiment**

Box plots can be useful to display differences between populations without making any assumptions of the underlying statistical distribution. The spacings between the different parts of the box help indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. Box plots can be drawn either horizontally or vertically.

For a data set, one constructs a horizontal box plot in the following manner:

- Calculate the first LQ($x_{.25}$), the median ($x_{.50}$) and third quartile ($x_{.75}$)
- Calculate the interquartile range (IQR) by subtracting the first quartile from the third quartile. ($x_{.75} - x_{.25}$)
- Construct a box above the number line bounded on the left by the first quartile ($x_{.25}$) and on the right by the third quartile ($x_{.75}$).
- Indicate where the median lies inside of the box with the presence of a symbol or a line dividing the box at the median value.
- The mean value of the data can also be labeled with a point.
- Any data observation which lies more than $1.5 \cdot \mathrm{IQR}$ lower than the first quartile or $1.5 \cdot \mathrm{IQR}$ higher than the third quartile is considered an outlier. Indicate where the smallest value that is not an outlier is by connecting it to the box with a horizontal line or "whisker". Optionally, also mark the position of this value more clearly using a small vertical line. Likewise, connect the largest value that is not an outlier to the box by a "whisker" (and optionally mark it with another small vertical line).
- Indicate outliers by open and closed dots. "Extreme" outliers, or those which lie more than three times the IQR ($3 \cdot \mathrm{IQR}$) to the left and right from the first and third quartiles

respectively, are indicated by the presence of a closed dot. "Mild" outliers - that is, those observations which lie more than 1.5 times the IQR from the first and third quartile but are not also extreme outliers are indicated by the presence of an open dot. (Sometimes no distinction is made between "mild" and "extreme" outliers.)

- Add an appropriate label to the number line and title the box plot.
- A box plot may be constructed in a similar manner vertically as opposed to horizontally by merely interchanging "bottom" for "left", "top" for "right" and "vertical" for "horizontal" in the above description.

A plain-text version might look like this (figure G-8):

```
                              +-----+-+
  *           o      |-------|     | |---|
                              +-----+-+

+---+---+---+---+---+---+---+---+---+---+---+---+   number line
0   1   2   3   4   5   6   7   8   9  10  11  12
```

**Figure G-8. Example Box Plot**

For this data set:

- smallest non-outlier observation = 5 (left "whisker") (left "whisker" would have been 4 had there been an observation with a value of 4 ($Q1 - 1.5 \cdot \mathrm{IQR}$))
- lower quartile ($Q1$, $x_{.25}$) = 7
- median ($Med$, $x_{.5}$) = 8.5
- upper quartile ($Q3$, $x_{.75}$) = 9
- largest non-outlier observation = 10 (right "whisker")
- interquartile range, IQR = $Q3 - Q1 = 2$
- the value 3.5 is a "mild" outlier, between $1.5 \cdot \mathrm{IQR}$ and $3 \cdot \mathrm{IQR}$ below $Q1$
- the value 0.5 is an "extreme" outlier, more than $3 \cdot \mathrm{IQR}$ below $Q1$
- the data are skewed to the left (*negatively skewed*)

The horizontal lines (the "whiskers") extend to at most 1.5 times the box width (the interquartile range) from either or both ends of the box. They must end at an observed value, thus connecting all the values outside the box that are not more than 1.5 times the box width away from the box. Three times the box width marks the boundary between "mild" and "extreme" outliers. In this box plot, "mild" and "extreme" outliers are differentiated by closed and open dots, respectively.

There are alternative implementations of this detail of the box plot in various software packages, such as the whiskers extending to at most the 5th and 95th (or some more extreme) percentiles. Such approaches do not conform to Tukey's definition, with its emphasis on the median in particular and counting methods in general, and they tend to produce "outliers" for all data sets larger than ten, no matter what the shape of the distribution. There are also several minor variations on how to calculate the quartiles (see also quantile), and Tukey (1977) originally proposed instead using another variant that he named "hinges". The

difference between the definitions is no more than the difference between two consecutive data values, however, so it is always dwarfed by sampling variability as so is of little practical consequence.

Box and whisker plots are uniform in their use of the box: the bottom and top of the box are always the 25th and 75th percentile (the lower and upper quartiles, respectively), and the band near the middle of the box is always the 50th percentile (the median). But the ends of the whiskers can represent several possible alternative values, among them (see figure G-9):

- the minimum and maximum of all the data
- the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile
- one standard deviation above and below the mean of the data
- the 9th percentile and the 91st percentile
- the 2nd percentile and the 98th percentile

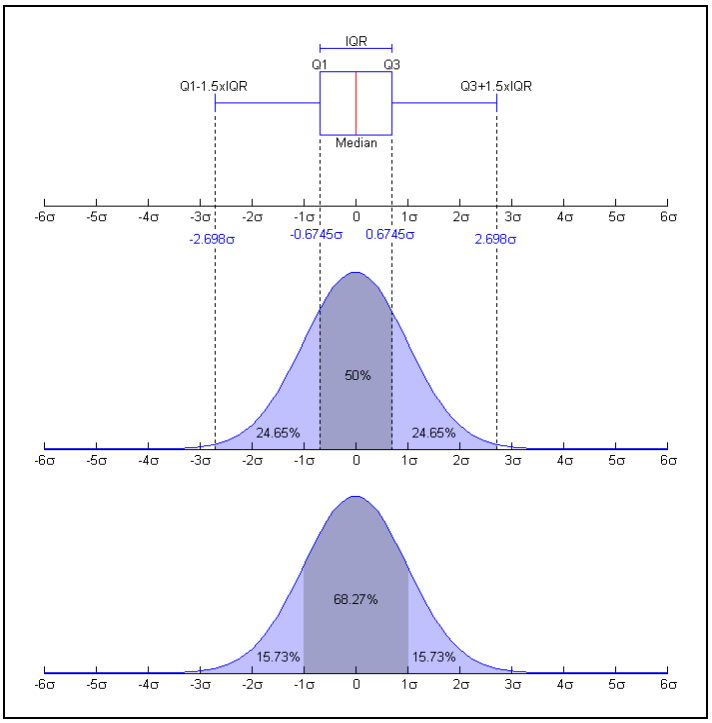



**Figure G-9. Box Plot and a Probability Density Function (PDF) of a Normal N(0,1σ2) Population**

Any data not included between the whiskers should be plotted as an outlier with a dot, small circle, or star, but occasionally this is not done.

Some box plots include an additional dot or a cross is plotted inside of the box, to represent the mean of the data in addition to the median.

On some box plots a crosshatch is placed on each whisker, before the end of the whisker.

Fairly rarely, box plots can be presented with no whiskers at all.

Because of this variability, it is appropriate to describe the convention being used for the whiskers and outliers in the caption for the plot.

The unusual percentiles 2%, 9%, 91%, 98% are sometimes used for whisker cross-hatches and whisker ends to show the seven-number summary. If the data are normally distributed the locations of the seven marks on the box plot will be equally spaced.[2]

**G.6.2 Hypothesis testing.** Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps.

- Formulate the null hypothesis (commonly, that the observations are the result of pure chance) and the alternative hypothesis (commonly, that the observations show a real effect combined with a component of chance variation).
- Identify a test statistic that can be used to assess the truth of the null hypothesis.
- Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the -value, the stronger the evidence against the null hypothesis.
- Compare the -value to an acceptable significance value (sometimes called an alpha value). If , that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

**G.6.3 Confidence intervals.** A range of values that contains (or covers) an unknown population or distribution parameter with pre-specified probability, typically set at 0.90 or higher. The values in a confidence interval can be thought of as "reasonable" values for the unknown parameter, in the sense that a two-sided hypothesis test for equality of the parameter to a particular number would not be rejected (at significance level equal to one minus the confidence level of the confidence interval) if and only if that number is within the confidence interval. A multi-dimensional generalization is a *confidence region* when several different parameters are involved.

**G.6.4 Regression Analysis.** A statistical method in which a dependent variable (or response) is assumed to depend on the values of other independent variables (or regressors). Regression then estimates the nature of this dependence, as well as assesses whether individual independent variables are statistically significant predictors of the dependent variable. The form of the regression model (linear, interactions, quadratic, etc.) is often developed with the help of plots of the dependent variable vs. the various possible independent variables.

The dependent variable in the **regression equation** is modeled as a function of the independent variables, corresponding parameters ("constants"), and an *error term*. The error term is treated as a *random variable*. It represents unexplained variation in the dependent

[2] www.en.wikipedia.org/wiki/Box_plots, 13 Jan 09. Wikipedia definition of Box plot.

variable. The parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method, but other criteria have also been used.

Regression can be used for prediction (including forecasting of time-series data), inference, hypothesis testing, and modeling of causal relationships. These uses of regression rely heavily on the underlying assumptions being satisfied.

Classical assumptions for regression analysis include:

- The sample must be representative of the population for the inference prediction.
- The error is assumed to be a random variable with a mean of zero conditional on the explanatory variables.
- The independent variables are error-free. If this is not so, modeling may be done using errors-in-variables model techniques.
- The predictors must be linearly independent, i.e., it must not be possible to express any predictor as a linear combination of the others. See Multicollinearity.
- The errors are uncorrelated, that is, the variance-covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant across observations (homoscedasticity). If not, weighted least squares or other methods might be used.

These are sufficient (but not all necessary) conditions for the least-squares estimator to possess desirable properties, in particular, these assumptions imply that the parameter estimates will be unbiased, consistent, and efficient in the class of linear unbiased estimators. Many of these assumptions may be relaxed in more advanced treatments.

It is convenient to assume an environment in which an experiment is performed: the dependent variable is then outcome of a measurement. The regression equation deals with the following variables:

- The **unknown parameters** denoted as $\boldsymbol{\beta}$. This may be a scalar or a vector of length $k$.
- The **independent variables**, $\mathbf{X}$.
- The **dependent variable**, $Y$.

Regression equation is a function of variables $\mathbf{X}$ and $\boldsymbol{\beta}$.

$$Y = f(X, \beta)$$

The user of regression analysis must make an intelligent guess about this function. Sometimes the form of this function is known; sometimes he must apply a trial and error process.

Assume now that the vector of unknown parameters, $\boldsymbol{\beta}$ is of length $k$. In order to perform a regression analysis the user must provide information about the dependent variable $Y$:

- If the user performs the measurement $N$ times, where $N < k$, regression analysis cannot be performed: there is not provided enough information to do so.
- If the user performs $N$ independent measurements, where $N = k$, then the problem reduces to solving a set of $N$ equations with $N$ unknowns $\boldsymbol{\beta}$.
- If, on the other hand, the user provides results of $N$ independent measurements, where $N > k$ regression analysis can be performed. Such a system is also called an overdetermined system;

In the last case the regression analysis provides the tools for:

1. finding a solution for unknown parameters $\boldsymbol{\beta}$ that will, for example, minimize the distance between the measured and predicted values of the dependent variable $Y$ (also known as method of least squares).
2. under certain statistical assumptions the regression analysis uses the surplus of information to provide statistical information about the unknown parameters $\boldsymbol{\beta}$ and predicted values of the dependent variable $Y$

**Independent measurements:**
Quantitatively, this is explained by the following example: Consider a regression model with, say, three unknown parameters $\beta_0$, $\beta_1$ and $\beta_2$. An experimenter performed 10 repeated measurements at exactly the same value of independent variables $\mathbf{X}$. In this case regression analysis fails to give a unique value for the three unknown parameters: the experimenter did not provide enough information. The best one can do is to calculate the average value of the dependent variable $Y$ and its standard deviation.

If the experimenter had performed five measurements at $\mathbf{X}_1$, four at $\mathbf{X}_2$ and one at $\mathbf{X}_3$, where $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ are different values of the independent variable $\mathbf{X}$ then regression analysis would provide a unique solution to unknown parameters $\boldsymbol{\beta}$.

In the case of general linear regression (see below) the above statement is equivalent to the requirement that matrix $\mathbf{X}^T\mathbf{X}$ is regular (that is: it has an inverse matrix).

**Statistical assumptions:**
When the number of measurements, $N$, is larger than the number of unknown parameters, $k$, and the measurement errors $\varepsilon_i$ (see below) are normally distributed then *the excess of information* contained in ($N$ - $k$) measurements is used make the following statistical predictions about the unknown parameters:

- confidence intervals of unknown parameters.

In linear regression, the model specification is that the dependent variable, $y_i$ is a linear combination of the *parameters* (but need not be linear in the *independent variables*). For example, in simple linear regression for modeling $N$ data points there is one independent variable: $x_i$, and two parameters, $\beta_0$ and $\beta_1$:

straight line: $$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,, \quad i = 1, \ldots, N$$

In multiple linear regression, there are several independent variables or functions of independent variables. For example, adding a term in $x_i^2$ to the preceding regression gives:

parabola: $$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i\,, \quad i = 1, \ldots, N$$

This is still linear regression; although the expression on the right hand side is quadratic in the independent variable $x_i$, it is linear in the parameters $\beta_0$, $\beta_1$ and $\beta_2$.

In both cases, $\varepsilon_i$ is an error term and the subscript $i$ indexes a particular observation. Given a random sample from the population, we estimate the population parameters and obtain the sample linear regression model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$$

The term $e_i$ is the residual, $\varepsilon_i = y_i - \hat{y}_i$.

One method of estimation is ordinary least squares. This method obtains parameter estimates that minimize the sum of squared residuals, SSE:

$$SSE = \sum_{i=1}^{N} e_i^2$$

Minimization of this function results in a set of normal equations, a set of simultaneous linear equations in the parameters, which are solved to yield the parameter estimators, $\hat{\beta}_0, \hat{\beta}_1$. See regression coefficients for statistical properties of these estimators. See Figure G-10 for an example of data points and the associated regression line.

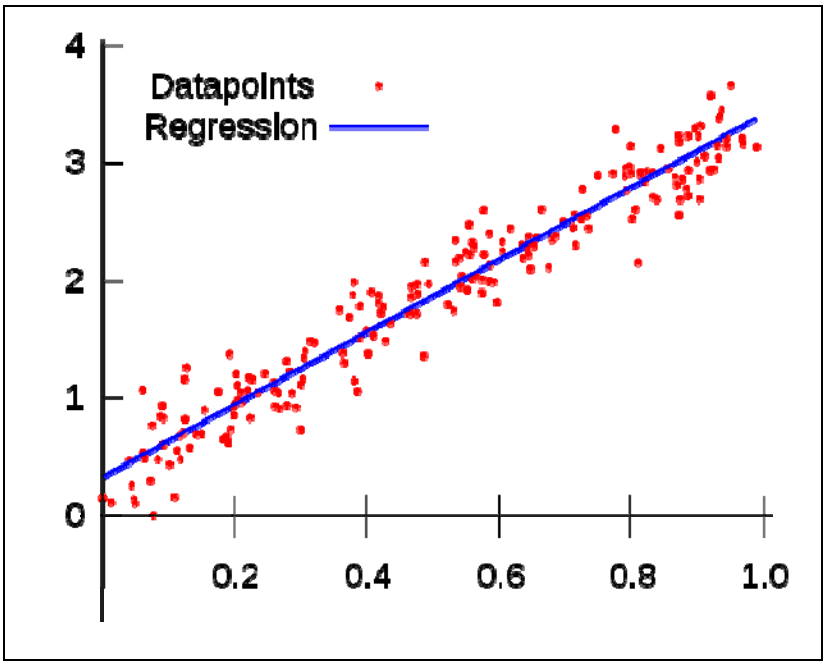



**Figure G-10. Illustration of Linear Regression on a Data Set (red points)**

In the case of simple regression, the formulas for the least squares estimates are:

$$\hat{\beta}_1 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2} \qquad \text{and} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where $\bar{x}$ is the mean (average) of the $x$ values and $\bar{y}$ is the mean of the $y$ values. See linear least squares(straight line fitting) for a derivation of these formulas and a numerical example. Under the assumption that the population error term has a constant variance, the estimate of that variance is given by:

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{SSE}{N-2}}$$

This is called the root mean square error (RMSE) of the regression. The standard errors of the parameter estimates are given by:

$$\hat{\sigma}_{\beta_0} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum\left(x_i - \bar{x}\right)^2}}$$

$$\hat{\sigma}_{\beta_1} = \hat{\sigma}_{\varepsilon} \sqrt{\frac{1}{\sum \left(x_i - \bar{x}\right)^2}}$$

Under the further assumption that the population error term is normally distributed, the researcher can use these estimated standard errors to create confidence intervals and conduct hypothesis tests about the population parameters.

**General Linear Model:**
In the more general multiple regression model, there are *p* independent variables:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

The least square parameter estimates are obtained by *p* normal equations. The residual can be written as

$$\varepsilon_i = yi - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \cdots - \hat{\beta}_p x_p$$

The normal equations are

$$\sum_{i=1}^{N} \sum_{k=1}^{p} X_{ij} X_{ik} \hat{\beta}_k = \sum_{i=1}^{N} X_{ij} y_i, \quad j = 1, \ldots, p$$

In matrix notation, the normal equations are written as

$$\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)\hat{\beta} = \mathrm{X}^{\mathrm{T}} y$$

**Regression diagnostics:**
Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Commonly used checks of goodness of fit include the R-squared, analyses of the pattern of residuals and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit, followed by t-tests of individual parameters.

Interpretations of these diagnostic tests rest heavily on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated. For example, if the error term does not have a normal distribution, in small samples the estimated parameters will not follow normal distributions, which complicates inference. With

relatively large samples, however, a central limit theorem can be invoked such that hypothesis testing may proceed using asymptotic approximations.

**Regression with limited dependant variables:**
The response variable may be non-continuous ("limited" to lie on some subset of the real line). For binary (zero or one) variables, if analysis proceeds with least-squares linear regression, the model is called the linear probability model. Nonlinear models for binary dependent variables include the probit and logit model. The multivariate probit model makes it possible to estimate jointly the relationship between several binary dependent variables and some independent variables. For categorical variables with more than two values there is the multinomial logit. For ordinal variables with more than two values, there are the ordered logit and ordered probit models. Censored regression models may be used when the dependent variable is only sometimes observed, and Heckman correction type models may be used when the sample is not randomly selected from the population of interest. An alternative to such procedures is linear regression based on polychoric or polyserial correlations between the categorical variables. Such procedures differ in the assumptions made about the distribution of the variables in the population. If the variable is positive with low values and represents the repetition of the occurrence of an event, count models like the Poisson regression or the negative binomial model may be used.

**Interpolation and extrapolation:**
Regression models predict a value of the *y* variable given known values of the *x* variables. If the prediction is to be done within the range of values of the *x* variables used to construct the model this is known as interpolation. Prediction outside the range of the data used to construct the model is known as extrapolation and it is more risky.

**Other methods:**
Although the parameters of a regression model are usually estimated using the method of least squares, other methods which have been used include:

- Bayesian methods, e.g. Bayesian linear regression
- Minimization of absolute deviations, leading to quantile regression
- Nonparametric regression. This approach requires a large number of observations, as the data are used to build the model structure as well as estimate the model parameters. They are usually computationally intensive.[3]

**G.6.5 ANOVA.** Common acronym for *Analysis of Variance*, a formal statistical method used to test whether the means of several different populations are all the same as each other, or if there is at least one mean that differs significantly from at least one other mean. If it appears that not all means are equal to each other, a follow-up analysis, *multiple comparisons*, is often done to identify which means differ significantly from which other means, and by how much and in which direction. There are many variations and generalizations of ANOVA, discussed in any statistics book, and usually available in statistical software. The name is somewhat of a misnomer, as we are analyzing means, not

[3] www.en.wikipedia.org/wiki/Regression_analysis, 13 Jan 09. Wikipedia definition of Regression analysis.

variances; however, this analysis is carried out via formulas that are similar to sample-variance formulas (weighted sums of squared differences).

In statistics, **analysis of variance** (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables. The initial techniques of the analysis of variance were developed by the statistician and geneticist R. A. Fisher in the 1920s and 1930s, and is sometimes known as **Fisher's ANOVA** or **Fisher's analysis of variance**, due to the use of Fisher's F-distribution as part of the test of statistical significance.

There are three conceptual classes of such models:

1. Fixed-effects models assumes that the data came from normal populations which may differ only in their means. (Model 1)
2. Random effects models assume that the data describe a hierarchy of different populations whose differences are constrained by the hierarchy. (Model 2)
3. Mixed-effect models describe situations where both fixed and random effects are present. (Model 3)

In practice, there are several types of ANOVA depending on the number of treatments and the way they are applied to the subjects in the experiment:

- One-way ANOVA is used to test for differences among two or more independent groups. Typically, however, the One-way ANOVA is used to test for differences among at least three groups, since the two-group case can be covered by a T-test (Gossett, 1908). When there are only two means to compare, the T-test and the F-test are equivalent; the relation between ANOVA and $t$ is given by $F = t^2$.
- One-way ANOVA for repeated measures is used when the subjects are subjected to repeated measures; this means that the same subjects are used for each treatment. Note that this method can be subject to carryover effects.
- **Factorial ANOVA** is used when the experimenter wants to study the effects of two or more treatment variables. The most commonly used type of factorial ANOVA is the 2×2 (read: two by two) design, where there are two independent variables and each variable has two levels or distinct values. Factorial ANOVA can also be multi-level such as 3×3, etc. or higher order such as 2×2×2, etc. but analyses with higher numbers of factors are rarely done by hand because the calculations are lengthy and the results are hard to interpret. However, since the introduction of data analytic software, the utilization of higher order designs and analyses has become quite common.
- When one wishes to test two or more independent groups subjecting the subjects to repeated measures, one may perform a factorial mixed-design ANOVA, in which one factor is a between subjects variable and the other is within subjects variable. This is a type of mixed effect model.
- Multivariate analysis of variance (MANOVA) is used when there is more than one dependent variable.

**Fixed-effects model:**
The fixed-effects model of analysis of variance applies to situations in which the experimenter applies several treatments to the subjects of the experiment to see if the response variable values change. This allows the experimenter to estimate the ranges of response variable values that the treatment would generate in the population as a whole.

**Random effects model:**
Random effects models are used when the treatments are not fixed. This occurs when the various treatments (also known as factor levels) are sampled from a larger population. Because the treatments themselves are random variables, some assumptions and the method of contrasting the treatments differ from ANOVA model 1.

Most random-effects or mixed-effects models are not concerned with making inferences concerning the particular sampled factors. For example, consider a large manufacturing plant in which many machines produce the same product. The statistician studying this plant would have very little interest in comparing the three particular machines to each other. Rather, inferences that can be made for *all* machines are of interest, such as their variability and the overall mean.

**Assumptions:**
- Independence of cases - this is a requirement of the design.
- Normality - the distributions in each of the groups are normal.
- Equality (or "homogeneity") of variances, called homoscedasticity — the variance of data in groups should be the same.

Levene's test for homogeneity of variances is typically used to confirm homoscedasticity. The Kolmogorov-Smirnov or the Shapiro-Wilk test may be used to confirm normality. Some authors claim that the F-test is unreliable if there are deviations from normality (Lindman, 1974) while others claim that the F-test is robust (Ferguson & Takane, 2005, pp.261-2). The Kruskal-Wallis test is a nonparametric alternative which does not rely on an assumption of normality.

These together form the common assumption that the errors are independently, identically, and normally distributed for fixed effects models, or:

$$\varepsilon \sim N\left(0, \sigma^2\right)$$

ANOVA models 2 and 3 have more complex assumptions about the expected value and variance of the residuals since the factors themselves may be drawn from a population.

**Partitioning of the sum of squares:**
The fundamental technique is a partitioning of the total sum of squares (abbreviated *SS*) into components related to the effects used in the model. For example, we show the model for a simplified ANOVA with one type of treatment at different levels.

$$SS_{Total} = SS_{Error} + SS_{Treatments}$$

The number of degrees of freedom (abbreviated *df*) can be partitioned in a similar way and specifies the chi-square distribution which describes the associated sums of squares.

$$df_{Total} = df_{Error} + df_{Treatments}$$

See also Lack-of-fit sum of squares.

**The F-test:**
The F-test is used for comparisons of the components of the total deviation. For example, in one-way, or single-factor ANOVA, statistical significance is tested for by comparing the F test statistic

$$F = \frac{\text{variance of the group means}}{\text{mean of the within-group variances}}$$

$$F^* = \frac{MSTR}{MSE}$$

where:

$$MSTR = \frac{SSTR}{I-1}, \text{ I = number of treatments}$$

and

$$MSE = \frac{SSE}{n_T - I}, \; n_T = \text{total number of cases}$$

to the F-distribution with $I\text{-}1, n_T$ degrees of freedom. Using the F-distribution is a natural candidate because the test statistic is the quotient of two mean sums of squares which have a chi-square distribution.

**ANOVA on ranks:**
As first suggested by Conover and Iman in 1981, in many cases when the data do not meet the assumptions of ANOVA, one can replace each original data value by its rank from 1 for the smallest to N for the largest, then run a standard ANOVA calculation on the rank-transformed data. "Where no equivalent nonparametric methods have yet been developed such as for the two-way design, rank transformation results in tests which are more robust to non-normality, and resistant to outliers and non-constant variance, than is ANOVA without the transformation." (Helsel & Hirsch, 2002, Page 177). However Seaman *et al.* (1994) noticed that the rank transformation of Conover and Iman (1981) is not appropriate for testing interactions among effects in a factorial design as it can cause an increase in Type I error (alpha error). Furthermore, if both main factors are significant there is little power to detect interactions.

A variant of rank-transformation is 'quantile normalization' in which a further transformation is applied to the ranks such that the resulting values have some defined distribution (often a normal distribution with a specified mean and variance). Further analyses of quantile-normalized data may then assume that distribution to compute significance values.


- Conover, W. J. & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124-129. [1] http://is.ba.ttu.edu/conover/Dr.Conover.htm [2] http://citeseer.ist.psu.edu/context/1743341/0

- Helsel, D. R., & Hirsch, R. M. (2002). *Statistical Methods in Water Resources: Techniques of Water-Resources Investigations, Book 4, Chapter A3*. U.S. Geological Survey. 522 pages.[3] http://pubs.water.usgs.gov/twri4a3

- Seaman, J. W., Walls, S. C., Wide, S. E., & Jaeger, R. G. (1994). Caveat emptor: Rank transform methods and interactions. *Trends Ecol. Evol.*, *9*, 261-263.


**Effect size measures:**
Several standardized measures of effect are used within the context of ANOVA to describe the degree of relationship between a predictor or set of predictors and the dependent variable.

**$\eta^2$**( *Eta-squared* ): Eta-squared describes the percentage of variance explained in the dependent variable by a predictor variable. It is a biased estimate of population variance explained.

**partial $\eta^2$** ( *Partial eta-squared* ): Partial eta-squared describes the percentage of variance explained in the dependent variable by a predictor controlling for other predictors. It is a biased estimate of the variance explained in the population. The following rules of thumb have emerged: small = 0.01; medium = 0.06; large = 0.14. These rules were taken from: Kittler, J. E., Menard, W., & Phillips, K., A. (2007). Weight concerns in individuals with body dysmorphic disorder. Eating Behaviors, 8, 115-120.

Omega Squared: Omega squared provides a relatively unbiased estimate of the variance explained in the population by a predictor variable.

**Cohen's *f*** This measure of effect size is frequently encountered when performing power analysis calculations. Conceptually it represents the square root of variance explained over variance not explained.

**Follow up tests:**
A statistically significant effect in ANOVA is often followed up with one or more different follow-up tests. This can be done in order to assess which groups are different from which other groups or to test various other focused hypotheses. Follow up tests are often distinguished in terms of whether they are planned (a priori) or post hoc. Planned tests are determined before looking at the data and post hoc tests are performed after looking at the data. Post hoc tests such as Tukey's test most commonly compare every group mean with every other group mean and typically incorporate some method of controlling of Type I errors. Comparisons, which are most commonly planned, can be either simple or compound. Simple comparisons compare one group mean with one other group mean. Compound comparisons typically compare two sets of groups means where one set has at two or more groups (e.g., compare average group means of group A, B and C with group D). Comparisons can also look at tests of trend, such as linear and quadratic relationships, when the independent variable involves ordered levels.

**Power analysis:**
Power analysis is often applied in the context of ANOVA in order to assess the probability of successfully rejecting the null hypothesis if we assume a certain ANOVA design, effect size in the population, sample size and alpha level. Power analysis can assist in study design by determining what sample size would be required in order to have a reasonable chance of rejecting the null hypothesis.[4]

**G.6.6 Post ANOVA Comparison – Tukey Post Hoc Comparison Test.** The Tukey Post Hoc comparison test is a procedure which allows for testing all the "pairs" of means present in an ANOVA test. Use of this test is pertinent when the ANOVA test statistic, $F_0$, was significant. The procedure is as follows:

First, compute the critical Tukey distance:

$$d_T = q_T \sqrt{\frac{MSE}{n_j}}$$

where:

$q_T$ is found from a statistical table,
MSE is taken from the ANOVA test,
and all $n_j$ are assumed to be equal. When the sample sizes are not equal, a technique to determine a representative $n_j$ is called the "harmonic" method. The generic equation for the "harmonic" method is:

[4] www.en.wikipedia.org/wiki/Anova, 13 Jan 09. Wikipedia definition for Analysis of variance.

$$\text{sample size} = \frac{j}{\left(\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_j}\right)}$$

where:

j represents the total number of samples,
and each *n* represents the size of each sample.

Second, calculate the absolute difference between each of the sample means from the ANOVA test (it may be beneficial to create a table – see Table G-5 for an example). The Tukey test compares the critical Tukey distance, $d_T$, against the absolute distance between each of the sample means. Any "pair" of differences greater than $d_T$ can be concluded as significant at the chosen $\alpha$ -level. In other words, if the difference of any "pair" of means exceeds $d_T$ , those two means are statistically different from each other.

**Table G-5. Example of Absolute differences of Sample Means**

| | $\bar{y}_1$ | $\bar{y}_2$ | $\cdots$ | $\bar{y}_n$ |
|---|---|---|---|---|
| $\bar{y}_1$ | 0 | | | |
| $\bar{y}_2$ | $\lvert \bar{y}_2 - \bar{y}_1 \rvert$ | 0 | | |
| : | | | | |
| $\bar{y}_n$ | $\lvert \bar{y}_n - \bar{y}_1 \rvert$ | $\lvert \bar{y}_n - \bar{y}_2 \rvert$ | | 0 |

**G.6.7 Correlation and cross-correlation matrices.** The correlation matrix of n random variables $X_1, \ldots, X_n$ is the n × n matrix whose i,j entry is $\mathrm{corr}(X_i, X_j)$. In probability theory and statistics, the term **cross-correlation** is also sometimes used to refer to the covariance cov(X, Y) between two random vectors X and Y, in order to distinguish that concept from the "covariance" of a random vector X, which is understood to be the matrix of covariances between the scalar components of X. It is commonly used to search a long duration signal for a shorter, known feature. For continuous functions *f* and *g* the cross-correlation is defined as**:**

$$(f \bullet g)(t) \stackrel{def}{=} \int_{-\infty}^{\infty} f^*(\tau)\, g(t+\tau)\, d\tau$$

where *f* * denotes the complex conjugate of *f*.

Similarly, for discrete functions, the cross-correlation is defined as**:**

$$(f \bullet g)(n) \stackrel{def}{=} \sum_{m=-\infty}^{\infty} f^*[m]\, g[n+m]$$

If $X$ and $Y$ are two independent random variables with probability distributions $f$ and $g$, respectively, then the probability distribution of the difference $X - Y$ is given by the cross-correlation $f \bullet g$. In contrast, the convolution $f * g$ gives the probability distribution of the sum $X + Y$. For example, consider two real valued functions $f$ and $g$ that differ only by a shift along the x-axis. One can calculate the cross-correlation to figure out how much $g$ must be shifted along the x-axis to make it identical to $f$. The formula essentially slides the $g$ function along the x-axis, calculating the integral of their product for each possible amount of sliding. When the functions match, the value of (f ● g) is maximized. The reason for this is that when lumps (positives areas) are aligned, they contribute to making the integral larger. Also, when the troughs (negative areas) align, they also make a positive contribution to the integral because the product of two negative numbers is positive. With complex-valued functions $f$ and $g$, taking the conjugate of $f$ ensures that aligned lumps (or aligned troughs) with imaginary components will contribute positively to the integral.[5]

**G.6.8 Classification and Regression Trees (CART).** CART builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification). In most general terms, the purpose of the analyses via tree-building algorithms is to determine a set of if-then logical (split) conditions that permit accurate prediction or classification of cases. Classification-type problems are generally those where one attempts to predict values of a categorical dependent variable (class, group membership, etc.) from one or more continuous and/or categorical predictor variables. Regression-type problems are generally those where one attempts to predict the values of a continuous variable from one or more continuous and/or categorical predictor variables. Tree classification techniques, when they "work" and produce accurate predictions or predicted classifications based on few logical if-then conditions, have a number of advantages over many of those alternative techniques. In most cases, the interpretation of results summarized in a tree is very simple. This simplicity is useful not only for purposes of rapid classification of new observations (it is much easier to evaluate just one or two logical conditions, than to compute classification scores for each possible group, or predicted values, based on all predictors and using possibly some complex nonlinear model equations), but can also often yield a much simpler "model" for explaining why observations are classified or predicted in a particular manner (e.g., when analyzing business problems, it is much easier to present a few simple if-then statements to management, than some elaborate equations).

Tree methods are nonparametric and nonlinear. The final results of using tree methods for classification or regression can be summarized in a series of (usually few) logical if-then

[5] www.en.wikipedia.org/wiki/Cross-correlation, 13 Jan 09. Wikipedia definition of Cross-correlation.

conditions (tree nodes). Therefore, there is no implicit assumption that the underlying relationships between the predictor variables and the dependent variable are linear, follow some specific non-linear link function [e.g., see Generalized Linear/Nonlinear Models (GLZ)], or that they are even monotonic in nature. For example, some continuous outcome variable of interest could be positively related to a variable Income if the income is less than some certain amount, but negatively related if it is more than that amount (i.e., the tree could reveal multiple splits based on the same variable Income, revealing such a non-monotonic relationship between the variables). Thus, tree methods are particularly well suited for data mining tasks, where there is often little a priori knowledge nor any coherent set of theories or predictions regarding which variables are related and how. In those types of data analyses, tree methods can often reveal simple relationships between just a few variables that could have easily gone unnoticed using other analytic techniques.

**G.6.9 Contour plots.** A contour plot is a graphical technique for representing a 3-dimensional surface by plotting constant z slices, called contours, on a 2-dimensional format. That is, given a value for z, lines are drawn for connecting the (x,y) coordinates where that z value occurs. The contour plot is formed by:

- Vertical axis: Independent variable 2
- Horizontal axis: Independent variable 1
- Lines: iso-response values

The independent variables are usually restricted to a regular grid. The actual techniques for determining the correct iso-response values are rather complex and are almost always computer generated. An additional variable may be required to specify the Z values for drawing the iso-lines. Some software packages require explicit values. Other software packages will determine them automatically. If the data (or function) do not form a regular grid, you typically need to perform a 2-D interpolation to form a regular grid.

**G.6.10 Response-surface methods**. Response surface methodology (RSM) explores the relationships between several explanatory variables and one or more response variables. The method was introduced by G. E. P. Box and K. B. Wilson in 1951. The main idea of RSM is to use a set of designed experiments to obtain an optimal response. Box and Wilson suggest using a first-degree polynomial model to do this. They acknowledge that this model is only an approximation, but use it because such a model is easy to estimate and apply, even when little is known about the process. An easy way to estimate a first-degree polynomial model is to use a factorial experiment or a fractional factorial designs. This is sufficient to determine which explanatory variables have an impact on the response variable(s) of interest. Once it is suspected that only significant explanatory variables are left, then a more complicated design, such as a central composite design can be implemented to estimate a second-degree polynomial model, which is still only an approximation at best. However, the second-degree model can be used to optimize (maximize, minimize, or attain a specific target for) a response.

**G.6.11 Advanced response-surface methodology.** Response surface methodology (RSM) is a collection of mathematical and statistical techniques that are useful for modeling and analysis of problems in which output or response influenced by several variables and the goal

is to find the correlation between the response and the variables. It can be used for optimizing the response. It is an empirical modelization technique devoted to the evaluation of relations existing between a group of controlled experimental factors and the observed results of one or more selected criteria.

**G.6.12 Nonparametric statistics.** If we have a basic knowledge of the underlying distribution of a variable, then we can make predictions about how, in repeated samples of equal size, this particular statistic will "behave," that is, how it is distributed. For many variables of interest, we simply do not know for sure the distribution. Nonparametric methods were developed to be used in cases when the researcher knows nothing about the parameters of the variable of interest in the population (hence the name nonparametric). In more technical terms, nonparametric methods do not rely on the estimation of parameters (such as the mean or the standard deviation) describing the distribution of the variable of interest in the population. Therefore, these methods are also sometimes (and more appropriately) called parameter-free methods or distribution-free methods.

**G.6.12.1 Kruskal-Wallis H Test.** The Kruskal-Wallis H test is a non-parametric test to compare k populations. The test may be described as:

$H_o$: The *k* populations are identical
$H_a$: At least two of the populations differ in location

Test Statistic: $$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

where:

$n_i$ = Number of measurements in sample from population *i*
$R_i$ = Rank sum for sample *i*, where the rank of each measurement is computed according to its relative size in the overall set of $n = n_1 + n_2 + \ldots n_k$ observations formed by combining the data from all *k* samples.

Rejection Region: Reject $H_o \; if \; H > \chi^2_\alpha$ with (k-1) degrees of freedom

Assumptions: The *k* samples are randomly and independently drawn. There are five or more measurements in each sample.

**G.6.13 Recursive Partitioning.** A statistical technique that "partitions data according to a relationship between the [independent variables and the dependent variables], creating a tree of partitions. [The technique] finds a set of cuts or grouping of [independent variables] that best predict [the independent variable]… [using] a tree of decision rules until the desired fit is reached." There are many variations of this technique. (SAS Institute, 2007).

Recursive partitioning is an exploratory technique for uncovering structure in data. Models are constructed by successively splitting a dataset into increasingly homogeneous subsets until it is infeasible to continue, based on a set of "stopping rules." The way the dataset is split is determined by the available covariates (independent variables). At each stage, all the covariates are examined and the one that gives the "best" split (e.g., greatest differences between groups) is chosen first. Each iteration of partitioning will select the best split to form a partitioning tree.

Covariates can consist of continuous data or categorical data. Continuous data is normally evaluated as a sample variance t-test. Categorical data may be evaluated as binary data $\hat{p}\left(1-\hat{p}\right)$ or multiple class data using Pearson's chi-square statistic.

**G.6.14 Neural networks.** Neural networks are very sophisticated modeling techniques capable of modeling extremely complex functions. In particular, neural networks are nonlinear (a term which is discussed in more detail later in this section). For many years linear modeling has been the commonly used technique in most modeling domains since linear models have well-known optimization strategies. Where the linear approximation was not valid (which was frequently the case) the models suffered accordingly. Neural networks also keep in check the curse of dimensionality problem that bedevils attempts to model nonlinear functions with large numbers of variables. Neural networks learn by example. The neural network user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data.

An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not exist in the taught list of input patterns, a firing rule is used to determine whether to fire or not.

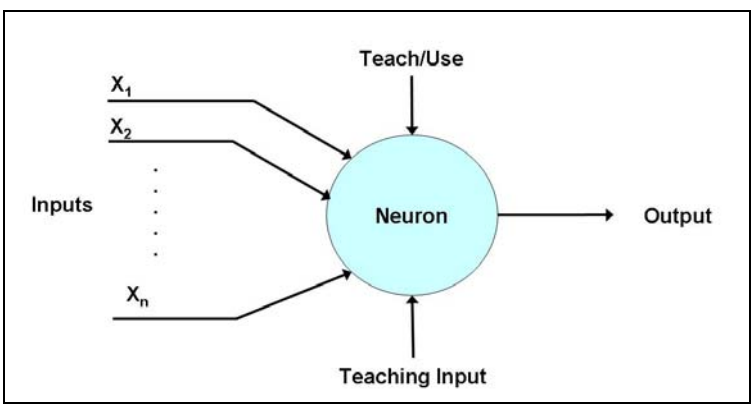



**Figure G-11. Simple neuron**

For example, a 3-input neuron is taught to output the value "1" when the input (X1,X2 and X3) is 111 or 101 and to output value "0" when the input is 000 or 001. Then, before applying a firing rule, the truth table G-6 is:

**Table G-6. Example Neuron Input-Output**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X1: | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| X2: | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| X3: | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| Output: | 0 | 0 | 0/1 | 0/1 | 0/1 | 1 | 0/1 | 1 |

A firing rule or set of rules may exist which determine the output when a taught input pattern is not present. A rule may be as simple as, "when a taught input set is not present, then output value 0".

A more sophisticated neuron (figure G-12) is the McCulloch and Pitts model (MCP). The difference from the previous model is that the inputs are *weighted*; the effect that each input has at decision making is dependent on the weight of the particular input. The weight of an input is a number which when multiplied with the input gives the weighted input. These weighted inputs are then added together and if they exceed a pre-set threshold value, the neuron fires. In any other case the neuron does not fire.

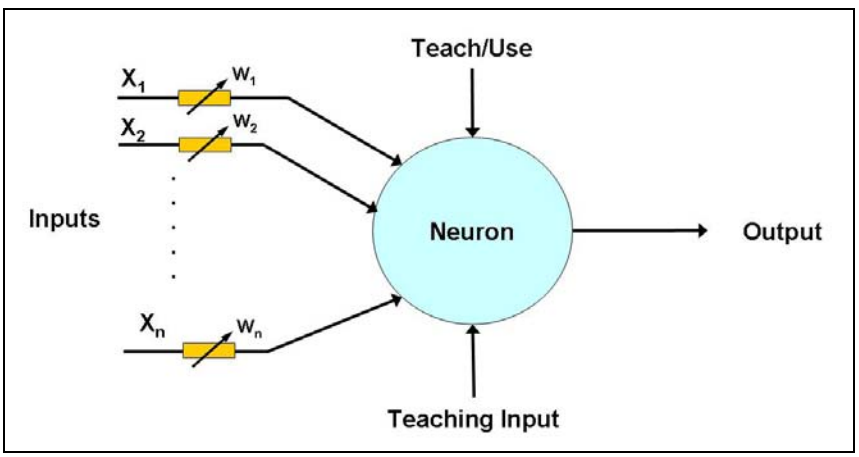



**Figure G-11. MCP neuron**

In mathematical terms, the neuron fires if and only if: $X_1W_1 + X_2W_2 + \cdots + X_nW_n > T$

Although the user does need to have some heuristic knowledge of how to select and prepare data, how to select an appropriate neural network, and how to interpret the results, the level of user knowledge needed to successfully apply neural networks is much lower than would be the case using (for example) some more traditional nonlinear statistical methods.

**G.6.15 Half Normal Probability Plot.** The half-normal probability plot answers the question: What are the important factors (including interactions) that impact a measure?

Related to this is a second question of a rank ordered list of important factors. Quantitatively, the estimated effect of a given main effect or interaction and its rank relative to other main effects and interactions is given via least squares estimation (that is, forming effect estimates that minimize the sum of the squared differences between raw data and the fitted values from such estimates). With those estimates, one can then construct a list of the main effects and interactions ordered by the effect magnitude. The half-normal probability plot is a graphical tool that uses these ordered estimated effects to help assess which factors are important and which are unimportant.

The half-normal probability plot can be used when for two level designs (where each factor has two discrete values). It is especially useful when the design is saturated and there are no degrees of freedom (i.e., n=8 runs and k=7 factors; d.f. = n-k-1). It is the most commonly-employed procedure for identification of important factors in 2-level full and factorial designs.

A half-normal distribution is the distribution of the |X| with X having a normal distribution. The outputs from the half-normal probability plot are grouping of factors and interactions into two categories: important and unimportant. For full factorial designs, interactions include the full complement of interactions of all orders; for fractional factorial designs, interactions are only some, and occasionally none, of the actual interactions (when they aren't estimable). A secondary output is a ranked list of factors and interactions from most important down to least important.

A half-normal probability plot is formed by an ordered (largest to smallest) absolute value of the estimated effects for the main factors and available interactions on the y-axis. If n data points (no replication) have been collected, then typically (n-1) effects will be estimated and the (n-1) effects will be plotted. On the x-axis is (n-1) theoretical order statistic medians from a half-normal distribution. These (n-1) values are not data-dependent. They depend only on the half-normal distribution and the number of items plotted (n-1). The theoretical medians represent an "ideal" typical ordered data set that would have been obtained from a random drawing of (n-1) samples in a half-normal distribution. If the design is a fractional factorial, the confounding structure is provided for main effects and 2-factor interactions. The basis for the half-normal probability plot comes from the normal probability plot (which also finds frequent use in 2-level designs). The normal probability plot estimated effects are the optimal least squares estimates. Because of the orthogonality of the 2k full factorial and the 2k-p fractional factorial designs, all least squares estimators for main effects and interactions simplify to the form:

$$\text{Estimated effect} = \bar{A}(+) - \bar{A}(-)$$

with $\bar{A}$ (+) the average of all response values for which the factor or interaction takes on a "+" value, and where $\bar{A}$ (-) is the average of all response values for which the factor or interaction takes on a "-" value. The sign of each estimate is completely arbitrary and will reverse depending on how the initial assignments were made (e.g., we could assign "-" to treatment A and "+" to treatment B or just as easily assign "+" to treatment A and "-" to treatment B). This arbitrariness is addressed by dealing with the effect magnitudes rather

than the signed effects. If the signed effects follow a normal distribution, the absolute values of the effects follow a half-normal distribution.

Thus, one tests for important versus unimportant factors by generating a half-normal probability plot of the absolute value of the effects. Linearity implies half-normality, which in turn implies all factors are unimportant. Unimportant factors are those that have near-zero effects and important factors are those whose effects are considerably removed from zero. Thus, unimportant effects tend to have a normal distribution centered near zero while important effects tend to have a normal distribution centered at their respective true large (but unknown) effect values. More typically, the half-normal probability plot will be only partially linear. Unimportant (that is, near-zero) effects manifest themselves as being near zero and on a line while important (that is, large) effects manifest themselves by being off the line and well-displaced from zero.

An example half-normal plot for defective springs data is provided figure G-13. [6]

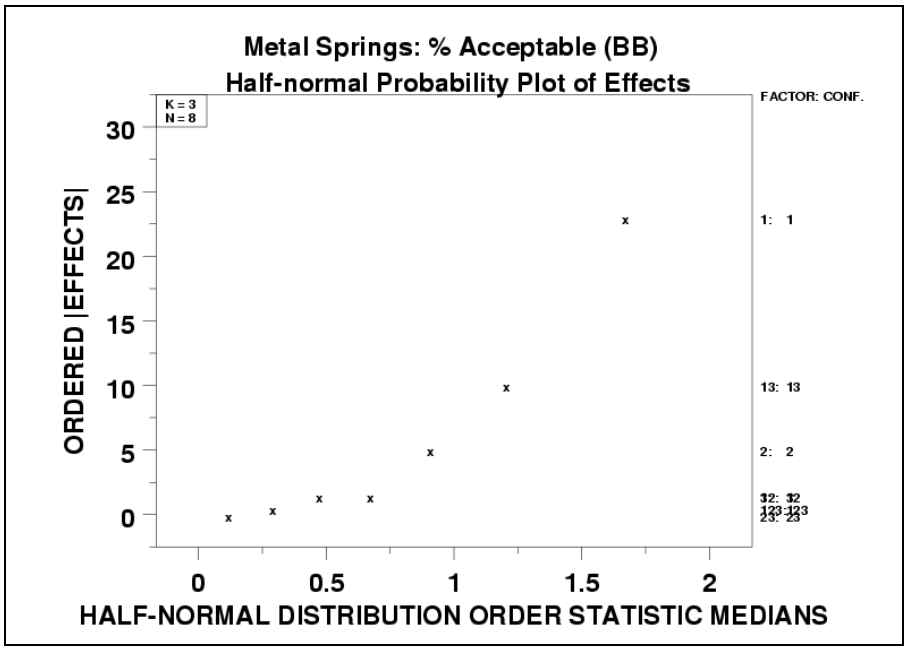



**Figure G-13. Illustrative Half Normal Distribution**

The unimportant factors will typically be on or close to a near-zero line, while the estimated effect of an important factor will typically be displaced well off the line. A rough rule-of-thumb would be to declare as important those factors/interactions whose far-right labels are easy to distinguish and to declare as unimportant those factors/interactions whose far-right labels are overwritten (crowded together) and hard to distinguish. The ranking of factors is simply an ordered list on the vertical y-axis from top to bottom with the most important at the top. In this example, factor X1 is most important, factor X1X2 second, and X2 third. All other factors are considered unimportant.

[6] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 7/18/2006

In summary, the signs of the estimated effects are arbitrary allowing the use of half-normal probability plot of effects over the normal probability plot of effects. These probability plots are commonly-employed procedures for identification of important factors in 2-level full and factorial designs.

**G.6.15 Stepwise Regression.** The Stepwise regression is an approach to selecting a subset of effects for a regression model. It is used when there is little theory to guide the selection of terms for a model and the modeler, in desperation, wants to use whatever seems to provide a good fit.

The approach is somewhat controversial. The significance levels on the statistics for selected models violate the standard statistical assumptions because the model has been selected rather than tested within a fixed model. On the positive side, the approach has been of practical use for 30 years in helping to trim out models to predict many kinds of responses. The book Subset Selection in Regression, by A. J. Miller (1990), brings statistical sense to model selection statistics.

Stepwise regression is a special case of forward selection: in addition to the steps performed in the forward selection algorithm, all variables are tested if their contribution is significant after a new variable has been added. This may lead to the elimination of an already selected variable if this variable has become superfluous because of its relationship to the other variables.

**Algorithm:**

1. Calculate the correlations of all independent variables, $X_i$, with the response variable Y. Use the variable with the highest correlation as the starting variable.
2. Add the variable with the highest partial F value.
3. Check all variables of the current model for their partial F values and remove any variable which falls below a predefined threshold.
4. Repeat the procedure with step 2 until some stopping criterion is met.

*This page intentionally left blank.*

# ENCLOSURE 1 TO ANNEX G - GLOSSARY OF ANALYTICAL TERMS

This is a lexicon with brief definitions of some key technical, mathematical, and statistical terms used in the *Analyst's Handbook for Testing in a Joint Environment* (Analyst's Handbook) and in some of its annexes. This is not meant to be exhaustive in these fields, and many such lists are available online. For example good and extensive glossaries of statistical terms are offered by the University of California, Berkeley (http://www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm), and by Youngstown State University (http://cc.ysu.edu/~eeusip/glossary.htm).

Where appropriate, terms in the descriptions are hyperlinked to other entries in the list. The hyperlinks are only applied to the first use of a particular term within a single definition. Subsequent uses of the term within the same definition are not hyperlinked.

There are two lists of definitions in this appendix. The first in Section 1, Pedagogical Ordering by Category, is a listing of terms by category. Within each category, the terms are listed in the logical order in which they should be read to understand their meaning (pedagogical order). The second listing in Section 2, Comprehensive Alphabetical Listing, is the entire list of terms in alphabetical order without regard to category. The definitions in both lists are identical; however, hyperlinks in the second list (alphabetical) will move the reader back to the first list to facilitate an understanding of the term in context.

## 1. Pedagogical Ordering by Category

### 1.1 General

- **Model**. A set of assumptions and approximations about how a process or system works, whether that process or system currently exists or not. There are many different kinds of models, including simulation, optimization, regression, and others. The assumptions and approximations should be reasonably accurate depictions of reality for the model to be valid.
- **Simulation**. A numerical exercising of a mathematical model used when an exact mathematical analysis of the model is not available. Typically provides estimates/approximations rather than exact results, but does not require perhaps-unrealistic strong assumptions that would be necessary to arrive at a mathematically tractable model, so allows for more complex models that are typically more valid than over-simplified models created for mathematical tractability.
- **Heuristic**. An approximate algorithm or method used when an exact mathematical approach is either unavailable, or would be too difficult or time-consuming to apply to a given problem.
- **Data Farming.** "Data Farming is the process of using a high performance computer or computing grid to run a simulation thousands or millions of times across a large parameter and value space. The result of data farming is a "landscape" of output that can be analyzed for trends, anomalies, [emergent behavior,] and insights in multiple parameter dimensions… [This] term comes from the idea of planting data in the

simulation and parameter/value space, and then harvesting the data that result from the simulation runs." ("Data Farming," 2008)

## 1.2 General Mathematical

- **Function**. A formula or rule that transforms one or several numbers (independent variables) into another (single) number (dependent variable). For each set of independent variables, a function can produce only one value of the dependent variable.
- **Variable**. A quantity that can take on a variety of different values, possibly within some range.
- **Independent Variable**. A variable that is free to vary on its own, possibly within some range, or within a range that is influenced by other independent variables. What is "fed into" a function.
- **Dependent Variable**. The "output" result of a function into which independent-variable values are "fed."
- **Continuous Variable**. A variable that can assume any real value, possibly within some range bounded on one or both sides. The number of possible values is always infinite.
- **Discrete Variable**. A variable that can assume only certain separated values in a list, which could be a finite or infinite list.
- **Binary Variable**. A variable that can take on one of only two values. Could be qualitative or quantitative, though often "coded" as 0 or 1.
- **Indicator Variable**. A binary variable that is actually a function of whether some other condition is true, and takes on the value of 1 if that condition is true, and takes on the value 0 if that condition is false.
- **Parameter**. (1) A quantity, usually regarded as a constant, that is part of the formula for a function. (Also, used in statistics to mean a number that is a characteristic of an entire population or distribution or process, and thus cannot be known with certainty from any finite sample.) (2) A quantity treated as an input to a model or simulation. It can be either known (e.g., number of personnel in a unit) or unknown (a distribution/population parameter). Model parameters may be changed as factors within an experimental design are varied. See also distribution/population parameter.
- **Causality**. A phenomenon when one condition or event directly results in another condition or event through an identifiable chain. Just because two values seem to be correlated, so tend to rise and fall together (or oppositely), does not mean that one is causing the other; for one thing, there could be a third value that is causing them both.
- **Stochastic**. Random, uncertain.
- **Uncertainty**. The quality that not everything is known exactly and with certainty.
- **Deterministic**. A value or model in which all aspects are assumed to be known with certainty, and without any uncertainty or possible ranges.

- **Linear Function**. A function is linear when it can be written in the form: $Y = c_0 + c_1X_1+c_2X_2+\ldots+c_kX_k$, where Y is the dependent variable, the terms $c_0$, $c_1$, $c_2$, etc. are coefficients involving the *k* independent variables.
- **Nonlinear Function.** A function that cannot be expressed in the form of a linear function.
- **Quadratic Terms.** Independent variables in a non-linear function that are squared.

## 1.3 Probability Terms

- **Sample Space**. The set of all possible values for the outcome of some random or uncertain activity. Subsets of the sample space are events, which have probabilities between 0 and 1 (inclusive on both ends).
- **Random Variable**. Mathematically, a numerical-valued function of points in the sample space. Practically speaking, a random variable is a number whose value depends on some random or uncertain activity, so we don't know its value before the activity; however, we might well know something about where it is likely and unlikely to be, as specified by its probability distribution.
- **Discrete Random Variable**. A random variable that can take on only certain separated values that can be listed, either finitely or infinitely. For example, the number of hits in ten shots would be a discrete random variable on the finite set of integers {0, 1, 2, …, 10}. The number of shots necessary to get the first hit would be a discrete random variable on the infinite set of integers {1, 2, 3, …}.
- **Continuous Random Variable**. A random variable that can take on any real number, possibly within a limited range. The number of possible values that a continuous random variable can take on is always infinite. For instance, the time to repair a piece of equipment would be a continuous random variable with positive values, possibly bounded above is there is an upper repair-time limit allowable before just scrapping the item.
- **Probability Distribution of a Random Variable**. A description of the relative probabilities of the values/ranges on/in which a random variable might fall. Could be expressed as a mathematical formula, a numerical table, or a graph.
- **Distribution/Population Parameter**. A number that helps characterize the population/distribution from which a sample is taken. In practice, a parameter is usually unknown so is the subject of estimation with a point estimator or confidence interval.
- **Point Estimator**. A formula into which observed data values are fed to produce an estimate of a distribution/population parameter. Once the values have been fed in, the resulting number is a *point estimate*. By themselves, point estimates are of limited value since there is no way to know how far away from the parameter they are likely to be; it is much better to combine a point estimate with a confidence interval to give some measure of precision to the point estimate.
- **Expectation**. The "average" value of a random variable, weighting its possible values by their relative probabilities of occurrence (from the random variable's probability distribution). The mathematical definition is different for discrete and continuous

random variables. Can be thought of as the usual arithmetic average of a very large (technically, infinite) number of observations on the random variable. In a sample of observations from a random variable, the expectation is point-estimated by the usual arithmetic average.

- **Mean**. Average, and is used both for expectation and sample average.
- **Variability**. The property that a quantity, usually modeled as a random variable, does not always take on the same value, depending on chance circumstances.
- **Variance**. A measure of variability, defined as the expected value of the square of the difference between a random variable and its own expectation. The units of measurement of the variance are the square of the random variable's (or data values') unit of measurement, which may or may not make sense (meters to square meters, but dollars to square dollars?). Estimated from a sample by the usual sample-variance formula.
- **Standard Deviation**. The (positive) square root of the variance. Units of measurement are the same as those for the random variable (or data values).

## 1.4 Statistics Terms

- **Correlation**. A measure of the degree of *linear* dependence between two variables, either random variables or variables in a data set. Correlations are always between –1 and +1, with the extremes' being perfect linear correlation. If the correlation is positive, the two variables tend to rise and fall together; if the correlation is negative one variable tends to rise while the other one falls, and vice-versa. Correlation is often over-interpreted in two ways. First, presence of correlation does not imply that one variable's movement is *causing* the other variable to move; both could be caused by some third variable, for instance. Second, zero correlation does not imply that there is no relationship at all between the variables, just that there is no *linear* relation; for instance, plot points on a symmetric parabola (or a smiley face) and see that their $x$ and $y$ coordinates are nearly uncorrelated, yet there is obviously a relationship.
- **Autocorrelation**. Also called *serial correlation*, the correlation between observations in a time series at different points in time. The time difference, or *lag*, between the two points can affect the sign and strength of the correlation. The autocorrelation between two points separated by a certain lag may or may not depend on where in time the points are; if not, the time series is called *covariance-stationary*.
- **Statistical Significance**. The property that an observation or value cannot reasonably be explained by mere chance, so that it is "real" beyond any natural variability that might be present. A statistically significant observation means that it is very unlikely that this could have happened by mere chance across different samples that might have been taken. In signal-processing terms, a statistically significant observation is one in which the signal component rises well above the noise component.
- **Statistical Independence**. The property that two variables or events have nothing to do with each other. Equivalently, knowing the value of one of the variables (or the occurrence of one of the events) tells you nothing about the value of the other variable (or the occurrence of the other event). Independence greatly simplifies mathematical manipulations (since joint probabilities factor into the product of marginal probabilities),

but assuming independence in situations where it is actually not physically true can lead to incorrect results (this is sometimes called the declaration of independence, since that is pretty much all it is).

- **Confidence Interval**. A range of values that contains (or covers) an unknown population or distribution parameter with pre-specified probability, typically set at 0.90 or higher. The values in a confidence interval can be thought of as "reasonable" values for the unknown parameter, in the sense that a two-sided hypothesis test for equality of the parameter to a particular number would not be rejected (at significance level equal to one minus the confidence level of the confidence interval) if and only if that number is within the confidence interval. A multi-dimensional generalization is a *confidence region* when several different parameters are involved.

- **Regression**. A statistical method in which a dependent variable (or response) is assumed to depend on the values of other independent variables (or regressors). Regression then estimates the nature of this dependence, as well as assesses whether individual independent variables are statistically significant predictors of the dependent variable. The form of the regression model (linear, interactions, quadratic, etc.) is often developed with the help of plots of the dependent variable vs. the various possible independent variables.

- **Regressor**. In regression, one of the independent variables on which the dependent variable might actually depend in some way.

- **Multicollinearity**. In regression, the property that one or more regressor independent variables are heavily linearly correlated with others, either positively or negatively. Causes problems in regression in terms of accuracy, stability, and interpretation of results. If multicollinearity is present, the regression coefficients cannot be interpreted as the effect of one independent variable while holding all others constant, because the other independent variables with which one is collinear cannot be held constant while the one changes. Can be detected by computing a correlation matrix of all regressors with each other (but not with the dependent variable); if found, all but one of the regressors in a multicollinear group should be dropped, since they are not carrying significantly additional information.

- **ANOVA**. Common acronym for *Analysis of Variance*, a formal statistical method used to test whether the means of several different populations are all the same as each other, or if there is at least one mean that differs significantly from at least one other mean. If it appears that not all means are equal to each other, a follow-up analysis, *multiple comparisons*, is often done to identify which means differ significantly from which other means, and by how much and in which direction. There are many variations and generalizations of ANOVA, discussed in any statistics book, and usually available in statistical software. The name is somewhat of a misnomer, as we are analyzing means, not variances; however, this analysis is carried out via formulas that are similar to sample-variance formulas (weighted sums of squared differences).

- **Recursive Partitioning.** A statistical technique that "partitions data according to a relationship between the [independent variables and the dependent variables], creating a tree of partitions. [The technique] finds a set of cuts or grouping of [independent

variables] that best predict [the independent variable]… [using] a tree of decision rules until the desired fit is reached." There are many variations of this technique. (SAS Institute, 2007).

- **Classification Tree.** A statistical techniques that involves a recursive partitioning for discrete/categorical variables.
- **Regression Tree.** A statistical techniques that involves a recursive partitioning for continuous variables.
- **Descriptive Statistics**. Several techniques that simply use the data from a sample to describe the sample or the population from which it was drawn, without doing any formal statistical inference (confidence intervals, hypothesis tests, regression, etc.). Examples include sample means, sample standard deviations, minima/maxima, and histograms.
- **Nonparametric Statistics**. A collection of statistical techniques that studiously avoid the assumption that the data come from a normally (Gaussian) distributed population. Used when histograms of the data reveal extreme skewness. Typically based on ranks or ordering of sorted data values.
- **Parametric Statistics**. "Standard" statistical methods that, unlike nonparametric statistics, *do* assume that the data come from a normal population.

## 1.5 Design of Experiments

- **Design of experiments (DOE) / experimental design**. The careful planning of exactly which combinations of experimental conditions (factors) are used in an experiment, either real or simulated. The goal is to enable reasonable estimates of factor main effects and interactions with the fewest number of design points possible.
- **Factor**. An input value, or independent variable, in a design of experiments. Could be quantitative or qualitative, discrete or continuous, controllable or uncontrollable in reality, and could take on as few as two but many more levels. Typically we want to identify which factors are important for which responses, and how they affect those responses.
- **Levels**. The values, either numerical or qualitative/categorical, that a factor in a design of experiments can take on. Choosing the levels for a factor must involve knowledge of the real situation, and often involves judgment. The simplest designs involve just two levels for a factor, but the more levels for a factor there are, the more information can be obtained about that factor's effect, such as nonlinear effects on the response; of course, having more levels implies more design points, and thus more effort.
- **Design Points**. The factor-level combinations at which a design of experiments will be run. At each design point, the experiment is run, perhaps replicated multiple times. The more design points, the more information can be obtained about factors' main effects on responses and the interactions between factors, but the more effort is involved. The particular design of experiments used determines the number of, and factor levels for, the design points.
- **Response**. The output or dependent variable in a design of experiments, typically a measure of performance (MOP) or a measure of effectiveness (MOE). It is usually of

interest to establish which factors have significant main effects on responses, and which factors might interact in their effect on responses. There could be multiple responses.

- **Main Effect**. In a design of experiments, the effect that a particular factor has, on its own, on a response. Typically, in a linear model, the main effect would be interpreted as the amount the response changes if this factor is increased by one unit. If there are interaction effects between this and other factors, the interpretation of the main effect is not this simple, since the values of the other factors can have an impact on the effect of this factor. Also, in nonlinear models, a factor's effect might vary depending on its level.
- **Interaction Effect**. In a design of experiments, a measure of the degree to which the level of one factor has an impact on the effect of another factor (also applies to multiple factors, where the levels of one group of factors have an impact on the effects of another group of factors). If there is no significant interaction between two factors, then the main effects of these factors can be interpreted as how they change the response on their own. Interactions can be two-way, three-way, etc., depending on the number of factors involved.
- **Replication**. Repeating the experiment in a way that is identical to, but independent of, past experiments. In a design of experiments, replications could be made at individual design points, or of the entire design itself.
- **Screening**. A type of design of experiments intended to identify which factors have a significant impact on a response and which do not. Those factors that appear not to matter can be dropped from further consideration, often greatly reducing further experimental effort if the original number of factors was high.
- **Randomization**. Randomly assigning factor-level "treatments" to "experimental units," e.g., randomly assigning which experimental unit gets which kind of weapon. This is meant to break up any systematic bias that could result from systematic assignment (e.g., the weapons are numbered best to worst, as are the units based upon training level). Generally not necessary in simulation experiments, provided that a good underlying random-number generator is used.

## 1.6 Factors in Design of Experiments

- **Range**. The possible values that a factor in a design of experiments can assume. A list of discrete values for a discrete factor, and a continuum of real numbers (possibly bounded on one or both ends) for a continuous factor.
- **Qualitative**. A factor in a design of experiments that is not naturally a number, such as presence or not of a particular weapons package. Can be (artificially) made quantitative by assigning numbers to different categories (e.g., 1 if the training level of an individual is low, 2 if medium, or 3 if high).
- **Quantitative**. A factor in a design of experiments that naturally takes on a numerical value with some scale of measurement, either discrete or continuous.
- **Controllable**. (1) A factor in a design of experiments that, in reality, can be controlled by a decision maker. For example, whether to include a particular weapons package, or how fast to operate a vehicle, would be controllable factors. (2) A factor that can be

controlled during experimentation (either live or simulation, or both). For example, the number of weapons assigned to a unit for a live experiment.

- **Uncontrollable**. (1) A factor in a design of experiments that, in reality, cannot be controlled by a decision maker. For example, the degree of enemy jamming of communications, or weather conditions, would be uncontrollable factors. Though uncontrollable in practice, such factors are often made controllable some way in a simulation or live test. (2) A factor that cannot be controlled during experimentation (either live or simulation, or both). For example, the relative humidity during a live experiment.

## 2. Comprehensive Alphabetical Listing

- **ANOVA**. Common acronym for *Analysis of Variance*, a formal statistical method used to test whether the means of several different populations are all the same as each other, or if there is at least one mean that differs significantly from at least one other mean. If it appears that not all means are equal to each other, a follow-up analysis, *multiple comparisons*, is often done to identify which means differ significantly from which other means, and by how much and in which direction. There are many variations and generalizations of ANOVA, discussed in any statistics book, and usually available in statistical software. The name is somewhat of a misnomer, as we are analyzing means, not variances; however, this analysis is carried out via formulas that are similar to sample-variance formulas (weighted sums of squared differences).
- **Autocorrelation**. Also called *serial correlation*, the correlation between observations in a time series at different points in time. The time difference, or *lag*, between the two points can affect the sign and strength of the correlation. The autocorrelation between two points separated by a certain lag may or may not depend on where in time the points are; if not, the time series is called *covariance-stationary*.
- **Binary Variable**. A variable that can take on one of only two values. Could be qualitative or quantitative, though often "coded" as 0 or 1.
- **Causality**. A phenomenon when one condition or event directly results in another condition or event through an identifiable chain. Just because two values seem to be correlated, so tend to rise and fall together (or oppositely), does not mean that one is causing the other; for one thing, there could be a third value that is causing them both.
- **Classification Tree.** A statistical techniques that involves a recursive partitioning for discrete/categorical variables.
- **Confidence Interval**. A range of values that contains (or covers) an unknown population or distribution parameter with pre-specified probability, typically set at 0.90 or higher. The values in a confidence interval can be thought of as "reasonable" values for the unknown parameter, in the sense that a two-sided hypothesis test for equality of the parameter to a particular number would not be rejected (at significance level equal to one minus the confidence level of the confidence interval) if and only if that number is within the confidence interval. A multi-dimensional generalization is a *confidence region* when several different parameters are involved.

- **Continuous Random Variable**. A random variable that can take on any real number, possibly within a limited range. The number of possible values that a continuous random variable can take on is always infinite. For instance, the time to repair a piece of equipment would be a continuous random variable with positive values, possibly bounded above is there is an upper repair-time limit allowable before just scrapping the item.
- **Continuous Variable**. A variable that can assume any real value, possibly within some range bounded on one or both sides. The number of possible values is always infinite.
- **Controllable**. (1) A factor in a design of experiments that, in reality, can be controlled by a decision maker. For example, whether to include a particular weapons package, or how fast to operate a vehicle, would be controllable factors. (2) A factor that can be controlled during experimentation (either live or simulation, or both). For example, the number of weapons assigned to a unit for a live experiment.
- **Correlation**. A measure of the degree of *linear* dependence between two variables, either random variables or variables in a data set. Correlations are always between –1 and +1, with the extremes' being perfect linear correlation. If the correlation is positive, the two variables tend to rise and fall together; if the correlation is negative one variable tends to rise while the other one falls, and vice-versa. Correlation is often over-interpreted in two ways. First, presence of correlation does not imply that one variable's movement is *causing* the other variable to move; both could be caused by some third variable, for instance. Second, zero correlation does not imply that there is no relationship at all between the variables, just that there is no *linear* relation; for instance, plot points on a symmetric parabola (or a smiley face) and see that their $x$ and $y$ coordinates are nearly uncorrelated, yet there is obviously a relationship.
- **Data Farming.** "Data Farming is the process of using a high performance computer or computing grid to run a simulation thousands or millions of times across a large parameter and value space. The result of data farming is a "landscape" of output that can be analyzed for trends, anomalies, [emergent behavior,] and insights in multiple parameter dimensions… [This] term comes from the idea of planting data in the simulation and parameter/value space, and then harvesting the data that result from the simulation runs." ("Data Farming," 2008)
- **Dependent Variable**. The "output" result of a function into which independent-variable values are "fed."
- **Descriptive Statistics**. Several techniques that simply use the data from a sample to describe the sample or the population from which it was drawn, without doing any formal statistical inference (confidence intervals, hypothesis tests, regression, etc.). Examples include sample means, sample standard deviations, minima/maxima, and histograms.
- **Design of experiments (DOE) / experimental design**. The careful planning of exactly which combinations of experimental conditions (factors) are used in an experiment, either real or simulated. The goal is to enable reasonable estimates of factor main effects and interactions with the fewest number of design points possible.

- **Design Points**. The factor-level combinations at which a design of experiments will be run. At each design point, the experiment is run, perhaps replicated multiple times. The more design points, the more information can be obtained about factors' main effects on responses and the interactions between factors, but the more effort is involved. The particular design of experiments used determines the number of, and factor levels for, the design points.
- **Deterministic**. A value or model in which all aspects are assumed to be known with certainty, and without any uncertainty or possible ranges.
- **Discrete Random Variable**. A random variable that can take on only certain separated values that can be listed, either finitely or infinitely. For example, the number of hits in ten shots would be a discrete random variable on the finite set of integers {0, 1, 2, …, 10}. The number of shots necessary to get the first hit would be a discrete random variable on the infinite set of integers {1, 2, 3, …}.
- **Discrete Variable**. A variable that can assume only certain separated values in a list, which could be a finite or infinite list.
- **Distribution/Population Parameter**. A number that helps characterize the population/distribution from which a sample is taken. In practice, a parameter is usually unknown so is the subject of estimation with a point estimator or confidence interval.
- **Expectation**. The "average" value of a random variable, weighting its possible values by their relative probabilities of occurrence (from the random variable's probability distribution). The mathematical definition is different for discrete and continuous random variables. Can be thought of as the usual arithmetic average of a very large (technically, infinite) number of observations on the random variable. In a sample of observations from a random variable, the expectation is point-estimated by the usual arithmetic average.
- **Factor**. An input value, or independent variable, in a design of experiments. Could be quantitative or qualitative, discrete or continuous, controllable or uncontrollable in reality, and could take on as few as two but many more levels. Typically we want to identify which factors are important for which responses, and how they affect those responses.
- **Function**. A formula or rule that transforms one or several numbers (independent variables) into another (single) number (dependent variable). For each set of independent variables, a function can produce only one value of the dependent variable.
- **Heuristic**. An approximate algorithm or method used when an exact mathematical approach is either unavailable, or would be too difficult or time-consuming to apply to a given problem.
- **Independent Variable**. A variable that is free to vary on its own, possibly within some range, or within a range that is influenced by other independent variables. What is "fed into" a function.

- **Indicator Variable**. A binary variable that is actually a function of whether some other condition is true, and takes on the value of 1 if that condition is true, and takes on the value 0 if that condition is false.
- **Interaction Effect**. In a design of experiments, a measure of the degree to which the level of one factor has an impact on the effect of another factor (also applies to multiple factors, where the levels of one group of factors have an impact on the effects of another group of factors). If there is no significant interaction between two factors, then the main effects of these factors can be interpreted as how they change the response on their own. Interactions can be two-way, three-way, etc., depending on the number of factors involved.
- **Levels**. The values, either numerical or qualitative/categorical, that a factor in a design of experiments can take on. Choosing the levels for a factor must involve knowledge of the real situation, and often involves judgment. The simplest designs involve just two levels for a factor, but the more levels for a factor there are, the more information can be obtained about that factor's effect, such as nonlinear effects on the response; of course, having more levels implies more design points, and thus more effort.
- **Linear Function**. A function is linear when it can be written in the form: $Y = c_0 + c_1X_1+c_2X_2+\ldots+c_kX_k$, where Y is the dependent variable, the terms $c_0$, $c_1$, $c_2$, etc. are coefficients involving the *k* independent variables.
- **Main Effect**. In a design of experiments, the effect that a particular factor has, on its own, on a response. Typically, in a linear model, the main effect would be interpreted as the amount the response changes if this factor is increased by one unit. If there are interaction effects between this and other factors, the interpretation of the main effect is not this simple, since the values of the other factors can have an impact on the effect of this factor. Also, in nonlinear models, a factor's effect might vary depending on its level.
- **Mean**. Average, and is used both for expectation and sample average.
- **Model**. A set of assumptions and approximations about how a process or system works, whether that process or system currently exists or not. There are many different kinds of models, including simulation, optimization, regression, and others. The assumptions and approximations should be reasonably accurate depictions of reality for the model to be valid.
- **Multicollinearity**. In regression, the property that one or more regressor independent variables are heavily linearly correlated with others, either positively or negatively. Causes problems in regression in terms of accuracy, stability, and interpretation of results. If multicollinearity is present, the regression coefficients cannot be interpreted as the effect of one independent variable while holding all others constant, because the other independent variables with which one is collinear cannot be held constant while the one changes. Can be detected by computing a correlation matrix of all regressors with each other (but not with the dependent variable); if found, all but one of the regressors in a multicollinear group should be dropped, since they are not carrying significantly additional information.

- **Nonlinear Function.** A function that cannot be expressed in the form of a linear function.
- **Nonparametric Statistics**. A collection of statistical techniques that studiously avoid the assumption that the data come from a normally (Gaussian) distributed population. Used when histograms of the data reveal extreme skewness. Typically based on ranks or ordering of sorted data values.
- **Parameter**. (1) A quantity, usually regarded as a constant, that is part of the formula for a function. (Also, used in statistics to mean a number that is a characteristic of an entire population or distribution or process, and thus cannot be known with certainty from any finite sample.) (2) A quantity treated as an input to a model or simulation. It can be either known (e.g., number of personnel in a unit) or unknown (a distribution/population parameter). Model parameters may be changed as factors within an experimental design are varied. See also distribution/population parameter.
- **Parametric Statistics**. "Standard" statistical methods that, unlike nonparametric statistics, *do* assume that the data come from a normal population.
- **Point Estimator**. A formula into which observed data values are fed to produce an estimate of a distribution/population parameter. Once the values have been fed in, the resulting number is a *point estimate*. By themselves, point estimates are of limited value since there is no way to know how far away from the parameter they are likely to be; it is much better to combine a point estimate with a confidence interval to give some measure of precision to the point estimate.
- **Probability Distribution of a Random Variable**. A description of the relative probabilities of the values/ranges on/in which a random variable might fall. Could be expressed as a mathematical formula, a numerical table, or a graph.
- **Quadratic Terms.** Independent variables in a non-linear function that are squared.
- **Qualitative**. A factor in a design of experiments that is not naturally a number, such as presence or not of a particular weapons package. Can be (artificially) made quantitative by assigning numbers to different categories (e.g., 1 if the training level of an individual is low, 2 if medium, or 3 if high).
- **Quantitative**. A factor in a design of experiments that naturally takes on a numerical value with some scale of measurement, either discrete or continuous.
- **Random Variable**. Mathematically, a numerical-valued function of points in the sample space. Practically speaking, a random variable is a number whose value depends on some random or uncertain activity, so we don't know its value before the activity; however, we might well know something about where it is likely and unlikely to be, as specified by its probability distribution.
- **Randomization**. Randomly assigning factor-level "treatments" to "experimental units," e.g., randomly assigning which experimental unit gets which kind of weapon. This is meant to break up any systematic bias that could result from systematic assignment (e.g., the weapons are numbered best to worst, as are the units based upon training level).

Generally not necessary in simulation experiments, provided that a good underlying random-number generator is used.

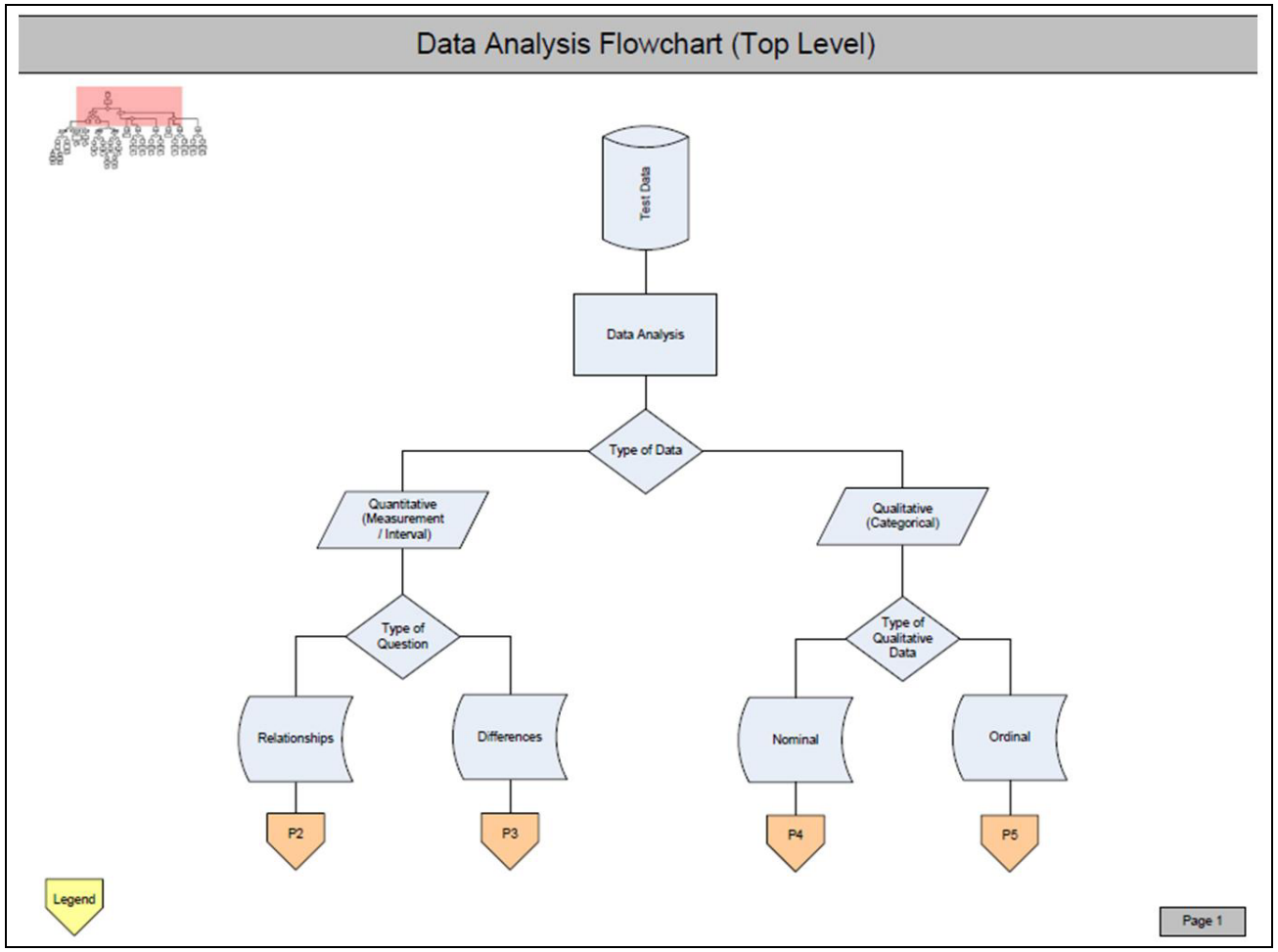
- **Range**. The possible values that a factor in a design of experiments can assume. A list of discrete values for a discrete factor, and a continuum of real numbers (possibly bounded on one or both ends) for a continuous factor.
- **Recursive Partitioning.** A statistical technique that "partitions data according to a relationship between the [independent variables and the dependent variables], creating a tree of partitions. [The technique] finds a set of cuts or grouping of [independent variables] that best predict [the independent variable]… [using] a tree of decision rules until the desired fit is reached." There are many variations of this technique. (SAS Institute, 2007).
- **Regression**. A statistical method in which a dependent variable (or response) is assumed to depend on the values of other independent variables (or regressors). Regression then estimates the nature of this dependence, as well as assesses whether individual independent variables are statistically significant predictors of the dependent variable. The form of the regression model (linear, interactions, quadratic, etc.) is often developed with the help of plots of the dependent variable vs. the various possible independent variables.
- **Regression Tree.** A statistical techniques that involves a recursive partitioning for continuous variables.
- **Regressor**. In regression, one of the independent variables on which the dependent variable might actually depend in some way.
- **Replication**. Repeating the experiment in a way that is identical to, but independent of, past experiments. In a design of experiments, replications could be made at individual design points, or of the entire design itself.
- **Response**. The output or dependent variable in a design of experiments, typically a measure of performance (MOP) or a measure of effectiveness (MOE). It is usually of interest to establish which factors have significant main effects on responses, and which factors might interact in their effect on responses. There could be multiple responses.
- **Sample Space**. The set of all possible values for the outcome of some random or uncertain activity. Subsets of the sample space are events, which have probabilities between 0 and 1 (inclusive on both ends).
- **Screening**. A type of design of experiments intended to identify which factors have a significant impact on a response and which do not. Those factors that appear not to matter can be dropped from further consideration, often greatly reducing further experimental effort if the original number of factors was high.
- **Simulation**. A numerical exercising of a mathematical model used when an exact mathematical analysis of the model is not available. Typically provides estimates/approximations rather than exact results, but does not require perhaps-unrealistic strong assumptions that would be necessary to arrive at a mathematically

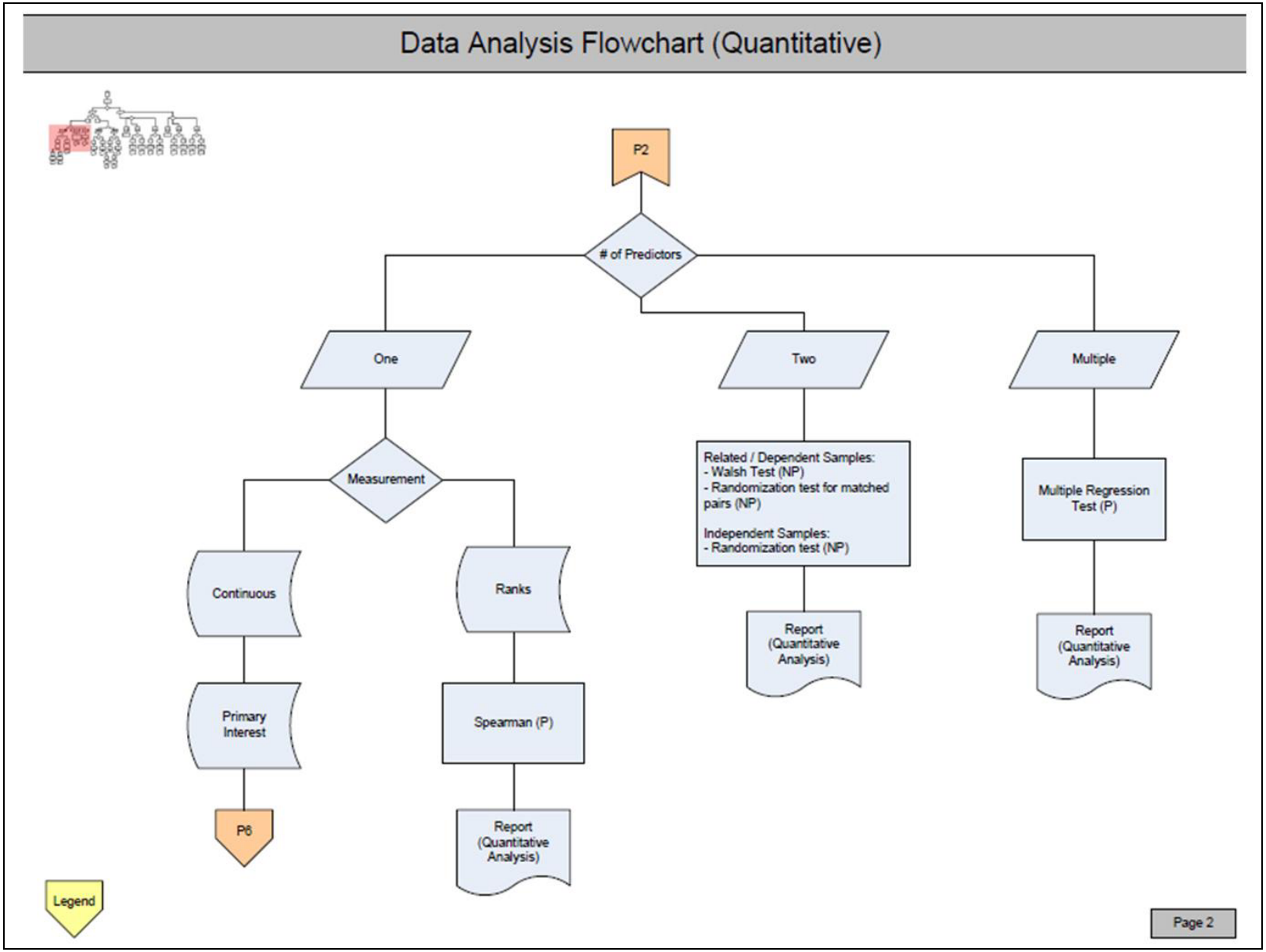
tractable model, so allows for more complex models that are typically more valid than over-simplified models created for mathematical tractability.

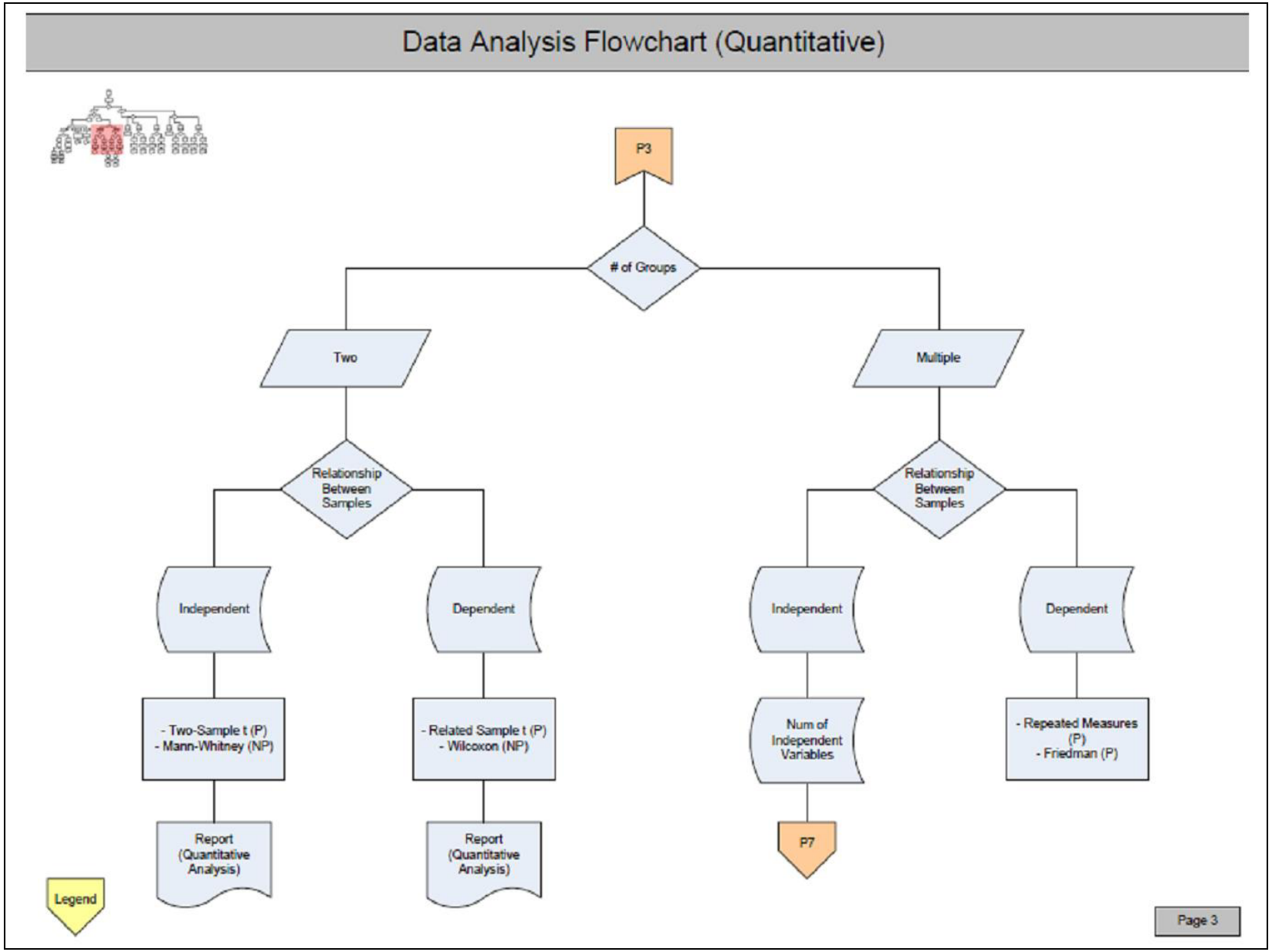
- **Standard Deviation**. The (positive) square root of the variance. Units of measurement are the same as those for the random variable (or data values).
- **Statistical Independence**. The property that two variables or events have nothing to do with each other. Equivalently, knowing the value of one of the variables (or the occurrence of one of the events) tells you nothing about the value of the other variable (or the occurrence of the other event). Independence greatly simplifies mathematical manipulations (since joint probabilities factor into the product of marginal probabilities), but assuming independence in situations where it is actually not physically true can lead to incorrect results (this is sometimes called the declaration of independence, since that is pretty much all it is).
- **Statistical Significance**. The property that an observation or value cannot reasonably be explained by mere chance, so that it is "real" beyond any natural variability that might be present. A statistically significant observation means that it is very unlikely that this could have happened by mere chance across different samples that might have been taken. In signal-processing terms, a statistically significant observation is one in which the signal component rises well above the noise component.
- **Stochastic**. Random, uncertain.
- **Uncertainty**. The quality that not everything is known exactly and with certainty.
- **Uncontrollable**. (1) A factor in a design of experiments that, in reality, cannot be controlled by a decision maker. For example, the degree of enemy jamming of communications, or weather conditions, would be uncontrollable factors. Though uncontrollable in practice, such factors are often made controllable some way in a simulation or live test. (2) A factor that cannot be controlled during experimentation (either live or simulation, or both). For example, the relative humidity during a live experiment.
- **Variability**. The property that a quantity, usually modeled as a random variable, does not always take on the same value, depending on chance circumstances.
- **Variable**. A quantity that can take on a variety of different values, possibly within some range.
- **Variance**. A measure of variability, defined as the expected value of the square of the difference between a random variable and its own expectation. The units of measurement of the variance are the square of the random variable's (or data values') unit of measurement, which may or may not make sense (meters to square meters, but dollars to square dollars?). Estimated from a sample by the usual sample-variance formula.

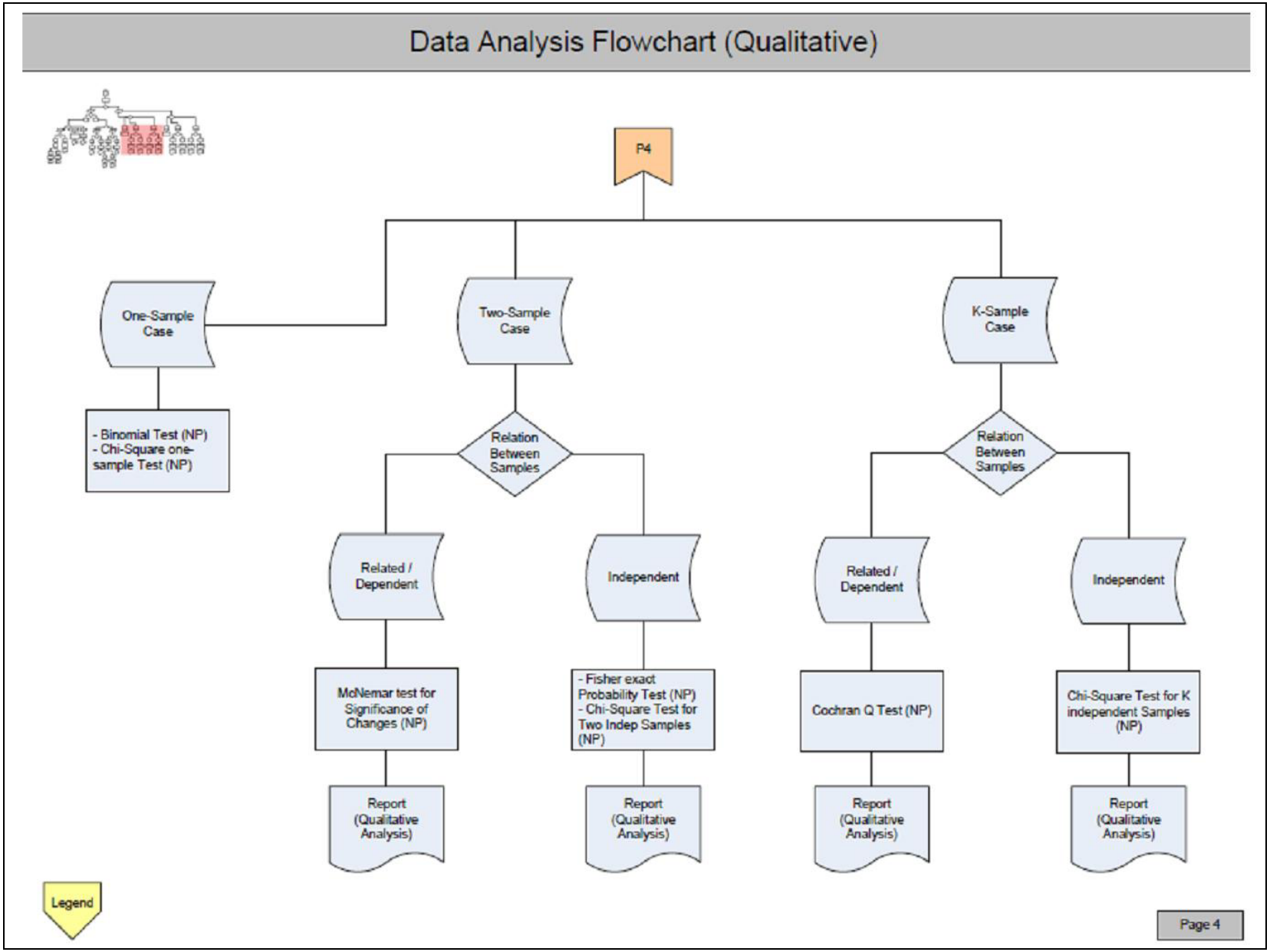
## ENCLOSURE 2 TO ANNEX G – ANALYTICAL TECHNIQUES DECISION FLOWCHART

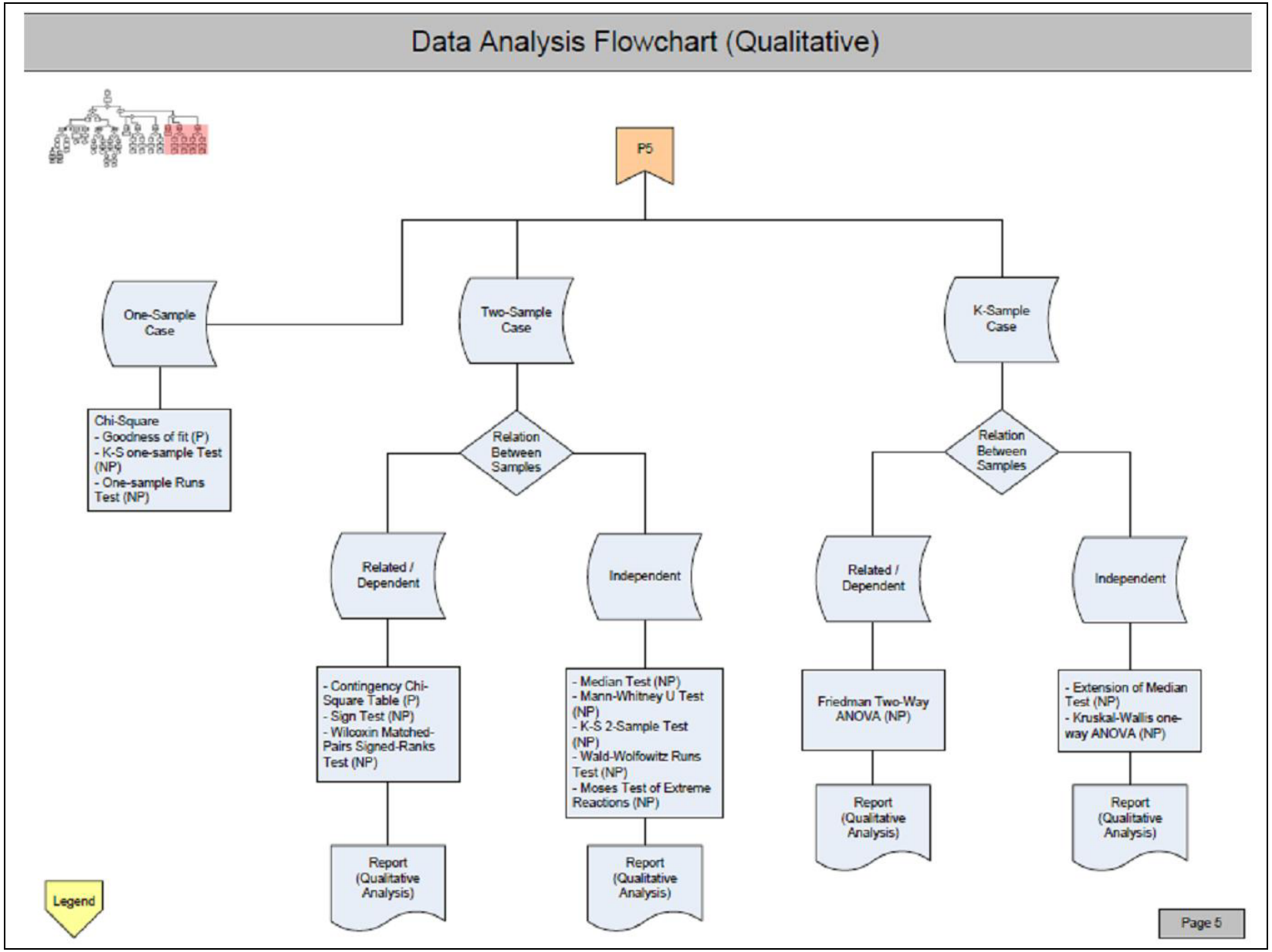
This enclosure provides a flowchart that can be used as a decision aid in identifying which analytical tool to use in evaluating data. The flowchart is broken down into eight separate slides as the decision tree entails numerous questions which must be answered. This flowchart is not intended to be an all inclusive chart but to provide the analyst with a logical process to decide an appropriate analytical technique.



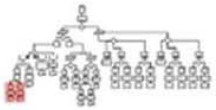

Data Analysis Flowchart (Quantitative)
P2
# of Predictors
One
Two
Multiple
Measurement
Related / Dependent Samples:
- Walsh Test (NP)
- Randomization test for matched pairs (NP)
Independent Samples:
- Randomization test (NP)
Multiple Regression Test (P)
Continuous
Ranks
Report (Quantitative Analysis)
Report (Quantitative Analysis)
Primary Interest
Spearman (P)
P6
Report (Quantitative Analysis)
Legend
Page 2

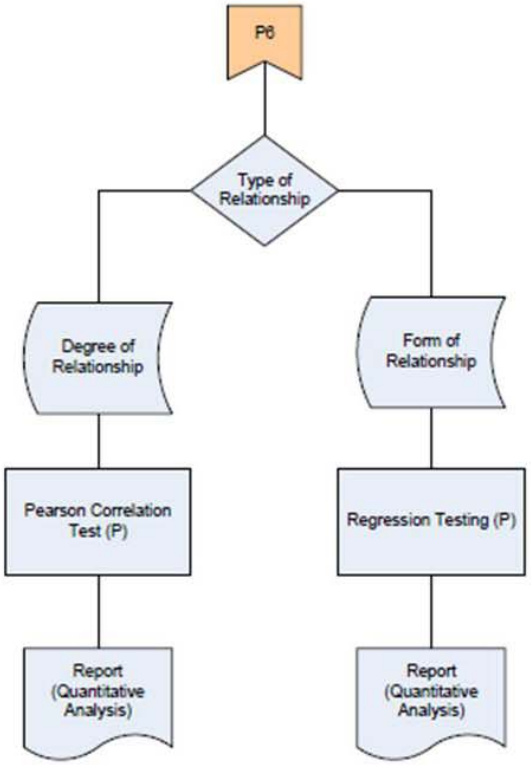

Data Analysis Flowchart (Quantitative)
P3
# of Groups
Two
Multiple
Relationship Between Samples
Relationship Between Samples
Independent
Dependent
Independent
Dependent
- Two-Sample t (P)
- Mann-Whitney (NP)
- Related Sample t (P)
- Wilcoxon (NP)
Num of Independent Variables
- Repeated Measures (P)
- Friedman (P)
Report (Quantitative Analysis)
Report (Quantitative Analysis)
P7
Legend
Page 3


Data Analysis Flowchart (Qualitative)
P4
One-Sample Case
Two-Sample Case
K-Sample Case
- Binomial Test (NP)
- Chi-Square one-sample Test (NP)
Relation Between Samples
Relation Between Samples
Related / Dependent
Independent
Related / Dependent
Independent
McNemar test for Significance of Changes (NP)
- Fisher exact Probability Test (NP)
- Chi-Square Test for Two Indep Samples (NP)
Cochran Q Test (NP)
Chi-Square Test for K independent Samples (NP)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Legend
Page 4


Data Analysis Flowchart (Qualitative)
P5
One-Sample Case
Two-Sample Case
K-Sample Case
Chi-Square
- Goodness of fit (P)
- K-S one-sample Test (NP)
- One-sample Runs Test (NP)
Relation Between Samples
Relation Between Samples
Related / Dependent
Independent
Related / Dependent
Independent
- Contingency Chi-Square Table (P)
- Sign Test (NP)
- Wilcoxin Matched-Pairs Signed-Ranks Test (NP)
- Median Test (NP)
- Mann-Whitney U Test (NP)
- K-S 2-Sample Test (NP)
- Wald-Wolfowitz Runs Test (NP)
- Moses Test of Extreme Reactions (NP)
Friedman Two-Way ANOVA (NP)
- Extension of Median Test (NP)
- Kruskal-Wallis one-way ANOVA (NP)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Report (Qualitative Analysis)
Legend
Page 5

# Data Analysis Flowchart (Quantitative)

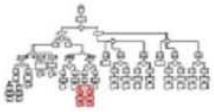


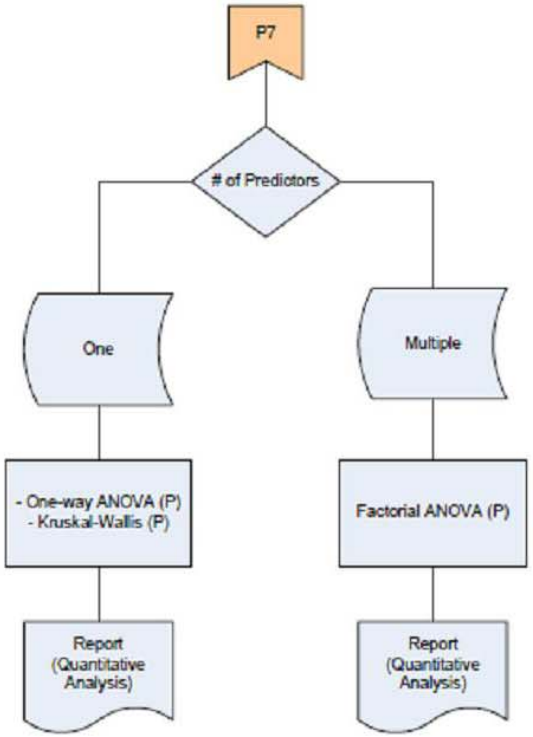



Legend

Page 6

# Data Analysis Flowchart (Quantitative)

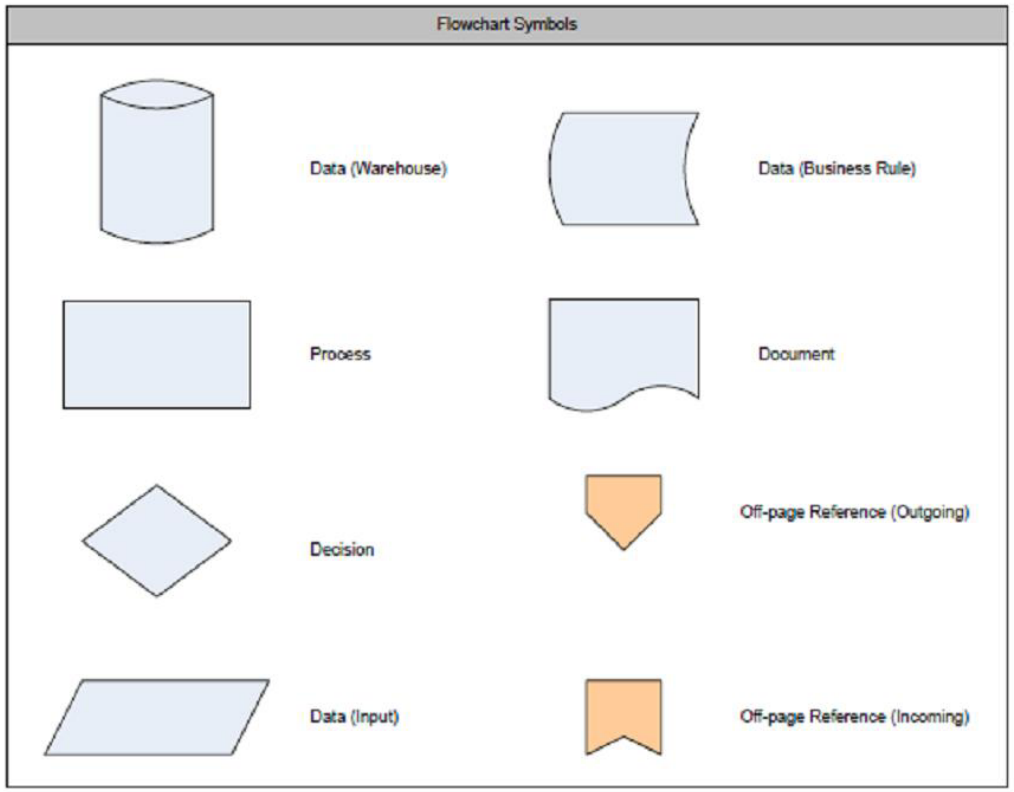





Legend

Page 7

# Legend




| Others | |
|---|---|
| (P) | Parametric |
| (NP) | Non-Parametric |

Page 8

## ANNEX H - DOD ARCHITECTURE FRAMEWORK (DoDAF)

### H.1 Architecture Views

The term *integrated architecture* refers to one in which architecture data elements are uniquely identified and consistently used across all products and views within the architecture. In most cases, an integrated architecture description has an OV, SV, TV, and an All View (AV) that are integrated with each other (i.e., there are common points of reference linking the OV and SV and also linking the SV and TV). The Operational Activity to Systems Functionality Traceability Matrix (SV-5), for example, relates operational activities from the Operational Activity Model (OV-5) to system functions from the Systems Functionality Description (SV-4); the SV-4 system functions are related to systems in the Systems Interface Description (SV-1), thus bridging the OV and SV. An architecture is defined as an *integrated architecture* when products and their constituent architecture data elements are developed, such that architecture data elements defined in one view are the same (i.e., same names, definitions, and values) as architecture data elements referenced in another view. Figure H-1 below, illustrates the major relationships.

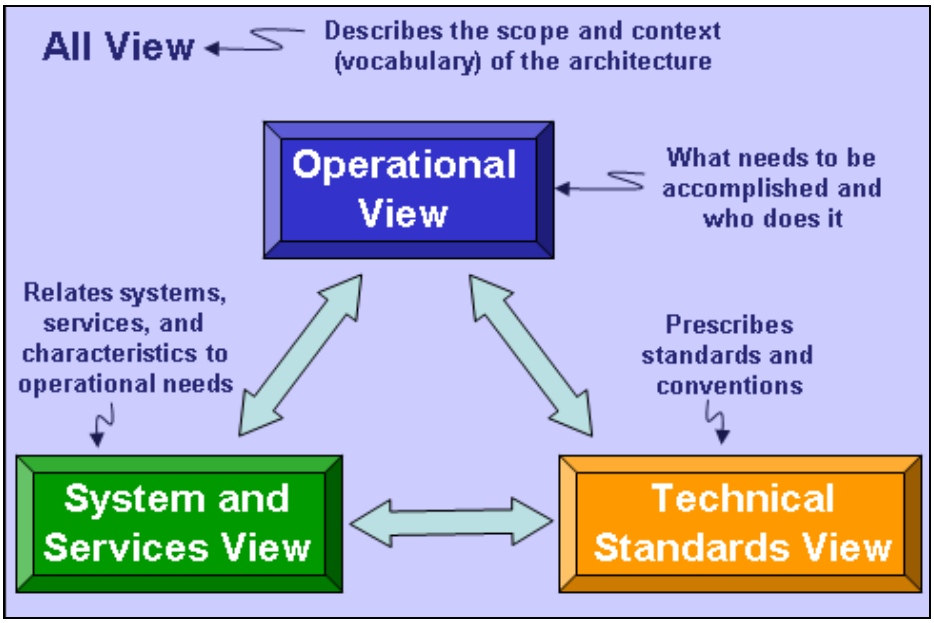



**Figure H-1. DoDAF All View**

#### H.1.1 Definition of the Operational View

The OV captures the operational nodes, the tasks or activities performed, and the information that must be exchanged to accomplish DoD missions. It conveys the types of information exchanged, the frequency of exchange, which tasks and activities are supported by the information exchanges, and the nature of information exchanges.

#### H.1.2 Definition of the Systems View

The SV captures system, service, and interconnection functionality providing for, or supporting, operational activities. DoD processes include warfighting, business, intelligence, and infrastructure functions. The SV system functions and services resources, and components may be linked to the architecture artifacts in the OV. These system functions

and service resources support the operational activities, and facilitate the exchange of information among operational nodes.

### H.1.3 Definition of the Technical Standards View

The TV is the minimal set of rules governing the arrangement, interaction, and interdependence of system parts or elements. Its purpose is to ensure that a system satisfies a specified set of operational requirements. The TV provides the technical systems implementation guidelines upon which engineering specifications are based, common building blocks are established, and product lines are developed. It includes a collection of the technical standards, implementation conventions, standards options, rules, and criteria that can be organized into profile(s) that govern systems and system or service elements for a given architecture.

### H.1.4 Definition of the All-Views

There are some overarching aspects of architecture that relate to all three views. These overarching aspects are captured in the AV products. The AV products provide information pertinent to the entire architecture but do not represent a distinct view of the architecture. AV products set the scope and context of the architecture. The scope includes the subject area and time frame for the architecture. The setting in which the architecture exists comprises the interrelated conditions that compose the context for the architecture. These conditions include doctrine; tactics, techniques, and procedures (TTP); relevant goals and vision statements; concepts of operations (CONOPS); scenarios; and environmental conditions.

Figures H-2 and H-3 show a table of all the DoDAF views.

## H.2 Using DoDAF Products

Test analysts can use DoD Architecture Framework products to support joint mission effectiveness (JMe) evaluation strategy development. The DoDAF products not only describe the relationship between system elements but also include a mapping of operational tasks to system functions. They reveal the allocation of system functions to hardware and software elements and how those functions support the tasks. That is, they can help address the CCI:

***Assess the ability to perform Task X by SoS Configuration Y under Condition Set A to achieve Mission Desired Effect Z.***

Architecture products are the natural artifacts of good systems engineering: translating a customer's needs into a set of technical requirements and then developing and fielding a system that meets the requirements and satisfies the needs [Miller]. Starting with that need means examining who the end user is (Operational Node), what he intends to accomplish in his domain (Operational Activities) and what information is exchanged between those activities. The operational activities and information exchanges are enabled by systems and their interfaces. Figure H-4 (inspired by Gunn, et al [2008]) is a simplified visualization that describes the underlying relationships.

| Applicable View | Framework Product | Framework Product Name | Net-Centric Extension | General Description |
|---|---|---|---|---|
| All View | AV-1 | Overview and Summary Information | ✓ | Scope, purpose, intended users, environment depicted, analytical findings |
| All View | AV-2 | Integrated Dictionary | ✓ | Architecture data repository with definitions of all terms used in all products |
| Operational | OV-1 | High-Level Operational Concept Graphic | ✓ | High-level graphical/textual description of operational concept |
| Operational | OV-2 | Operational Node Connectivity Description | ✓ | Operational nodes, connectivity, and information exchange need lines between nodes |
| Operational | OV-3 | Operational Information Exchange Matrix | ✓ | Information exchanged between nodes and the relevant attributes of that exchange |
| Operational | OV-4 | Organizational Relationships Chart | ✓ | Organizational, role, or other relationships among organizations |
| Operational | OV-5 | Operational Activity Model | ✓ | Capabilities, operational activities, relationships among activities, inputs, and outputs; overlays can show cost, performing nodes, or other pertinent information |
| Operational | OV-6a | Operational Rules Model | ✓ | One of three products used to describe operational activity—identifies business rules that constrain operation |
| Operational | OV-6b | Operational State Transition Description | ✓ | One of three products used to describe operational activity—identifies business process responses to events |
| Operational | OV-6c | Operational Event-Trace Description | ✓ | One of three products used to describe operational activity—traces actions in a scenario or sequence of events |
| Operational | OV-7 | Logical Data Model | ✓ | Documentation of the system data requirements and structural business process rules of the Operational View |
| Systems and Services | SV-1 | Systems Interface Description<br>Services Interface Description | ✓ | Identification of systems nodes, systems, system items, services, and service items and their interconnections, within and between nodes |
| Systems and Services | SV-2 | Systems Communications Description<br>Services Communications Description | ✓ | Systems nodes, systems, system items, services, and service items and their related communications lay-downs |
| Systems and Services | SV-3 | Systems-Systems Matrix<br>Services-Systems Matrix<br>Services-Services Matrix | ✓ | Relationships among systems and services in a given architecture; can be designed to show relationships of interest, e.g., system-type interfaces, planned vs. existing interfaces, etc. |
| Systems and Services | SV-4a | Systems Functionality Description | | Functions performed by systems and the system data flows among system functions |
| Systems and Services | SV-4b | Services Functionality Description | ✓ | Functions performed by services and the service data flow among service functions |
| Systems and Services | SV-5a | Operational Activity to Systems Function Traceability Matrix | | Mapping of system functions back to operational activities |
| Systems and Services | SV-5b | Operational Activity to Systems Traceability Matrix | | Mapping of systems back to capabilities or operational activities |
| Systems and Services | SV-5c | Operational Activity to Services Traceability Matrix | ✓ | Mapping of services back to operational activities |
| Systems and Services | SV-6 | Systems Data Exchange Matrix<br>Services Data Exchange Matrix | ✓ | Provides details of system or service data elements being exchanged between systems or services and the attributes of that exchange |

**Figure H-2. DoDAF View Listing**

| Applicable View | Framework Product | Framework Product Name | Net-Centric Extension | General Description |
|---|---|---|---|---|
| Systems and Services | SV-7 | Systems Performance Parameters Matrix<br>Services Performance Parameters Matrix | ✓ | Performance characteristics of Systems and Services View elements for the appropriate time frame(s) |
| Systems and Services | SV-8 | Systems Evolution Description<br>Services Evolution Description | ✓ | Planned incremental steps toward migrating a suite of systems or services to a more efficient suite, or toward evolving a current system to a future implementation |
| Systems and Services | SV-9 | Systems Technology Forecast<br>Services Technology Forecast | ✓ | Emerging technologies and software/hardware products that are expected to be available in a given set of time frames and that will affect future development of the architecture |
| Systems and Services | SV-10a | Systems Rules Model<br>Services Rules Model | ✓ | One of three products used to describe system and service functionality—identifies constraints that are imposed on systems/services functionality due to some aspect of systems design or implementation |
| Systems and Services | SV-10b | Systems State Transition Description<br>Services State Transition Description | ✓ | One of three products used to describe system and service functionality—identifies responses of a system/service to events |
| Systems and Services | SV-10c | Systems Event-Trace Description<br>Services Event-Trace Description | ✓ | One of three products used to describe system or service functionality—identifies system/service-specific refinements of critical sequences of events described in the Operational View |
| Systems and Services | SV-11 | Physical Schema | ✓ | Physical implementation of the Logical Data Model entities, e.g., message formats, file structures, physical schema |
| Technical Standards | TV-1 | Technical Standards Profile | ✓ | Listing of standards that apply to Systems and Services View elements in a given architecture |
| Technical Standards | TV-2 | Technical Standards Forecast | | Description of emerging standards and potential impact on current Systems and Services View elements, within a set of time frames |

**Figure H-3. DoDAF View Listing**

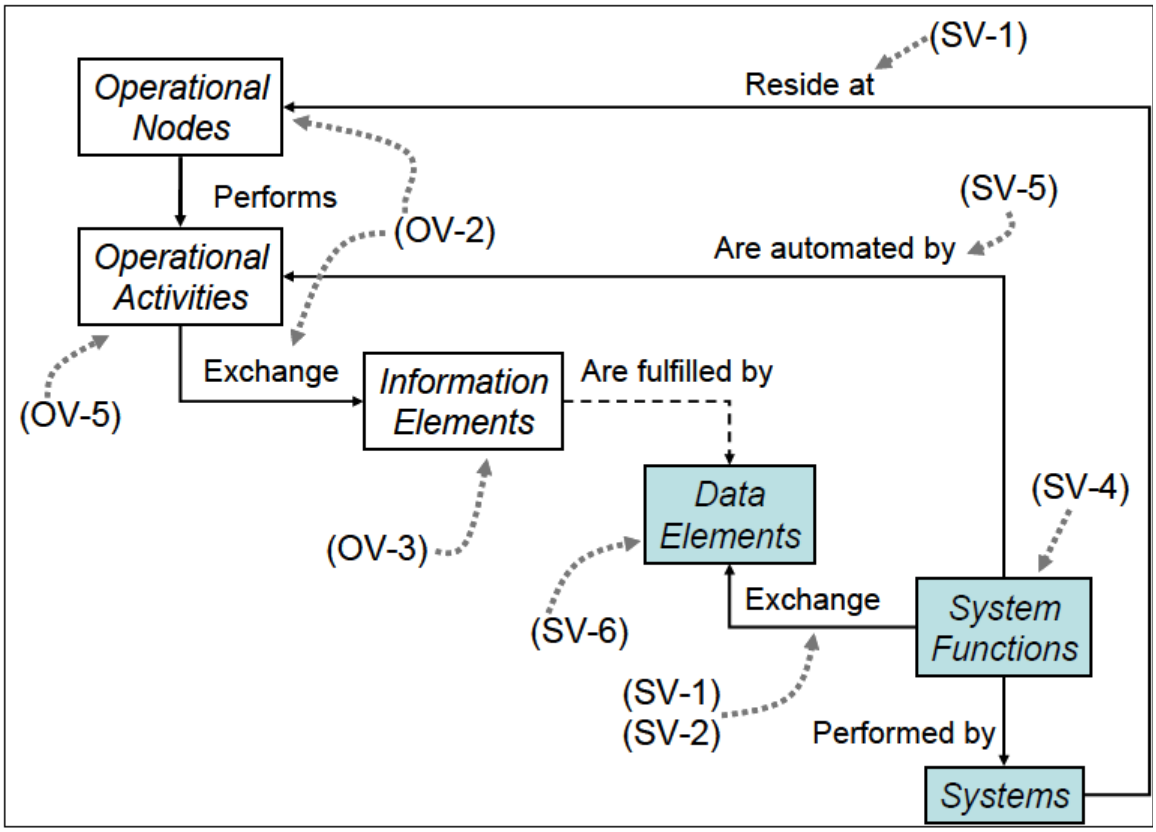



**Figure H-4. Simplified Representation of DoDAF Product Relationships**

Beginning with the doctrinally-supported (UJTL, Service-level Mission Essential Tasks, etc.) set of operational tasks captured in the OV-5 Operational Activity Model, one understands the mission area and joint capability to be supported by the SoS under test. Most activities require the exchange of information between operational nodes to support decision-making and proper execution, which is partially captured in the OV-2 Operational Node Connectivity Description and detailed in the OV-3 Operational Information Exchange Matrix. To this point, the architecture focuses on the operations domain. Objectives are achieved by the accomplishment of activities and tasks which work together because they produce and consume information [Lee, et al]. The exchange of information at the human level is supported, automated, and enabled by hardware and software systems. The functionality of the systems is described in the SV-4 Systems Functionality Description, which is mapped to the operational activities they support by the SV-5 Operational Activity to Systems Function Traceability Matrix. The proper functioning of those systems requires the exchange of data, which are captured by SV-1 Systems Interface Description and SV-2 Systems Communication Description, and further detailed in the SV-6 Systems Data Exchange Matrix. The SV products depict the tools utilized to facilitate the conduct of the operations. They capture the communications and hardware-centric requirements of the SoS in support of the mission [Lee, et al]. Further, assignment of systems (with their interface) to operational nodes (with their information need lines) is depicted in SV-1 Systems Interface Description. By identifying the facilities which host system elements, we see which operational organizations operate which system elements.

(Note: the above discussion is a simplification of much more complex concepts; it does not differentiate the network-centric extensions of DoDAF by calling out SV-4 and SV-4b or SV-5a, SV-5b and SV-5c separately.)

For products that are created within this integrated schema, the connections between system- and SoS-level measures (KPP, KSA, etc), Task MOPs and joint Mission MOEs are explicitly known. Further, the relationships between the activities and the relationships between the system nodes can help identify the measure relationships that exist between different hierarchical levels of MOEs and MOPs. They provide insight into SoS emergent behavior in which lateral influences dominate hierarchical relationships and causal interactions are not obvious or direct. Figure H-5 is a more comprehensive illustration of how one product integrates with others.

The integrated nature of a complete set of the products is apparent. No product stands alone. It is that very nature of interconnectedness that illuminates the SoS and operational relationships to support a solid composition of joint operational context. One possible scenario involves the application of artillery fires. Tasks like Conduct Fire Support Coordination, Select Target to Attack and Request Attack as decompositions of the higher level task Conduct Joint Fires are captured in the OV-5 along with related tasks like Collect Target Information and Provide Battle Damage Assessment. Measures for these tasks are available in the UJTL: a measure at the tactical level for Conduct Joint Fires is the percentage of targets attacked in accordance with requests for fires. Other measures are available in Service task lists. The operational nodes involved may include an infantry battalion fire support coordination center (FSCC), an artillery battalion fire direction center (FDC), an artillery battery FDC and a forward observer. The information they exchange includes Fire Request, Fire Order, Firing Data, Adjust Fires, and Battle Damage Assessment. Those nodes and the information exchanges are displayed

graphically in the OV-2. The details of each information exchange (content, information producer and consumer, periodicity, timeliness, security, etc) are in the OV-3.

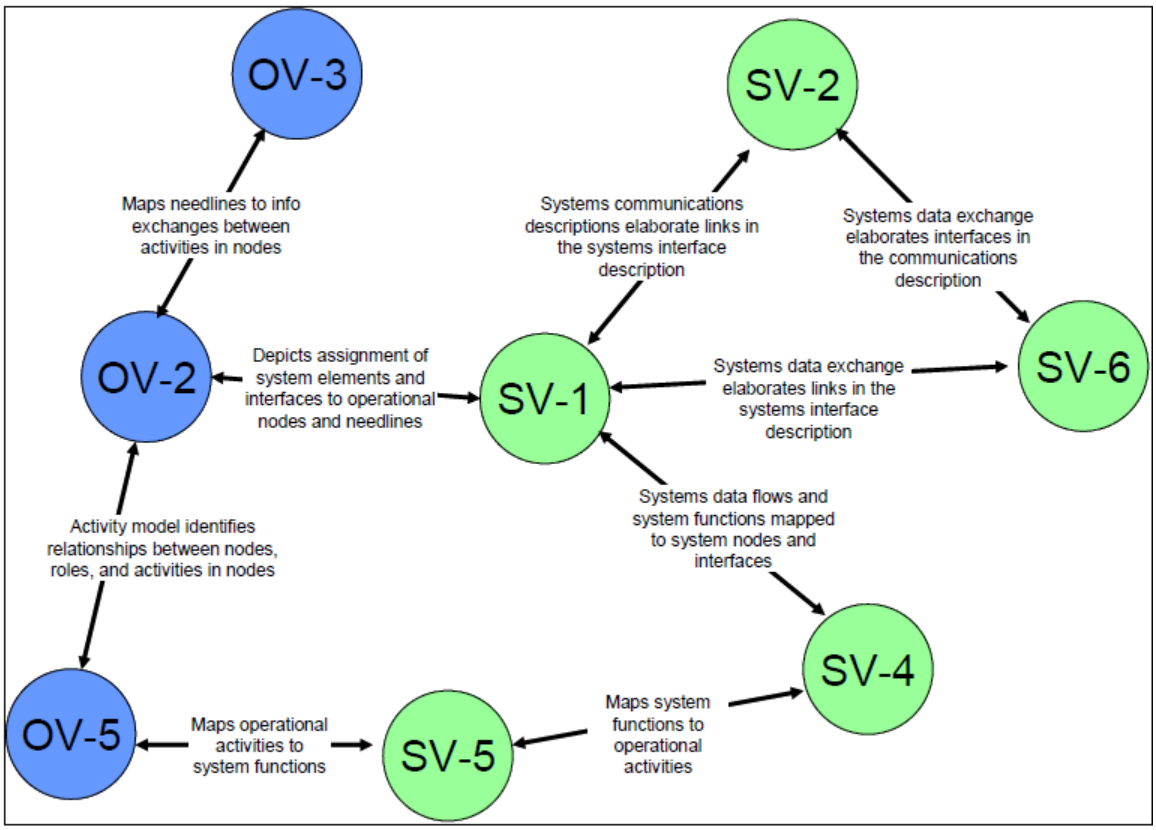



**Figure H-5. Selected DoDAF Product Relationships**

Automated systems supporting the tasks include AFATDS, EPLRS and SINCGARS and are shown on the SV-1 with their communication connections on the SV-2. The details of each data exchange element (format, media, data standard, etc) are in the SV-6. The functions those systems perform include receive fire request, identify ordnance, de-conflict airspace, verify trajectories and generate fire order and are described in the SV-4. A direct mapping of the system functions to the operational tasks they support is provided in the SV-5 matrix.

Ultimately, we should also consider where the DoDAF products fit in the larger scheme of CEM. Figure H-6 is a mapping of how DoDAF products can serve as a source of information in JOC-T. Some caution should be exercised, because not all the information required in JOC-T is captured in the DoDAF products and not all the information in the products is applicable in JOC-T.

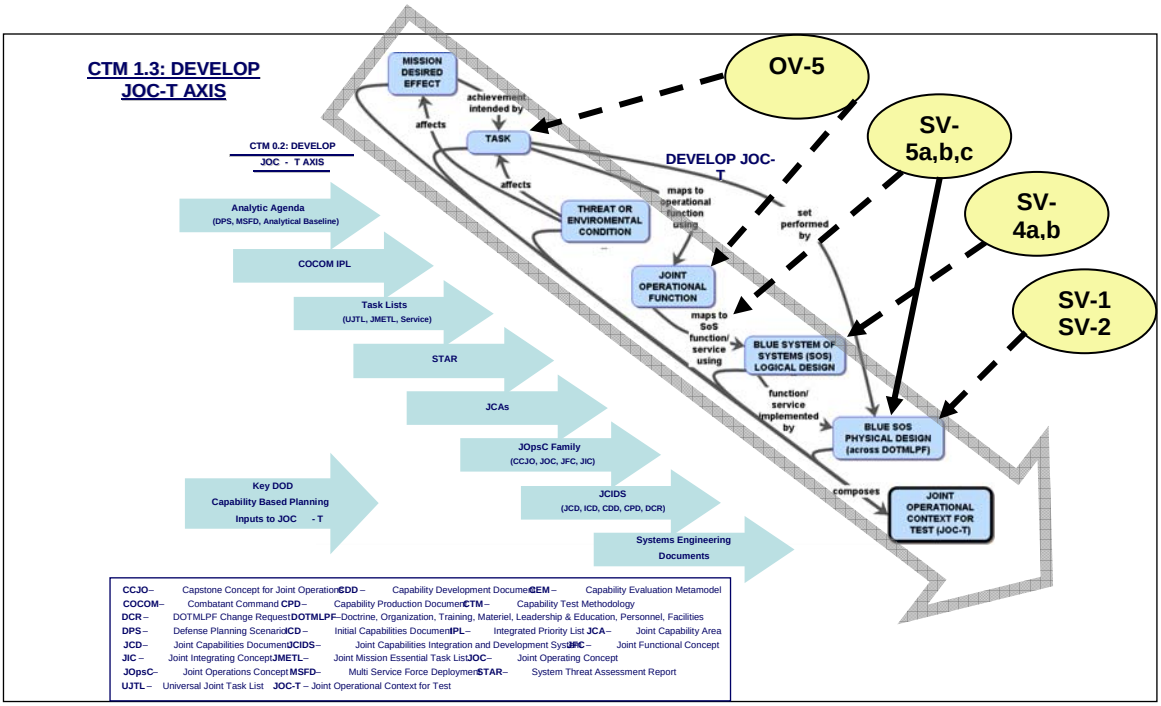



**Figure H-6. The JOC-T CEM and DoDAF Products**

When evaluating a SoS composed of elements from different acquisition program offices, each system may have its own set of DoDAF products. While they will be integrated within a set, they may not be completely consistent with other programs' products.

The SoS under test may include System X fielded by PEO ABC and System Y fielded by PEO DEF, and both systems are required to achieve the task or desired mission effect. However, because the DoDAF products for those systems were developed, vetted, and approved by different organizations, they may exhibit some differences. The opportunity of a critical examination of the products allows all stakeholders to identify such issues early in the process of developing an evaluation strategy and to take appropriate action.

## H.3 DoDAF Product Enhancements

During the development of the CTM, some enhancements to DoDAF products were suggested to the DoDAF working group for consideration into future DoDAF revisions.

### H.3.1. Operational View 1 (OV-1)

The OV-1 provides an operational view of the overall joint operating area. It was recommended to include mission and task information to enhance the view. Figure H-7 is an example of an enhanced OV-1.

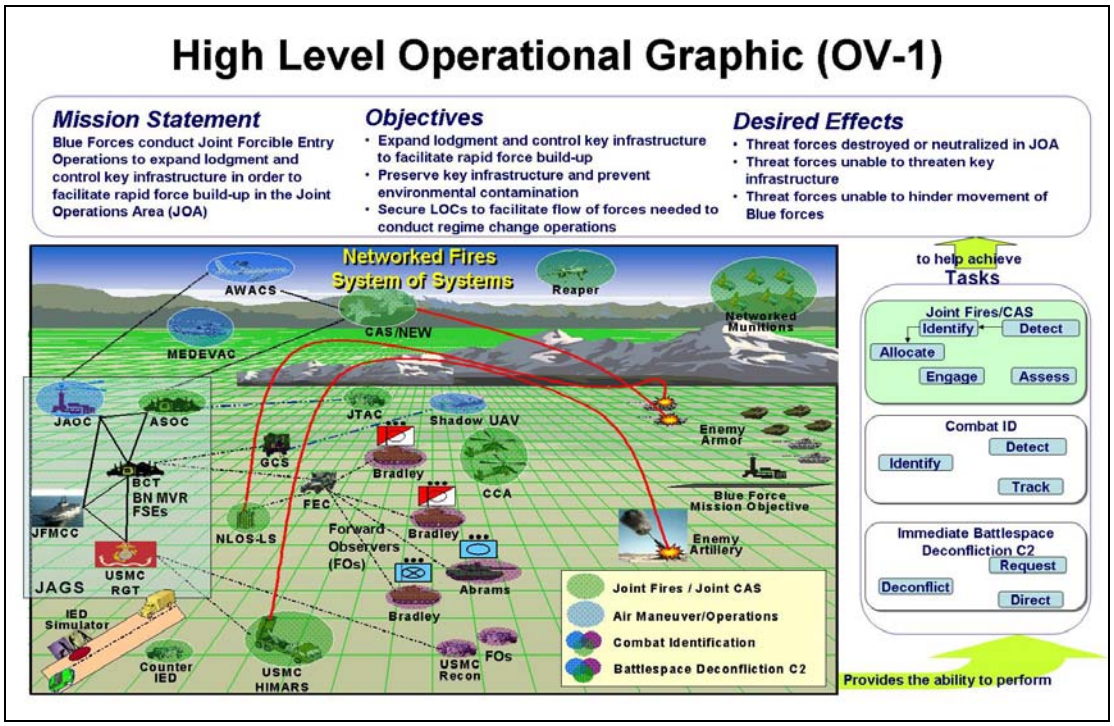



**Figure H-7. Enhanced OV-1**

### H.3.2. Evaluation View (EV)

The addition of Evaluation Views (EV) provides detailed information to the analyst and test team on measures, data elements, and analysis requirements. EVs were first in use by the Army's Cross Command Collaboration Effort (3CE). The JTEM team built on the idea to develop evaluation views for each measure that would support the IDRL development and populate the Data Analysis Plan (DAP). EVs would support the measures framework to evaluate joint mission effectiveness (JMe) in terms of the following three evaluations: (1) SOS performance, (2) joint task performance, and (3) accomplishment of mission desired effects. EVs consist of a front page EV-1 and back page EV-2 for each measure.

Figure H-8 depicts an example EV-1 for a Mission MOE. As seen in the figure, the EV-1 includes an overall cartoon of the SoS and the JME. Mission information (mission statement, end state, and desired effect) are also provided to add traceability of the Mission MOE to joint mission effectiveness. A Task MOP EV-1 may alternatively provide task description, standards, etc. A measure of SoS Attributes (MOSA) may have an attribute description and relevant key characteristic information. Additionally, the EV-1 includes a time horizon and list of key events that are critical to the evaluation of the measure and the collection of relevant data elements.

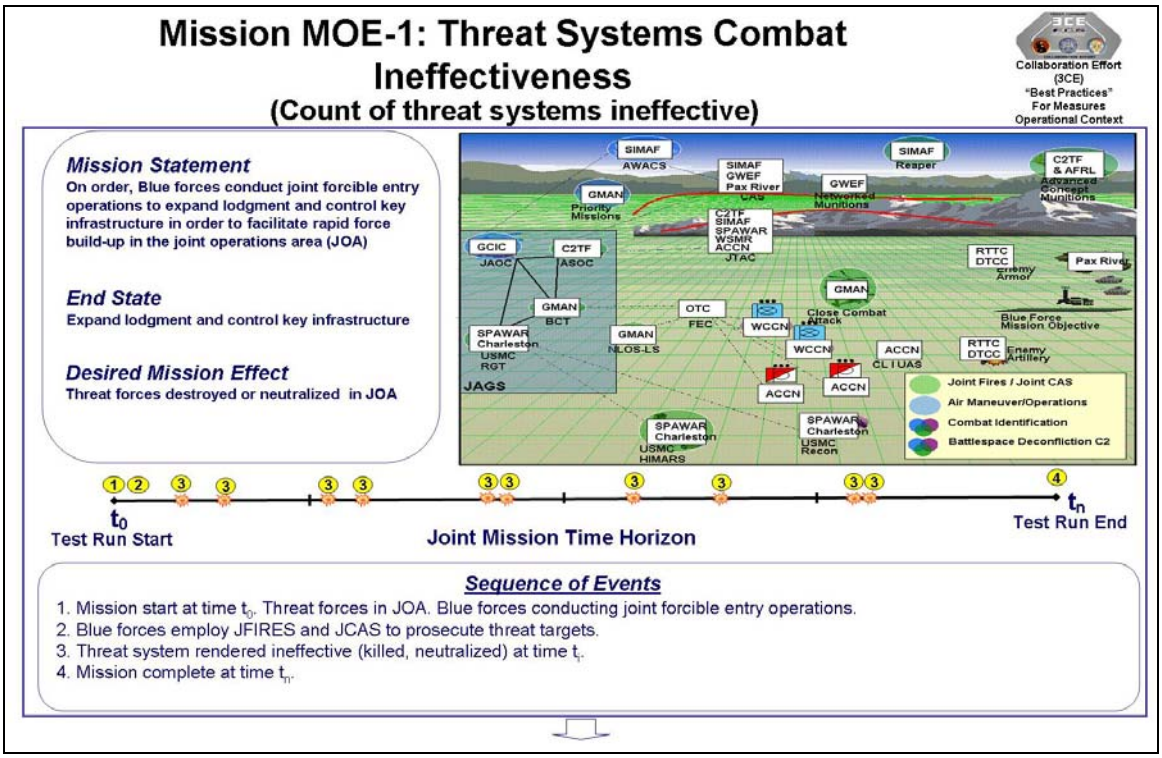



**Figure H-8. Example Mission MOE EV-1**

Figure H-9 depicts an example Mission MOE EV-2. This view provides the detailed information required to ensure complete and accurate data is collected as required by the measurement evaluation. The first section provides a list of required data elements along with a short description, data element notation, and units of measure. The second section provides detailed description of key terms. The descriptions must be complete enough to eliminate any misunderstandings or vagueness with the terms. The last section provides a calculation and success criteria so as to inform test team members on the intended use of the data elements and evaluation of the measure.

The primary value to the development of EVs is in providing sufficient detailed information for the analyst, data collector, and test team to ensure complete, consistent, and relevant data that will support the evaluation of the measure. A secondary value to EVs is in its reuse. Once an EV is created, it may be reused in future test events for the capability or for supporting capabilities. It is suggested that the analyst always archive EVs for possible future use.

# Mission MOE-1: Threat Systems Combat Ineffectiveness
## (Count of threat systems ineffective)

Collaboration Effort (3CE) "Best Practices" For Measures Operational Context

### Data Elements

1. Blue Desired Threshold Value ($P_D$): Percentage Threat Systems combat ineffectiveness
2. Total number Threat Systems in JOA at start time ($X_d$)
3. Count of Threat System ineffective (Killed, Neutralized) at test run end ($X_i$)

### Key Terms

1. **Threat Systems**: Those threat systems that are in the JOA at test run start time. It is assumed that the number of threat systems in the JOA remain constant over the Test Run period of time; that no threat systems enter or leave the JOA. (Unit **ID**: identifying ID number; **Effectiveness**: E – Effective, I – Ineffective).
2. **Threat System kill**: Those threat systems that can no longer move, shoot, and sense. System no longer possesses a capability. Assumes cannot be fixed and remains a kill until Test Run end. (Unit **ID**: identifying ID number; **Effectiveness**: E – Effective, I – Ineffective; **Status**: A – Active, K – Killed, N – Neutralized; **Ineffective Time**: Time which entered Ineffective state).
3. **Threat System Neutralized**: Those threat systems that have been attacked and can no longer shoot but are not killed. System is combat ineffective but still retains some capability to move or sense. (Unit **ID**: identifying ID number; **Effectiveness**: E – Effective, I – Ineffective; **Status**: A – Active, K – Killed, N – Neutralized; **Ineffective Time**: Time which entered Ineffective state).

### Calculation

$$\frac{\frac{X_i}{X_d}}{P_D}$$

where:
$X_i$ = Count of Threat systems ineffective at test run end time
$X_d$ = Total number Threat systems at test run start time
$P_D$ = Desired percent of Threat systems ineffectiveness

### Success Criteria

MMOE ≥ 1.0

**Figure H-9. Example Mission MOE EV-2**

# SAMPLE
# DATA MANAGEMENT PLAN
# FOR
# JOINT TEST AND EVALUATION EVENTS

## MARCH 2009

DISTRIBUTION STATEMENT:
Distribution limited to US Government agencies and …….. Other requests for this document should be referred to Organizations Name, Office Symbol, ATTN: Department, Address, Zip Code

*This page intentionally left blank.*

## TABLE OF CONTENTS

**1. Introduction ..... 1**

1.1. Overview ..... 1
1.2. Purpose of the Data Management Plan ..... 1
1.3. Data Collection System Description ..... 1

**2. Data Collection ..... 2**

2.1. Data Requirements ..... 2
2.2. Types of Data Collection ..... 2
2.2.1. Manual Data Collection ..... 2
2.2.2. Automated Data Collection ..... 2
2.3. Procedures ..... 3
2.3.1. Tactical Message Collection ..... 3
2.3.2. Simulation Message Collection ..... 3
2.4. Tools ..... 3
2.4.1. Remote Reconfigurable Intelligent Instrumentation to Control, Collect, Simulate and Stimulate. ..... 3
2.4.2. DPU Parser ..... 4
2.4.3. Test Talk ..... 4
2.4.4. Distributed Capabilities Integration Toolbox (DCIT) ..... 4

**3. Data Storage ..... 5**

3.1. Overview ..... 5
3.2. Procedures ..... 5
3.2.1. Data Harvesting ..... 5
3.3. Tools ..... 6
3.3.1. Data Warehouse System ..... 6
3.3.2. Simulation Collection System ..... 6

**4. Data Reduction ..... 7**

4.1. Overview ..... 7
4.2. Procedures ..... 7
4.2.1. Using the DCARS client (example follows) ..... 7
4.3. Tools ..... 7
4.3.1. DCARS Client (example follows) ..... 7

**5. Data Reporting ..... 9**

5.1. Overview ..... 9
5.2. Procedures ..... 9
5.3. Tools ..... 9
5.4. Data querying ..... 9

5.5. Overview ..... 9

**6. Data Archival ..... 10**

6.1. Overview ..... 10
6.2. Procedures ..... 10
6.3. Test Metadata ..... 10
6.4. Data Analysis and Presentation ..... 10

**Annex A. Acronyms and Abbreviations ..... A-1**

**Annex B. DCARS Software Component Functions ..... B-1**

**Annex C. Sample Data Collection Network ..... C-1**

**Annex D. Data Levels ..... D-1**

# 1. Introduction

## 1.1. Overview

Department of Defense (DoD) Strategic Planning Guidance demanded creation of joint environment testing capability. This demand led to creation of the Capability Test Methodology (CTM) and the Joint Mission Environment Test Capability (JMETC) as part of a testing in a joint environment. The Test Event will provide a high fidelity, real-time, rapidly configurable, distributed network including virtual and constructive models linked with live systems, to evaluate a joint capability as well as support the development and testing initiatives for the partnered organizations. The event design is based on a Joint Operational Context for Test (JOC-T) drawing from authoritative planning sources. The test will execute from TBD through TBD.

## 1.2. Purpose of the Data Management Plan

The purpose of the Data Management Plan (DMP) is to provide procedures for the collection, reduction, quality assurance, collation, storage, and archival of data gathered to support determination of the operational effectiveness and suitability for the system under test. This document aligns with the test plan and the analysis in terms of contribution to a successful test. The DMP is both a planning tool to ensure procedures are in place for data collection and a data management tool for tracking and assessing data collection during test execution. The purpose of this plan is to formally document the intended course of action for collection, reducing, reporting, and archiving of data products relating to the joint test event.

## 1.3. Data Collection System Description

Provide a description of the system or systems used for data collection. List and define the risks in the data requirements, collection of data, and analysis. Discuss possible mitigation processes to minimize the impacts. Discuss the tools or process for the data capture, the flow of the data and the process the data goes through in preparation for analysis. If the system is automated, list the strengths and weaknesses and the risk reduction steps to mitigate any deficiencies for the automated process, such as redundant observations or paper based surveys, etc. List any requirements for logistical or information technology support for the systems. Provide a graphical depiction of the data flow and the nodes in the data collection system. See Annex C for a sample data collection system diagram.

# 2. Data Collection

## 2.1. Data Requirements

Describe the source of all data elements necessary to resolve the measure. Cite required data parameters (data fields) and their characteristics to include supporting data, measurement units, measurement frequency, etc. Include logs, evaluator comments, questionnaires, operator or maintenance comments, multimedia documentation, and any other input items necessary for the evaluation. Provide the sampling plan, including the measurement frequency, sample size per measure, and the total sample size.

## 2.2. Types of Data Collection

Describe, in detail, the methods and instrumentation used to collect the data, and any manual or automated processing to be used to transform the data into a form for analysis. Include a description of any agreements with other organizations (either in place or required) to obtain the required data. Discuss range infrastructure adequacy, if necessary, and highlight those data elements whose source must yet be range certified. Discuss the data configuration management process (i.e., procedures for (1) marking/tagging data sources especially hard copy questionnaires, maps, or video documentation; (2) verifying the data are complete and accurate; and (3) ensuring the integrity of processed data. Include a summary of data collection dry-run procedures for all steps/areas.

### 2.2.1. Manual Data Collection

- Time Ordered Event List (TOEL) data
- Test Incident Reports
- Test Logs (recorded using Test Talk)
- System reboot logs
- Subject Matter Expert (SME) logs

### 2.2.2. Automated Data Collection

- Messages and HLA data collected on all LANs using RICS2 and DCARS
- Ping logs recorded by STARSHIP
- Message logs from the systems under test
- Simulation/Stimulation logs
- Data elements from live platforms, both stationary and moving, including platform or system identification and TSPI
- Camera and sensor video

## 2.3. Procedures

List the supporting information contained in the sub-bullets below for each type of data collection system such as tactical messages or simulation messages or modeling output.

### 2.3.1. Tactical Message Collection

- List data collection location and the organization responsible for ensuring the collection occurs properly.
- Describe each type of data element's flow into the data collection system.
- Describe any compiling of data within the data collection system
- List the location or locations where the compiled data is accessible.
- List any storage or archival systems or locations, and how to access data from the owning organizations.

### 2.3.2. Simulation Message Collection

- List data collection location and the organization responsible for ensuring the collection occurs properly.
- Describe each type of data element's flow into the data collection system.
- Describe any compiling of data within the data collection system
- List the location or locations where the compiled data is accessible.
- List any storage or archival systems or locations, and how to access data from the owning organizations.

## 2.4. Tools

Provide descriptions for all of the tools used in the data collection system.

### 2.4.1. Remote Reconfigurable Intelligent Instrumentation to Control, Collect, Simulate and Stimulate.

(RICS)2 is a controller, collector, simulator, stimulator executing on a Windows NT platform capable of running in diverse environments. The software and hardware architecture suite maximizes Commercial-off-the-Shelf software and hardware (COTS) with internal GPS and data radios to provide time tagging and a remote instrumentation network. Primary uses of the RICS2 are to send and receive ground truth data for training, and performance tests of Tactical C4I systems and networks. Functions of the RICS2 include: the physical interface to tactical radios and ABCS for STORM and DCARS, Traffic Generator unformatted free text stimulation, DAUVS (Digital Army USMTF (US Message Text Format) Variable Message Format (VMF) Stimulator), Joint Variable Message Format (JVMF) stimulation, Interface Network Controller (INC) Sniffer for INC ground truth, Standard Query Language (SQL) for Army Tactical Command & Control System (ATCCS) platform ground truth, HTTP ATCCS ground truth, High Speed LAN TAP (HSLT) LAN data collection, Wide Area Network (WAN) FBCB2 (TI) data collection, Pinger and Bandwidth Monitor for network monitoring.

### 2.4.2. DPU Parser

Data Processing Unit is a Raw Data Parser for (RICS)2 Data Sniffer. Takes binary raw files and converts them into Access Database or Text file. Can convert HSLT LAN Files, IFDC Tapes, IFDC Raw Files, Real Time Socket, WANTAP Files and MCREWS Log Files.

### 2.4.3. Test Talk

Test Talk is a stand-alone, requirements-driven tool that displays a common test time and event schedule, called a TOEL, and a user message board.

### 2.4.4. Distributed Capabilities Integration Toolbox (DCIT)

Connectivity Matrix Utility (CMU) The CMU facilitates the connectivity check procedure for a test network connecting multiple sites with multiple nodes with multiple simulation components. The application tabulates in matrix form the connectivity verification input from operators at each remote site. The CMU can be used for network, simulation and live vehicle connectivity checks. Please note: DCIT will be used to document all test results immediately after the completion of each test.

# 3. Data Storage

## 3.1. Overview

Develop procedures and tools that filter all collected data to ensure data production in accordance with the data element requirements specified in the Data Analysis Plan (DAP), or the Data Collection Matrix (DCM).

## 3.2. Procedures

### 3.2.1. Data Harvesting

To upload LDC log files to DCARS Website, rename your log file to your LDC znumber_date which is provided in Table 1.

**Table 1. LDC Numbers by Site**

| Site | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| LDC # | XXXX | XXXX | XXXX | XXXX | XXXX | XXXX | XXXX |

- Open Internet Explorer
- Type IP for the DCARS web server in the address bar. IP Address is available on the classified wiki.
- Create a login if you have not logged into this particular DCARS web server.
- Click on the link on the left side of the web page that says “User File Uploads.”
- Radio button will default to “Public.”
- Click on the link that says “Open File Upload Page.”
- Once the Java Window opens, click on “Add” button.
- Another Java window will open which will allow you to browse to the file that you want to upload.
- Select (highlight) the file to be uploaded, click the “OK” button.
- Click on the “upload Selected Files” button.

NOTE: Do NOT check the zip files check box as this will rename your zip file to “upload.zip”.

To verify your upload, follow these instructions:

- On the DCARS web site, click on the link on the left side of the web page that says "User File Downloads."
- Ensure the radio button is on "Public."
- Click on the link that says "Show files in Selected Folder."

NOTE: Do not forget to logout of the website before shutting down your web browser Any questions should be directed to system point of contact.

## 3.3. Tools

### 3.3.1. Data Warehouse System

The Data Warehouse System (DWS) links directly to the data collection instrumentation and other DCARS components to gather, consolidate, and archive the collected data. The DWS performs the major portion of the data processing. It also archives the collected information in the message database and the fusion database.

The DWS receives and stores the data collected from the C4I system workstations and tactical networks via the data collection instrumentation and from simulation and range instrumentation networks via the Simulation Collection System (SCS). It streams location data to the Geographic Information System (GIS) for live plotting and stores all the data in its own database repository for near real-time retrieval and fusion and to prepare customer products. This data includes location information for units and supply activities, overlays and military graphics, fire mission and target information, personnel and equipment resource status, and attack criteria, among much other information.

### 3.3.2. Simulation Collection System

The SCS links directly to the simulation network to gather the entity data and interactions from the simulation sources and the live telemetry from range instrumentation systems. It collects information and data from the simulations used to control or drive an event and to stimulate responses in various C4I systems or command and staff elements. These simulations include the High Level Architecture (HLA) and Distributed Interactive Simulation (DIS) networks. In addition, it collects data and information from the Test and Training Enabling Architecture (TENA) and from range instrumentation used to track or control live entities.

# 4. Data Reduction

## 4.1. Overview

Develop procedures and tools that filter all collected data to ensure data production in accordance with the data element requirements specified in the DAP, or DCM.

## 4.2. Procedures

### 4.2.1. Using the DCARS client (example follows)

- Analyst will utilize the DCARS client to view both the tactical and simulation message database through designated workstations in the JATTL.
- Analyst logs onto the JATTL workstation.
- Access the ABDC document on the desktop for JDAS/DCARS to get DCARS client log-on and server connection information.
- Click the DCARS client icon on the workstation desktop.
- Enter information from ABDC document into the DCARS client window.
- DCARS client will initialize with a connection to the server which was selected at start-up.
- In the DCARS client, use the menu on the left side of the client to select either the tactical or simulation message database related to the information.
- More details for using the DCARS client can be found in the DCARS user's manual. A copy is available at the workstations in the JATTL.

## 4.3. Tools

### 4.3.1. DCARS Client (example follows)

The DCARS Client software resides on Analyst Workstations and performs three primary functions:

a. Importing processed data files into directories or commercial off-the-shelf (COTS) databases
b. Reporting the processed data in product format
c. Plotting the processed data.

The DCARS Client is the primary analyst interface to the DCARS-II system. It provides mapping displays, data and message query and analysis, and live-plotting functionality. The analyst uses the Client to query the processed data and messages based on a variety of selection criteria including data source, time frame, element type, echelon, category, location, and identification, among others. The Client can display message contents and data in textual format. It can also display data in graphical format on background maps. The maps cover a multiple theaters with a variety of selectable layers. The Client can also

plot live streaming data simultaneously with existing maps and historical data queries. The DCARS Client can perform some of its functions while the other DCARS-II modules are inactive.

The DCARS Client provides rapid report response and the capability to search multiple databases. This enables analysts to compare data among a variety of systems and sources. This tool can produce several types of reports depending on the selected product, including: a summary view on the client, a formal report pre-formatted in accordance with the specifications of the analysis team, an unformatted delimited text report suitable for export to other applications, and a plot file for the GIS tool.

- **Military Message Traffic**

The DCARS-II instrumentation can collect all electronic message packets traveling over the military C4I networks. The common message types exchanged among the C4I system workstations and their supporting systems include: United States Message Text Format (USMTF), Joint Variable Message Format (JVMF), legacy C4I system, Over-The-Horizon Gold (OTG), Tactical Digital Information Link (TADIL), and other miscellaneous messages. The event analysts access the collected messages using the Message Handling System (MHS) on the DCARS-II Client.

The MHS provides a tool for the analyst-user to search for and view captured electronic messages. In addition, the MHS provides the capability for each analyst to set up profiles to automatically select messages for viewing based on their type, content, or other features. In addition to providing answers to specific inquiries, intelligent selection and viewing of these messages often can give the analyst an insight into the condition of the unit, esprit of its personnel, and its operational situation. There is no practical limit to the number of different profiles and searches a user can establish. Each profile or search is unique to the user who establishes it.

- **Screen Capture**

The DCARS-II provides a remote Screen Capture capability that enables the analyst-user to capture and view the display screens from any selected ABCS workstation accessible on the military C4I network. These products are useful for comparing the information that was available to the information that the commander was actually using. Analysts should coordinate screen captures with on-site observers. The captures do not significantly interfere with system operation, but done blindly, usually result in useless images that do not show the desired information.

# 5. Data Reporting

## 5.1. Overview

Develop procedures and tools that organize and align the data according to analyst and evaluator requirements in the DCMP. This data should align to technical, operational, or analytical categories specified by the lead analyst/evaluator.

## 5.2. Procedures

## 5.3. Tools

## 5.4. Data querying

Solicit the available data to make the data available in analyst format requirements in accordance with queries identified to categorize raw data in the form of measurement space requirements.

## 5.5. Overview

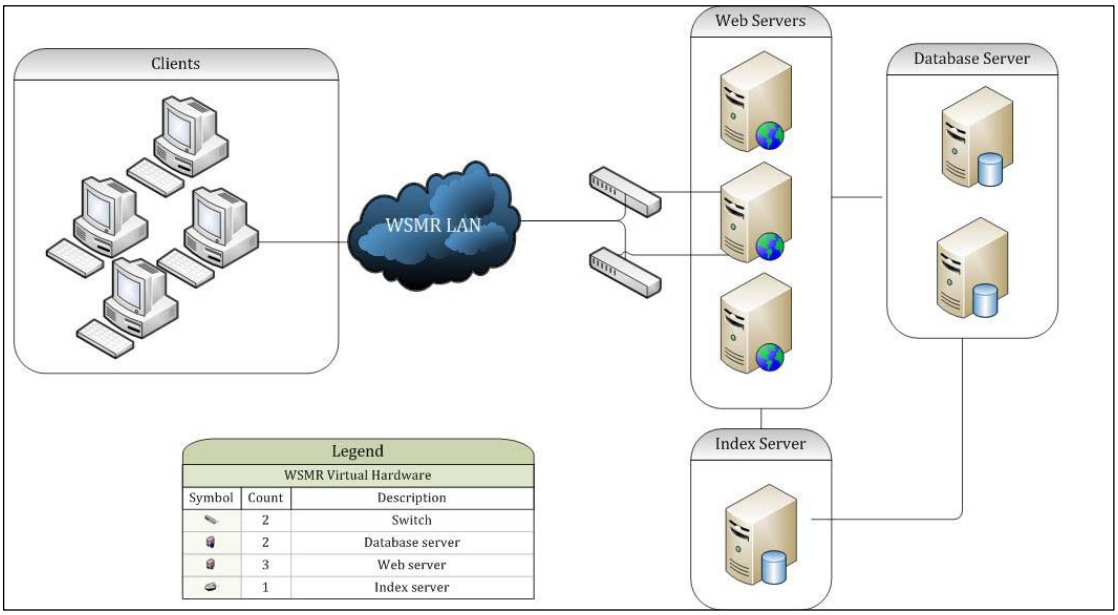
Develop procedures and tools that harvest and store all collected data.

# 6. Data Archival

## 6.1. Overview

Develop processes and tools that enable the archiving and remote access of collected data for a period of X years.

All test data will be retained at the storage facility.




## 6.2. Procedures

- During test execution, simulation and tactical data will be stored on the DCARS servers. These stored databases can be accessed via the DCARS clients as described in the next section, Using the DCARS client.
- DCARS archive all test data at the designated storage facility.

## 6.3. Test Metadata

Depends on type of information that should be captured to describe data collected, such as:

- Data Source
- Day of Test
- Type of Test
- Project Name

## 6.4. Data Analysis and Presentation

- Data Analysis. Describe how the processed data will be analyzed for the measure and what techniques or tools will be used (i.e. statistical tests, software tools, etc) Give suitable references when necessary.
- Data Evaluation. Discuss the evaluation techniques or procedures to explain the test results and data analysis. Discuss the impacts to the Critical Joint Issue(s).
- Data Presentation. Describe how the data will be presented in the data document supporting the test findings (i.e., graphs, narratives). Show an example of how the measure results will be reported in the final report using sample tables or plots if appropriate.

*This page intentionally left blank.*

## Annex A. Acronyms and Abbreviations

| | |
|---|---|
| ATCCS | Army Tactical Command & Control System |
| CMU | Connectivity Matrix Utility |
| COTS | Commercial Off-the-Shelf |
| CTO | Combined Test Organization |
| CTM | Capability Test Methodology |
| DAP | Data Analysis Plan |
| DAUVS | Digital Army USMTF (US Message Text Format) Variable Message Format (VMF) Stimulator |
| DCIT | Distributed Capabilities Integration Toolbox |
| DCM | Data Collection Matrix |
| DCARS | Digital Collection Analysis and Review System |
| DIS | Distributed Interactive Simulation |
| DMP | Data Management Plan |
| DWS | Data Warehouse System |
| FCS | Future Combat Systems |
| GIS | Geographic Information System |
| HLA | High Level Architecture |
| HSLT | High Speed LAN TAP |
| IDRL | Integrated Data Requirements List |
| INC | Interface Network Controller |
| JBD2 | Joint Battlespace Dynamic Deconfliction |
| JME | Joint Mission Environment |
| JMETC | Joint Mission Environment Test Capability |
| JOC-T | Joint Operational Context for Test |
| JTEM | Joint Test and Evaluation Methodology |
| JVMF | Joint Variable Message Format |
| LVC-DE | Live, Virtual, Constructive Distributed Environment |
| MHS | Message Handling System |
| M&P | Methods and Processes |
| M&S | Modeling and Simulation |

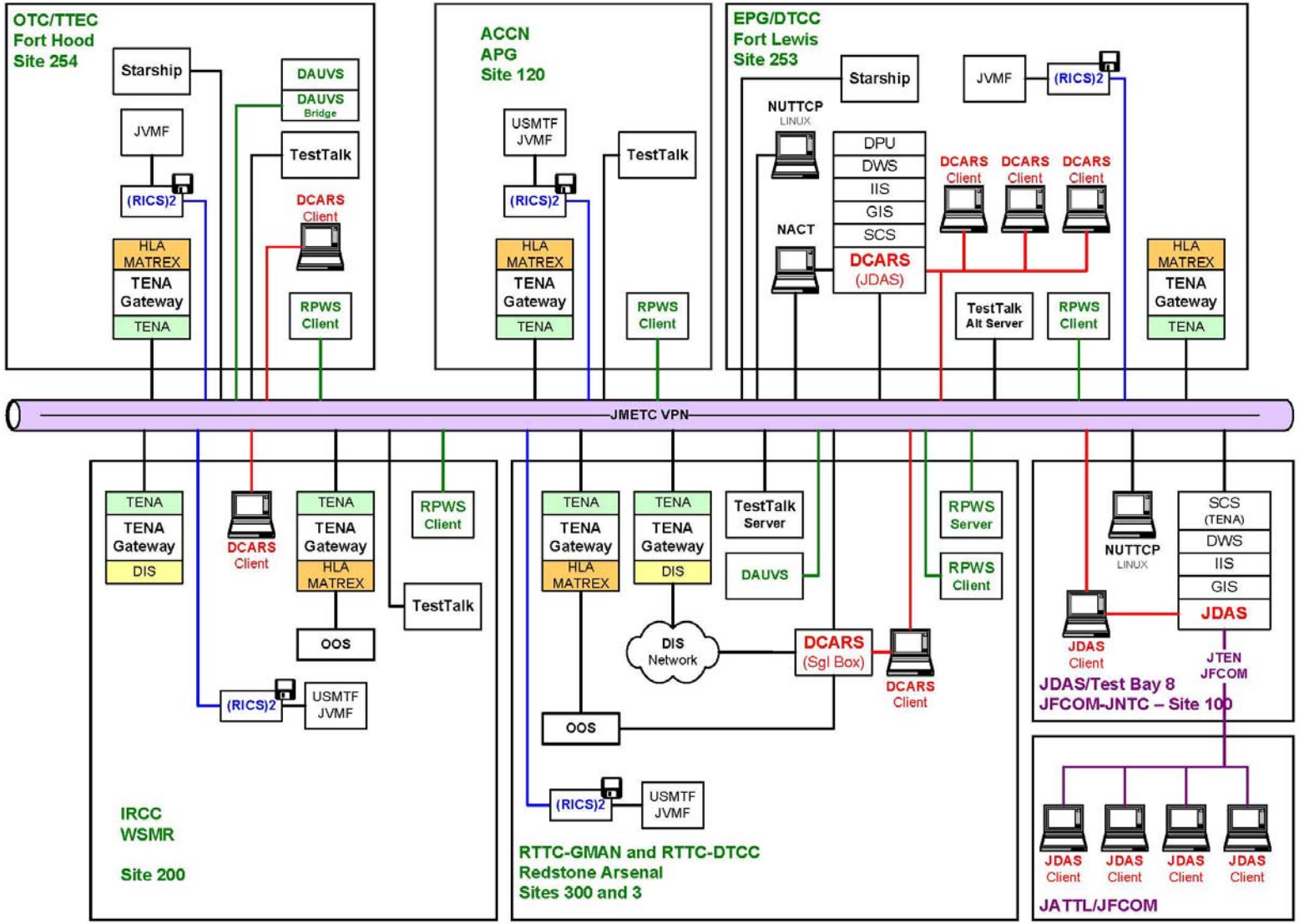
| | |
|---|---|
| MMOE | Mission Measure of Effectiveness |
| OT&E | Operational Test and Evaluation |
| OTG | Over-The-Horizon Gold |
| SCS | Simulation Collection System |
| SME | Subject Matter Expert |
| SQL | Standard Query Language |
| TADIL | Tactical Digital Information Link |
| TENA | Test and Training Enabling Architecture |
| TOEL | Time Ordered Event List |
| TP | Test Plan |
| TMOP | Task Measure of Performance |
| USMTF | United States Message Text Format |
| WAN | Wide Area Network |

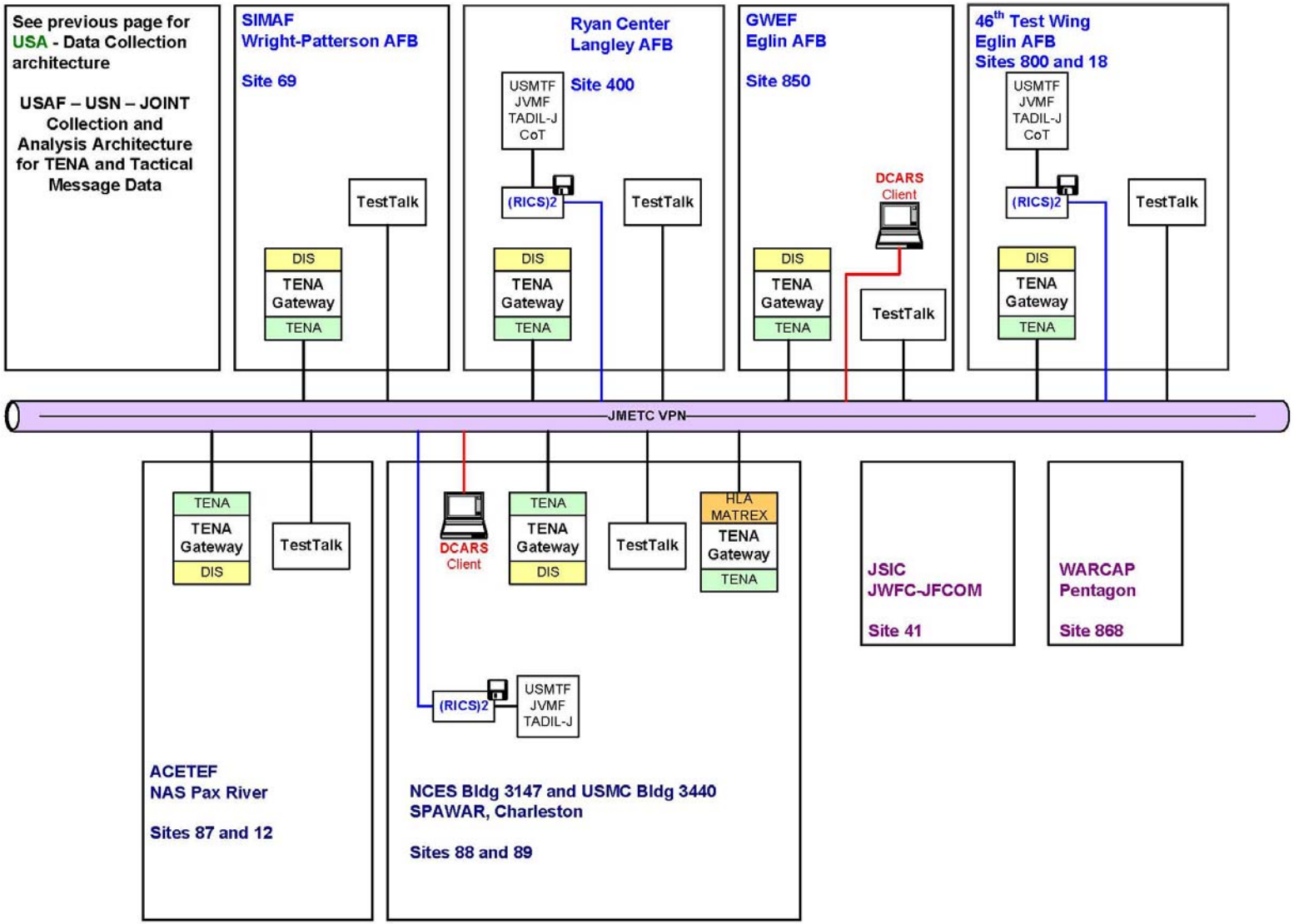
# Annex B. DCARS Software Component Functions

**Software Component Functions**

- Analyst Client
  - Battlefield Visualization
  - Static & Live Element Plotting
  - Replay Animation
  - After-Action Review
  - Activity & Performance Reports
  - Message Analysis & Review
  - Web Browser
  - Graphical User Interface
  - Office Applications
- Network Analysis and Collection Tool
  - Automatic collection
  - Network Link Status
  - Network Traffic History
- Geographic Information System (GIS)
  - Digital Terrain
  - Static Plotting Service
  - Live Stream Plotting Service
  - Client Control & Management
- IIS Web Server (IIS)
  - Active Web Browsing
  - Web Services
  - Client Interface
  - Security Classification
  - DCARS Web Site
- Analyst Client
  - Battlefield Visualization
  - Static & Live Element Plotting
  - Replay Animation
  - After-Action Review
  - Activity & Performance Reports
  - Message Analysis & Review
  - Web Browser
  - Graphical User Interface
  - Office Applications
- Network Analysis and Collection Tool
  - Automatic collection
  - Network Link Status
  - Network Traffic History

# Annex C. Sample Data Collection Network




See previous page for USA - Data Collection architecture
USAF – USN – JOINT Collection and Analysis Architecture for TENA and Tactical Message Data
SIMAF
Wright-Patterson AFB
Site 69
TestTalk
DIS
TENA Gateway
TENA
Ryan Center
Langley AFB
Site 400
USMTF
JVMF
TADIL-J
CoT
(RICS)2
TestTalk
DIS
TENA Gateway
TENA
GWEF
Eglin AFB
Site 850
DCARS Client
DIS
TENA Gateway
TENA
TestTalk
46th Test Wing
Eglin AFB
Sites 800 and 18
USMTF
JVMF
TADIL-J
CoT
(RICS)2
TestTalk
DIS
TENA Gateway
TENA
JMETC VPN
TENA
TENA Gateway
DIS
TestTalk
ACETEF
NAS Pax River
Sites 87 and 12
DCARS Client
TENA
TENA Gateway
DIS
TestTalk
HLA MATREX
TENA Gateway
TENA
(RICS)2
USMTF
JVMF
TADIL-J
NCES Bldg 3147 and USMC Bldg 3440
SPAWAR, Charleston
Sites 88 and 89
JSIC
JWFC-JFCOM
Site 41
WARCAP
Pentagon
Site 868

## Annex D. Data Levels

| Level | Description | Possible forms | Example of content | Disposition |
|---|---|---|---|---|
| Level 1<br><br>“Raw Data” | Data in their original form. Results of field trials just as recorded. | Complete data collection sheets, exposed camera film, voice recording tapes, original instrumentation magnetic tape or printouts, original videotapes, completed questionnaires, and/ or interview notes. | 1. All reported target presentations and detection.<br><br>2. Clock times of all events.<br><br>3. Azimuth and vertical angle from each flash base for each flash.<br><br>4. Recording tapes of interviews. Accumulated during trials for processing. | Usually discarded after use. Not published. |
| Level 2<br><br>“Reduced Data” | Data taken from the raw form and consolidated. Invalid or unnecessary data points deleted. Trials declared “No Test” deleted. | Confirmed and corrected data collection sheets, film with extraneous footage deleted, corrected tapes or printouts, and original raw data with “No Test” events marked out. | 1. Record of all valid detections.<br><br>2. Start and stop times of all applicable events.<br><br>3. Computed impact points of each round flashed.<br><br>4. Confirmed interview records. | Produced during processing. Usually discarded after use. Not published. |

| Level | Description | Possible forms | Example of content | Disposition |
|---|---|---|---|---|
| Level 3<br><br>“Ordered Data” | Data that have been checked for accuracy and arranged in convenient order for handling. Operations limited to counting and elementary arithmetic. | Spread sheet, tables, typed lists, ordered and labeled printouts, purified and ordered tape, edited film, and/or edited magnetic tapes. | 1. Counts of detections arranged in sets showing conditions under which detections occurred.<br><br>2. Elapsed times by type of event.<br><br>3. Impact points of rounds by condition under which fired.<br><br>4. Interview comments categorized by type. | Not usually published but made available to analysts. Usually stored in institutional databanks. All or part may be published as supplements to the test report. |
| Level 4<br><br>“Findings” or<br><br>“Summary Statistics” | Data that have been summarized by elementary mathematical operations. Operations limited to descriptive summaries without judgments or inferences. Does not go beyond what was observed in the test. | Tables or graphs showing totals, means, medians, modes, maximums, minimums, quartiles, deciles, percentiles, curves, or standard deviations. Qualitative data in form of lists, histograms, counts by type, or summary statements. | 1. Percentage of presentations detected.<br><br>2. Mean elapsed times.<br><br>3. Calculated probable errors about the centers of impact.<br><br>4. Bar graph showing relative frequency of each category of comment. | Published as the basic factual findings of the test. |

| Level | Description | Possible forms | Example of content | Disposition |
|---|---|---|---|---|
| Level 5<br><br>“Analysis” or<br><br>“Inferential Statistics” | Data resulting from statistical tests of hypothesis or interval estimation. Execution of planned analysis data. Includes both comparisons and statistical significance levels. Judgments limited to analysts’ selection of techniques and significant levels. | Results of primary statistical techniques such as T-tests, Chi-square, F-test, analysis of variance, regression analysis, contingency table analyses and other associated confidence levels. Follow- on tests of hypotheses arising from results of earlier analysis or fallback to alternate nonparametric technique when distribution of data does not support assumption of normality. Qualitative data in the form of prevailing consensus. | 1. Inferred probability of detection with its confidence interval.<br><br>2. Significance of difference between two mean elapsed times.<br><br>3. Significance of difference between observed probable error and criterion threshold.<br><br>4. Magnitude of difference between categories of comments. | Published in system evaluation reports. (If system evaluation report is part of test report, the level 5 analysis results are presented separately from the level 4 findings.) |
| Level 6<br>“Extended analysis” or operations | Data resulting from further analytic treatment going beyond primary statistical analysis, combination of analytic results from different sources, or exercise of simulation or models. Judgments limited to analysts’ choices only. | Insertion of test data into a computational model or a combat simulation, aggregation of data from different sources observing required disciplines, curve fitting and other analytic generalization, or other operations research techniques such as application of queuing theory, inventory theory, cost analysis, or decision analysis techniques. | 1. Computation of probability of hit based on target detection data from test combined with separate data or probability of hit given detection.<br>2. Exercise of attrition model using empirical test time’s distribution.<br>3. Determination of whether a trend can be identified from correlation of flash base accuracy data under stated conditions from different sources.<br>4. Delphi technique treatment of consensus of interview comments. | Published as appropriate in system evaluation reports. |

| Level | Description | Possible forms | Example of content | Disposition |
|---|---|---|---|---|
| Level 7 “Conclusion” or Evaluation | Data conclusions resulting from applying evaluative military judgments to analytic results. | Stated conclusions as to issues, position statements, and challenges to validity or analysis. | 1. Conclusion as to whether probability of detection is adequate.<br>2. Conclusion as to timeliness of system performance.<br>3. Conclusion as to military value of flash base accuracy.<br>4. Conclusion as to main problems identified by interviewees. | Published as the basic evaluative conclusions of system evaluation reports. |

# ANNEX J – SAMPLE DATA ANALYSIS PLAN (DAP)

CHAPTER 1 – INTRODUCTION
1.0 Overview
1.1 Background
1.2 Capability Gap Description
1.3 Test Goal & Objectives
1.4 Test Description
1.4.1 Mission Description
1.4.2 System of System Description
1.4.3 Threat Description
1.4.4 Environment Description

CHAPTER 2 – TEST CONCEPT
2.0 Overview
2.1 Issues
2.2 Evaluation Strategy
2.2.1 Scope
2.2.1.1 Assumptions
2.2.1.2 Constraints and Limitations
2.2.2 Test Methodology
2.2.2.1 Test Factors
2.2.2.2 Test Measures and Data Elements
2.2.2.3 Test Crosswalk
2.3 Risks

CHAPTER 3 – DATA ANALYSIS AND EVALUATION
3.0 Overview
3.1 Test Data Processing and Analysis
3.2 Joint Mission Effectiveness Evaluation
3.3 Data Management
3.3.1 Overview
3.3.2 Data Management Requirements
3.3.3 Data Validation Requirements
3.3.4 Data Processing

CHAPTER 4 – REPORTING REQUIREMENTS
4.0 Overview
4.1 Reports
4.2 Reporting Timelines
4.3 Capability Evaluation
4.3.1 Goal-Objectives Resolution
4.3.2 Issues Resolution

Appendixes:
A. IDRL

B. Crosswalk Matrix
C. Evaluation Views
D. Acronyms
E. Analysis Schedule

# SAMPLE DATA ANALYSIS PLAN (DAP)

## JOINT CAPABILITY
## UNDER TEST TITLE

DISTRIBUTION STATEMENT: Distribution limited to U.S. Government agencies and their supporting contractors. Other requests for this document shall be referred to …

*This page intentionally left blank*

# CHAPTER 1 – INTRODUCTION

## 1.0 Overview

Department of Defense (DoD) Strategic Planning Guidance demanded the creation of a joint environment testing capability. This demand led to the creation of a test methods and processes project and Joint Mission Environment Test Capability (JMETC) as part of the larger testing in a joint environment initiative. This evolved into the Joint Battlespace Dynamic Deconfliction test event whose architecture included multiple sites and multiple fully integrated systems, as well as a few partially integrated systems.

## 1.1 Background

Current joint force doctrine, from Joint Publication 3-09, Joint Fire Support; Joint Publication 3-52, Joint Doctrine for Airspace Control in the Combat Zone; Joint Publication 3-30, Command and Control for Joint Air Operations, provided the overview of the operational context for the test event will.

Combat zone airspace control increases combat effectiveness by promoting the safe, efficient, and flexible use of airspace with minimum restraint placed upon the airspace users. Airspace control includes coordinating, integrating, and regulating airspace to increase operational effectiveness. The airspace of the combat zone is a crucial part of the battlespace and is used by all components of the joint and multinational forces. A high concentration of friendly surface, subsurface, and air-launched weapon systems must share this airspace without unnecessarily hindering the application of combat power in accordance with the Joint Force Commander's intent. The primary goal of combat zone airspace control is to enhance the combat effectiveness of the joint force.

Airspace control procedures provide maximum flexibility through an effective mix of positive and procedural control measures. The control structure must encourage close coordination between joint force components to allow rapid concentration of combat power. The methods of airspace control vary throughout the range of military operations. They range from positive control of all air assets in an airspace control area to procedural control of all such assets, or any effective combination of the two. Airspace control procedures and systems need to accommodate these methods based on component, joint, and national capabilities and requirements.

## 1.2 Capability Gap Description

The capability gap under analysis in this test is a well integrated, synchronized C2 capability for immediate trigger events. The Joint Force must be able to quickly and accurately portray and quantify a complete integrated air picture and execute rapid deconfliction in response to a dynamically changing battlespace to support networked fires, air maneuver operations, integrated air and missile defense, command and control, and battlespace combat identification in a joint and coalition operational environment. This test will evaluate Joint Fires (JFIRES) and Joint Close Air Support (JCAS).

## 1.3 Test Goal & Objectives

The goal of this test is to evaluate the Joint Air Ground System (JAGS) solution for the conduct of Close Air Support and Joint Fires. This systems of systems (SoS) is a C2 capability which will analyze two variables. The first variable (factor) will be the battlespace management capability (Current versus Future). The second variable (factor) will be the C2 processes utilized by the forces (Current processes versus Future processes).

The test has two evaluation objectives:

- Determine which SoS configuration provides the best response to improvement in joint mission effectiveness.
- Determine if the SoS performed the task at the desired standards.

## 1.4 Test Description

The Test Event will provide a high fidelity, real-time, rapidly configurable, distributed network including virtual and constructive models linked with live systems, to evaluate C2 for JFIRES and JCAS. The event design is based on a Joint Operational Context for Test (JOC-T) drawing from authoritative planning sources. The result has been the design of a joint mission environment (JME) to stress the boundaries of a live, virtual, and constructive distributed environment (LVC-DE).

### 1.4.1 Mission Description

Commander's intent: On order, Blue forces conduct joint forcible entry operations to expand lodgment and control key infrastructure in order to facilitate rapid force build-up in the joint operations area (JOA).

The objectives of the mission with associated endstates are identified as:

- Expand lodgment and control key infrastructure to facilitate rapid force build-up – Expanded lodgment and control of key infrastructure
- Preserve key infrastructure and prevent environmental contamination – Key infrastructure preserved without environmental contamination
- Secure LOCs to facilitate flow of forces needed to conduct regime change operations – LOCs secured

To accomplish these objectives and reach the anticipated endstates, the Joint Force Commander is planning for the following effects:

- Threat forces destroyed or neutralized in JOA
- Threat forces unable to threaten key infrastructure
- Threat forces unable to hinder movement of Blue forces

***For purposes of this event, the focus will be on the first mission objective only - Expand lodgment and control key infrastructure to facilitate rapid force build-up.***

The Force Brigade Combat Team (FBCT) attacks to clear the enemy in zone and seize the airfield to secure key lines of communication for the rapid advance and passage of follow-on

forces. O/O FBCT conducts stability and reconstruction operations within its assigned zone of operations.

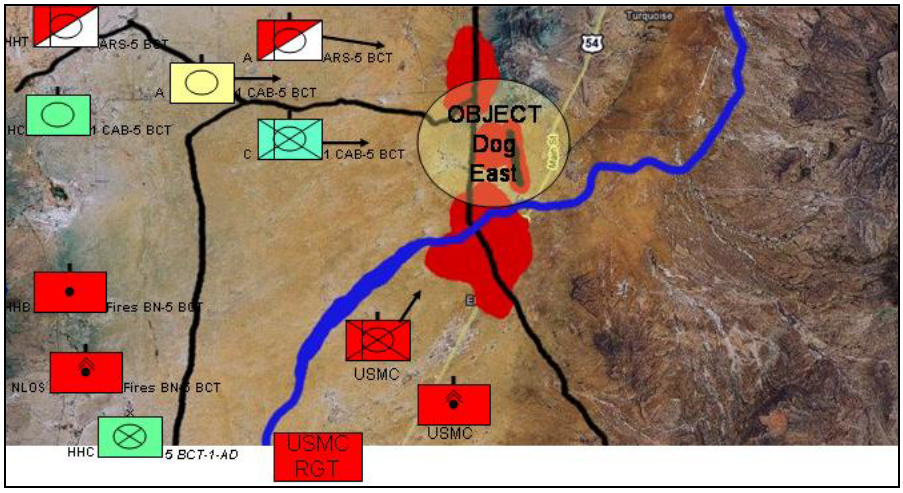



**Figure 1-1. Joint Forcible Entry Mission Overview**

The Test Event is focusing on six key JFIRES/JCAS mission task threads occurring simultaneously and/or sequentially as the battle unfolds:

<u>JFIRES Variations</u>

- USA MVR Observer to USMC HIMARS
- USMC MVR Observer to USA NLOS-LS
- USAF UAS to USA NLOS-LS
- USAF UAS/USA FSE to USAF Weaponized UAS

<u>JCAS Variations</u>

- USAF JTAC to Fixed Wing (USAF/USN/USMC)
- Marine JTAC to Fixed Wing (USAF/USN/USMC)

The commander plans to follow the mission task flow outlined in Figure 1-2. This task flow has been broken into detect, decide, deconflict, deliver, and assess portions. This process decomposition will allow for consistent analysis of the mission through all eight mission task threads.

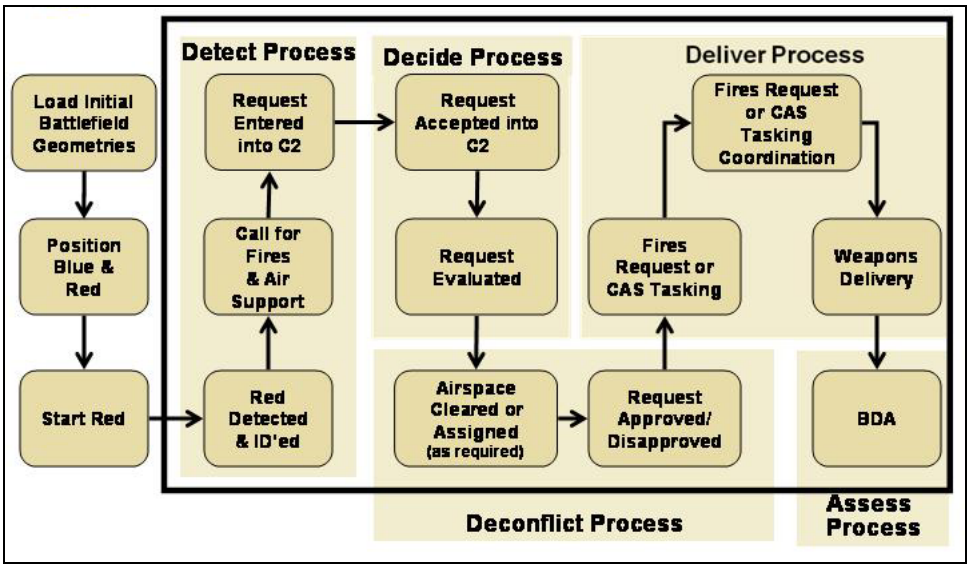



**Figure 1-2. Standard Mission Task Flow Diagram**

### 1.4.2 System of System Description

The Joint Air Ground System of Systems (JAGS) provides the capability to deconflict near-real-time tactical changes during the full range of military operations. Figure 1-3 depicts the JAGS configuration.

- A material solution depicting different configurations of the THE TEST EVENT SoS where the current SoS includes base processes and systems, while the future version includes future processes and systems.
- A non-material solution evaluates the command and control (C2) processes as either: (1) current procedural-based processes or (2) a future abbreviated process.

Of primary interest to the evaluation are the systems briefly described below. A full description of all the systems encompassing the JAGS can be found in the Event Management Plan.

A system provides the combat air forces and the joint/combined forces with an automated and integrated capability to plan and execute the air battle plan for operations and intelligence personnel at the combined AOC (CAOC) and individual unit levels. It provides the air commander with the means to plan, direct, and control all theater air operations in support of command objectives. It also coordinates with engaged ground and maritime elements.

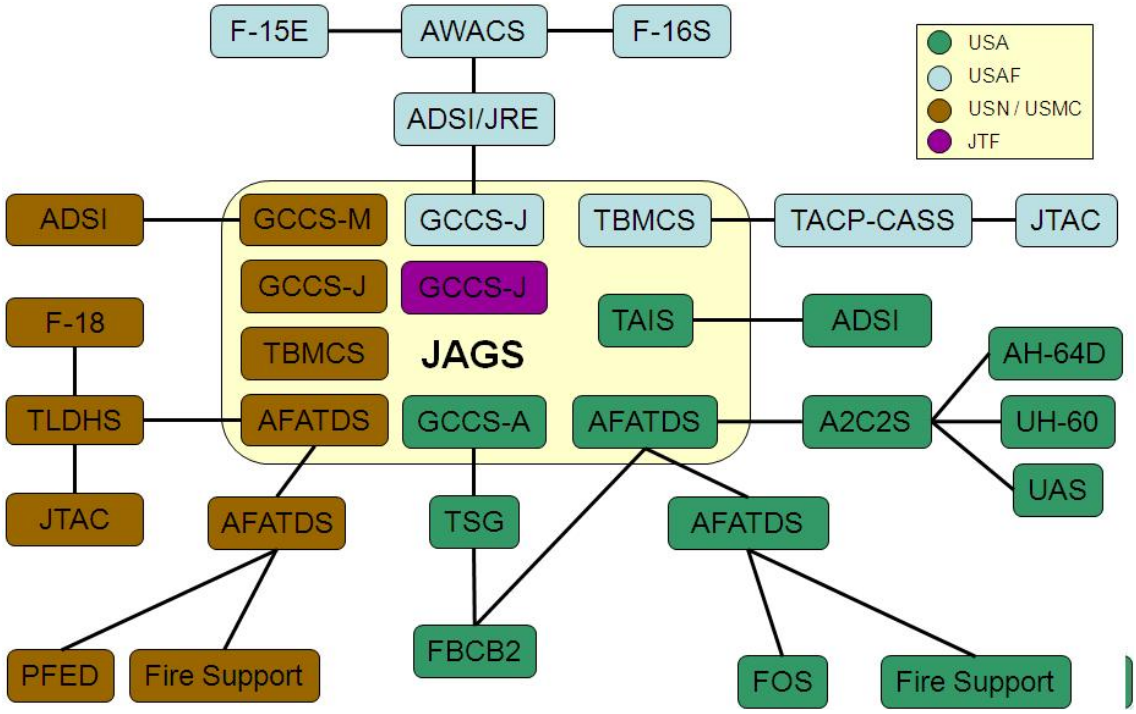



**Figure 1-3. Joint Air to Ground System Overview**

The Tactical Airspace Integration System (TAIS) is the Army's materiel solution for the integration and synchronization of Airspace Command and Control (AC2) and enroute Air Traffic Services (ATS) within the Army Battle Command System (ABCS). TAIS revolutionizes the management of airspace and conduct of air traffic services. TAIS focuses on interactions between aviation, missiles and unmanned systems and the development of infrastructure to examine airspace management in a joint context. Two versions of this system will be used during the test event to represent current and future processes.

### 1.4.3 Threat Description

The notional Red force is depicted in figure 1-4 and includes mechanized units, tanks, artillery, militia and air defense artillery.

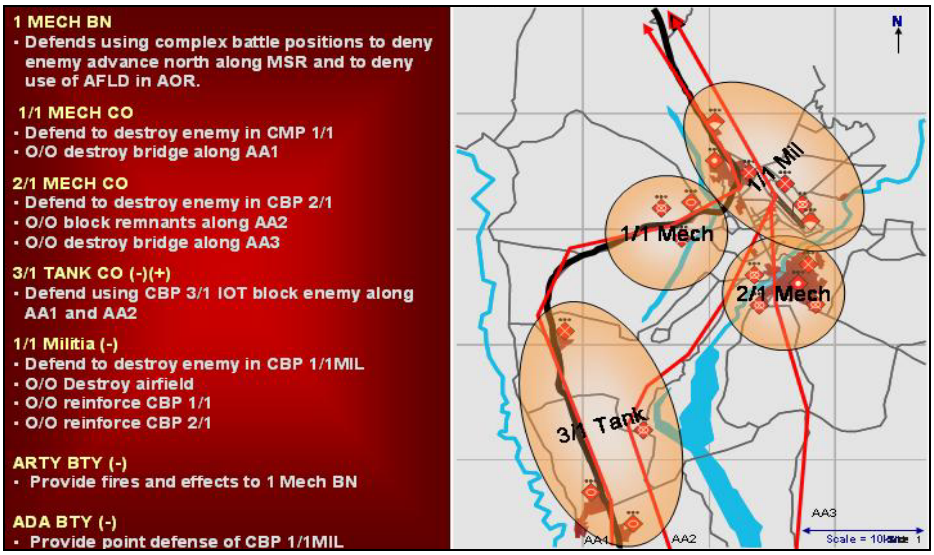



**Figure 1-4. THE TEST EVENT JOC-T**

### 1.4.4 Environment Description

The test assumes a non-urban setting comprised of primarily flat desert terrain. A small mountain range is in the region, but it does not impact on the operational planning or effects. Weather is clear and not a factor to the event.

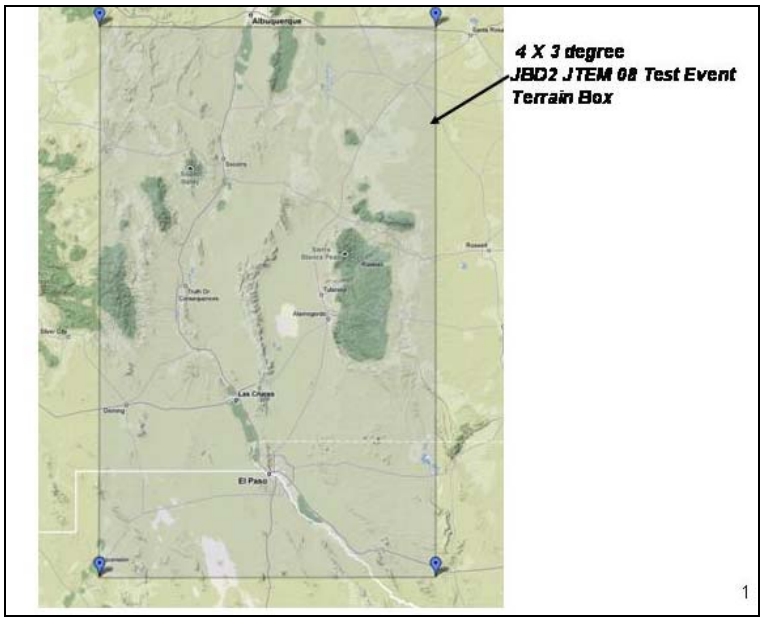



**Figure 1-5. THE TEST EVENT Environment View**

# CHAPTER 2 – TEST CONCEPT

## 2.0 Overview

## 2.1 Issues

As mentioned in chapter 1, the overall objective of the THE TEST EVENT Test Event is to assess the degree the Blue force can successfully conduct identified missions in the context of controlled changes to specific material and non-material factors, holding all other potential variables constant. From an operational view, the specific joint issues of the test are:

- Assess the ability to perform immediate battlespace deconfliction C2 by JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA
- Assess the ability to perform JCAS with JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA
- Assess the ability to perform JFIRES with JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA

## 2.2 Evaluation Strategy

An evaluation strategy is developed based on identified warfighter needs and includes the capability gaps, solutions, and implementations from the appropriate source documents (including DPS, Analytical Agenda, ICD, and CDD). This strategy involves deriving critical joint issues (CJIs) from the warfighter needs and mapping these CJIs to an evaluation framework consisting of initial test factors, measures, and relationships in order to demonstrate the warfighter needs are adequately addressed. CJIs are used to assess performance pertaining to capabilities supporting joint missions. A CJI for test and evaluation should be carefully structured as a question to address the key capability characteristics described in the joint capability documentation. It is important to state how the test issue contributes to achieving the desired mission end state outcomes in terms of mission desired effects. The CJIs should address the SoS capability to contribute to mission effectiveness, perform joint operational tasks and/or the SoS, system, or service attribute performance. CJIs are of primary importance to the appropriate authority when deciding whether to allow the SoS to advance to the next phase of development. There are three test factor dimensions in the evaluation strategy framework: joint mission(s) and task(s); threat and environmental conditions; and system of systems (SoS) configuration options across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) resources.

This analysis plan draws from an initial evaluation strategy described in annex G (this would normally be documented in a Test and Evaluation Strategy (TES) document). As stated in the Defense Acquisition Guidebook, "TES is an early T&E planning document that describes the T&E activities starting with Technology Development and continuing through System Development and Demonstration into Production and Deployment". Initial measures are also provided in the evaluation strategy framework to support evaluation of CJIs. Information gained by the evaluation will provide feedback to the acquisition life cycle mangers, and help mitigate development and other programmatic risks. The evaluation strategy should indicate a decision point of focus, (e.g., MS B, MS C, LRIP decision, etc.) for the evaluation data. The lowest level

of test measures are system and SoS attributes, which measure SUT and SoS performance, and include Critical Test Parameters and Key Performance Parameters. These are used to assess SoS and SUT performance and limitations based on specified system requirements and additional characteristics such as reliability, maintainability, human factors, and susceptibility characteristics. The next level of test measures are task measures of performance (TMOPs). TMOPs are used to assess friendly actions tied to measuring task accomplishment. The third level of test measures are mission measures of effectiveness (MMOEs). MMOEs are used to assess changes in system behavior, capability, or operational environment tied to measuring the attainment of a mission end state, achievement of a mission objective, or creation of a mission effect.

### 2.2.1 Scope

#### 2.2.1.1 Assumptions

Based upon discussions between THE TEST ORGANIZATION and CTO personnel, the following assumptions will be made during execution of the Test Event:

- The "Threat Units" participating in the Test Event will be static.
- "Threat Units" participating at the start of a Test Run will remain in the JOA for the duration of the Test Run. Additionally, no new "Threat Units" will enter the JOA once the Test Run has begun.
- "Threat Units" are either effective (fully functional) or ineffective (not functional); there will not be "in-between" levels of battle damage.
- An ACMREQ will be required for each THE TEST EVENT process.
- The number of "Threat Units" will be the same for each Test Run.
- All task threads will follow the Standard Mission Task Flow diagram (reference figure 1-2).
- For simplification in tracking responses to the CFFs/ASRs, the number of targets identified in the CFF/ASR will equate to the number of targets attacked by the respective JFIRES/JCAS mission (only applies to TMOPs 6, 7, 8).
- Assessment of all hypotheses will be evaluated against an alpha of 0.1.
- If a CFF/ASR is accepted but not decided (e.g., a CFF/ASR is accepted at the 58 minute mark and the trial ends prior to the issuance of a decision from the C2 authority), this occurrence will be collected, annotated as incomplete (with the reason), and not included as a data point in the analysis (only applies to TMOPs 1, 2, 6, 9 and MOSA 1).
- Each threat unit will have equal priority.
- The threat system initial starting positions will vary with each test run. This will minimize the possibility of players "gaming" the trials.
- The expected standards and the set of threat and environmental conditions will remain consistent throughout the test.

#### 2.2.1.2 Constraints and Limitations

- Due to the proposed threads developed for this test, the expected number of "Inter-Service" data points is projected to be almost three times greater than the number of "Intra-Service" data points. This may have an impact on the statistical calculations, e.g., if the disparity in sample size affects the variances of the

respective data sets, parametric methods to assess the data will not be statistically valid (only applies to TMOPs 2, 3, 4).

- The current test plan states a total of 16 runs will be conducted (four per day over four days), with each run requiring one hour to accomplish. Based upon this, the sample size is constrained to the number of data points eligible to be conducted over this time period.
- Each run will allow "free-play" to dictate the events occurring within them. This poses an issue if certain "test threads" are not selected as a solution to the "free-play," thereby creating the possibility a thread (or threads) is not exercised.

### 2.2.2 Test Methodology

#### 2.2.2.1 Test Factors

Two factors will be evaluated during the Test Event; one material, one non-material. The material factor attempts to discern if a difference will exist between the current and future Joint AGS SoS being implemented. The current JAGS SoS will be comprised of the Theater Battle Management Control System (TBMCS) and TAIS version __; the future will employ TBMCS and TAIS version __. The non-material factor will focus on the C2 processes, where one C2 process is procedural in nature and the other C2 process is more "expedient", where some of the procedures will be automated.

#### 2.2.2.2 Test Measures and Data Elements

The test design follows a traditional scientific methodology of *empirical-inductive* reasoning to evaluate casual relationships between capabilities and increased warfighting effectiveness. The scientific method employs a basic experimental[1] design process to determine if a proposed capability **A** causes the anticipated military effect **B** [TCCP GUIDEx 2006, pg 8]. This can be stated in terms of an experimental hypothesis, "If the proposed capability, then an improved mission effectiveness". There are three treatment dimensions in the capability evaluation test design: joint mission and tasks; threat and environmental conditions; and system of systems (SoS) configuration options across DOTMLPF resources. Evaluation dependent response measures are structured in three levels: mission measures of effectiveness (MMOEs), task measures of performance (TMOPs), and system/SoS attributes. These measures are outlined in terms of their nature (e.g., qualitative and quantitative), measurement units, and desired fidelity in the IDRL (annex B), as well as detailed in chapter three.

##### 2.2.2.2.1 Measures

The three different levels of measures and their relationships are highlighted in figure 2-1.

[1] An experiment can be defined as a process "…*to explore the effects of manipulating a variable*." [Shandish, et al 2002, pg 507]

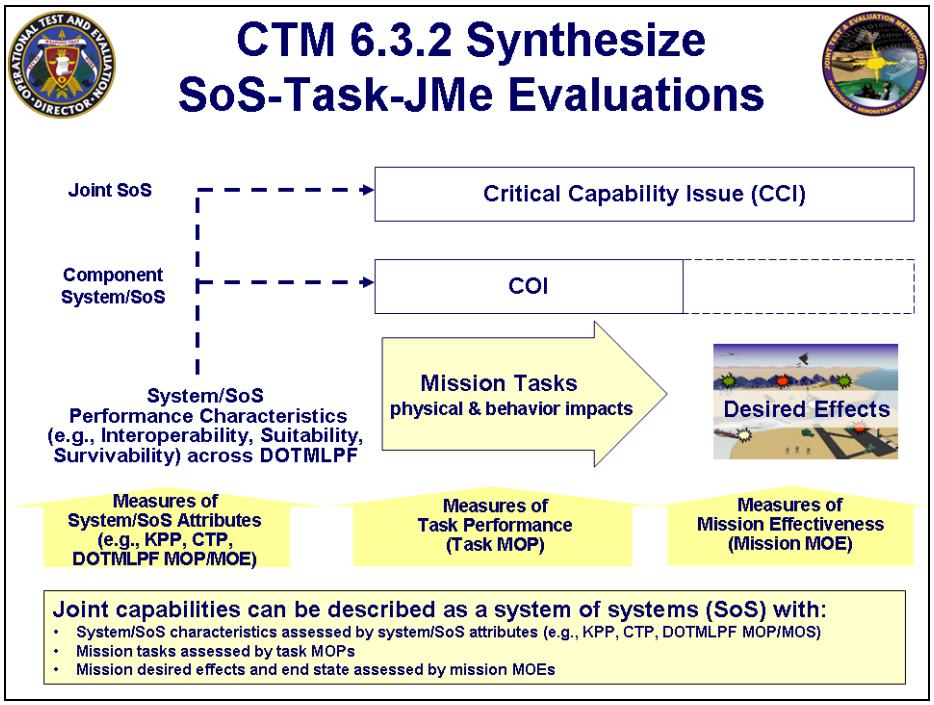



**Figure 2-1. Measures Framework**

These measures are described as:

(1) Mission Measures of Effectiveness (MMOE): Those measures evaluating achievement of desired mission end state outcomes in terms of mission desired effects. The MMOEs should measure both material and non-material factors to the SoS configuration.

(2) Task Measure of Performance (TMOP): In order to analyze Task Performance, mission-level data will be used to produce quantitative and qualitative mission effectiveness test results, using statistical and operational measures reflecting mission-level task outcomes. These measures will provide the statistical rigor and operational context to produce results analysts and SMEs will use for capability evaluations concerning SoS task performance to achieve desired mission outcomes. TMOPs are used to quantify mission task accomplishment. TMOPs are defined using the Universal Joint Task List (UJTL) structure of task, condition, and standard and are based on joint force commander mission tasks deemed essential to mission accomplishment using specified conditions and standards. The TMOPS are required to meet a preset threshold value.

(3) Measure of SoS attribute (MOSA): The MOSAs are quantitative or qualitative performance characteristic of a system or system of systems that make a significant contribution to the characteristics of the future joint force. Since the SoS attribute areas are required to have associated performance measures, the mapping below ties the individual systems to SoS mission areas and to measurable and testable SoS attributes.

Specific measures for this test are listed in table 2-1 by measurement type and supporting issue. Detailed information on each measure is provided later in this chapter.

**Table 2-1. Measures Overview**

| Measure Type | Measure | Supporting issue |
|---|---|---|
| MMOE 1 | Threat System Ineffectiveness | Mission Desired Effect – Destroy/neutralize threat forces |
| MMOE 2 | Threat System Ineffectiveness Rate | Mission Desired Effect – Rate threat forces are destroyed / neutralized |
| TMOP 1 | Time to C2 THE TEST EVENT Indirect Fires Missions | THE TEST EVENT UJTL 5.x – Execute Immediate Deconfliction C2 |
| TMOP 2 | Time to C2 THE TEST EVENT Indirect Fires Missions by Task Thread | THE TEST EVENT UJTL 5.x – Execute Immediate Deconfliction C2 |
| TMOP 3 | Time to C2 THE TEST EVENT Indirect Fires Missions when airspace is controlled by requesting Service | THE TEST EVENT UJTL 5.x – Execute Immediate Deconfliction C2 |
| TMOP 4 | Time to C2 THE TEST EVENT Indirect Fires Missions when airspace is controlled by Service different from requesting Service | THE TEST EVENT UJTL 5.x – Execute Immediate Deconfliction C2 |
| TMOP 5 | Number of Indirect Fires missions denied (with reason for denial) | THE TEST EVENT UJTL 5.x – Execute Immediate Deconfliction C2 |
| TMOP 6 | Time to get ordnance on target after initiating air-to-surface request | UJTL TA 3.2.2 – Conduct Close Air Support |
| TMOP 7 | Percent of targets attacked IAW JCAS | UJTL TA 3.2.2 – Conduct Close Air Support |
| TMOP 8 | Percent of targets attacked IAW JFIRES | UJTL TA 3.2.1 – Conduct JFIRES |
| TMOP 9 | Time to get ordnance on target after initiating task (immediate CFF) | UJTL TA 3.2.1 – Conduct JFIRES |
| MOSA 1 | Speed of Decision | SoS Characteristic: Speed of Command |
| MOSA 2 | Speed of Deconfliction | SoS Characteristic: Speed of Command |
| MOSA 3 | Force Simultaneity | SoS Characteristic: Force Agility |

#### 2.2.2.2.2 Data Elements

Each measure requires different sets of data elements to be collected during the test event in order to do the analysis and evaluation of the measure. The Integrated Data Requirements List (IDRL) in annex B lists the complete set of data elements required for all measures. In summary, data is required for:

- The execution of all test runs to include start and stop times
- The status of all threat systems
- The instantiation of every indirect fires request
- Every THE TEST EVENT process to include sub-process times and status

Additional discussion of the data elements and the analysis are included later in this chapter and in chapter 3.

#### 2.2.2.3 Test Crosswalk

There are multiple issues, factors, standards and conditions affecting the outcome of a test. Due to a variety of constraints (such as time, limited resources, etc), it is not possible to test every aspect mentioned. For THE TEST EVENT, the standards and conditions will be held constant throughout all 16-runs. Additionally, the test event will focus on the first issue, "Assess the ability to perform immediate battlespace deconfliction C2 by JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA". To document and obtain meaningful results from the remaining aspects, a matrix was configured to ensure each aspect would be presented in combination with the others (reference annex C for the Crosswalk Matrix).

### 2.2.3 Test Measure Descriptions

#### 2.2.3.1 MMOE 1: Threat Systems Combat Ineffectiveness

(1) **Description**: "Threat systems combat ineffectiveness" measures a count of threat systems that become ineffective during a test run. Quantifying threat systems ineffectiveness leads to a mission desired effect of threat forces destroyed or neutralized in the JOA resulting in a mission end state of expanded lodgment and control of key infrastructure in the conduct of a joint forcible entry mission.

(2) **Additional Assumptions and Constraints**:

- The level of threat is consistent across test runs. This will reduce variability in the results from outside influences not related to the change in independent variables.
- The number of test runs is constrained by resources and available time.

(3) **Evaluation Criteria**: The evaluation will occur across different material and non-material configurations of the event's C2 SoS including a current SoS and a future SoS. Each test trial will utilize a different SoS configuration expected to impact threat systems combat ineffectiveness. Hypothetically, a more efficient, automated system of systems will provide a timely and accurate C2 process resulting in effective target acquisition and processing. No documented criteria or standards are available for this measure.

(4) **Response Variables**: The response (dependent) variable is the count of threat systems rendered ineffective (killed or neutralized) at the end of each test run. Threat systems are those enemy systems of interest in the JOA at trial start time. A threat system is considered killed when it can no longer move, shoot, and sense. The system no longer possesses a capability. Once killed, the threat system remains killed until trial end. A threat system is considered neutralized when it has been attacked and can no longer shoot, but is not killed. The threat system becomes combat ineffective but still retains some capability to move or sense.

(5) **Scope and Methodology**: Recording the number of ineffective threat systems at the start and stop of each test run will meet the minimum data requirements for this measure. Additional data beneficial to understanding the measure and impacts of the capability on mission effectiveness would include the time each threat system became ineffective. Therefore, part of the methodology should include a means to collect the time each threat system became ineffective. This measure should be evaluated across the different test factors and thus requires collection for every test run.

(6) **Expected Response**: It is expected the threat systems combat ineffectiveness value for each test trial will vary, but not significantly. Since the sensors will not vary across the test event, the number of threat systems detected is expected to remain somewhat consistent. Additionally, the kill chain is not expected to alter the results and will remain consistent within the test event. The C2 process will change. Based on time differences in the C2 processes, this will have some impact on threat system ineffectiveness. A more efficient THE TEST EVENT C2 process will enable faster and earlier prosecution of threat targets. Therefore, it is expected SoS configurations which are more effective will result in earlier threat systems ineffectiveness times, and thus a incremental improvement in mission effectiveness. It is desirable for the future SoS configurations to result in a more efficient THE TEST EVENT process, otherwise a transition to a future SoS may not be valid.

#### 2.2.3.3 TMOP 1: C2 THE TEST EVENT Indirect Fire (JCAS & JFIRES) Tasks

(1) **Description**: This measure will ascertain how long it takes from the issuance of a call-for-fire (CFF) request / air support request (ASR) to the time it takes the C2 authority to approve or deny it. Collecting information on this TMOP will contribute to the evaluation of the notional event C2 Joint Task (UJTL TA 5.x).

(2) **Additional Assumptions and Constraints:**
None

(3) **Evaluation Criteria**

- **Task Level:** $T_D \geq T_{ei} - T_{si}$**.** The amount of time it takes from acceptance of a CFF/ASR, $T_{si}$, until the C2 authority issues its response, $T_{ei}$, will be less than or equal to the Desired Blue Time, $T_D$. The requirement is for <u>all</u> C2 times to be less than or equal to $T_D$. (Reference: Notional – Need operational community to establish the "Desired Blue Time, $T_D$", to C2 indirect fire missions by task thread.)

- **Mission Level:** To determine if the material and non-material factors selected to emulate the current and future SoS participating in THE TEST EVENT are statistically different. The data supporting this TMOP will be used to test the hypothesis:
  - $H_a$: $\mu_{CT} \neq \mu_{FT}$ : The material and non-material factors selected to emulate the future SoS will be different in the time from acceptance of a CFF/ASR to the time it takes the C2 authority to approve or deny it, than those factors emulating the current SoS over the same interval ($\mu_{CT}$ represents the average time it takes the current SoS and $\mu_{FT}$ represents the average time it takes the future SoS).
  - $H_o$: $\mu_{CT} = \mu_{FT}$ : The material and non-material factors selected to emulate the current and future SoS will be statistically equal in the time it takes from acceptance of a CFF/ASR to the time it takes the C2 authority to approve or deny it.

(4) **Response Dependent Variables:** C2 time. The start time, $T_{si}$, equates to arrival of the CFF/ASR into the C2 Center of the $i^{th}$ CFF/ASR. The end time, $T_{ei}$, represents the time the C2 authority issues its decision whether the $i^{th}$ CFF/ASR has been approved or denied. The average time for the current SoS will be represented by $\bar{x}_{CT}$ (representing the sample mean of the data relevant to the current factors – which will be the statistic used in place of $\mu_{CT,}$ the population mean); the average time for the future SoS will be represented by $\bar{x}_{FT}$ (representing the data relevant to the future factors – which will be the statistic used in place of $\mu_{FT,}$ the population mean of the future factors).

(5) **Scope and Methodology:** The start time for this measurement will occur when a CFF/ASR is accepted within the C2 Center ($T_{si}$); its corresponding end time will occur when the C2 authority transmits the decision to the supported unit regarding whether the CFF/ASR is approved or denied ($T_{ei}$). Each CFF/ASR will be assigned a number to differentiate it from other CFFs/ASRs (represented by the counter, i). The respective current and future averages ( $\bar{x}_{CT}$ and $\bar{x}_{FT}$ ) will be calculated by summing the times relevant to each and dividing by the number in each sample test run.

(6) **Expected Response:** Upon completion of each test run, there should be "n" times collected, one for each CFF/ASR accepted and decided. Variability across test runs in the same test trial is expected and will provide insight on the distribution of the C2 times. The variability of the C2 times across test trials (current and future SoS) is expected to be most influenced by whether the weapon system selected in response to the CFF/ASR conflicts with the current "owner" of the airspace the weapon will have to traverse. For example, if the response to the ASR involves fighter aircraft (a JCAS mission), and the airspace the aircraft must fly is currently "owned" by the ground forces performing a JFIRES mission, time must be allocated to ensure the airspace gets deconflicted.

### 2.2.3.12 MOSA 1: Speed of Decision

(1) **Description**: This measure will ascertain how long it takes from when a CFF request / ASR enters the C2 queue to the time the C2 authority has evaluated it (not including the

time taken to deconflict the airspace). Collecting information on this MOSA will contribute to the evaluation of the time required to complete one full cycle of Boyd's observe-orient-decide-act (OODA) loop.

(2) **Additional Assumptions and Constraints:**
None

(3) **Evaluation Criteria**

- **Task Level:** $T_D \geq T_{ei} - T_{si}$. The amount of time it takes from acceptance of a CFF/ASR, $T_{si}$, until the C2 authority has evaluated the CFF/ASR, $T_{ei}$, (minus time taken to deconflict the airspace, $T_{di}$), will be less than or equal to the Desired Blue Time, $T_D$. The requirement is for <u>all</u> decision times to be less than or equal to $T_D$. (Reference: Network Centric Warfare – Its Nature and Modeling, Fewell, M.P. and Hazen, Mark G.) (Need operational community to establish the "Desired Blue Time, $T_D$", to C2 indirect fire missions by task thread.)
- **Mission Level:** To determine if the material and non-material factors selected to emulate the current and future SoS participating in THE TEST EVENT are statistically different. The data supporting this TMOP will be used to test the hypothesis:
  - $\underline{H_a}$: $\mu_{CT} > \mu_{FT}$ : The material and non-material factors selected to emulate the current SoS will have a longer average time, from the acceptance of a CFF/ASR to the time it takes the C2 authority to evaluate it, than the same process executed under the future SoS (where, $\mu_{CT}$ represents the average time it takes the current SoS and $\mu_{FT}$ represents the average time it takes the future SoS).
  - $\underline{H_o}$: $\mu_{CT} \leq \mu_{FT}$ : The material and non-material factors selected to emulate the current SoS will have a shorter average time (or statistically equivalent), from the acceptance of a CFF/ASR to the time it takes the C2 authority to evaluate it, than the same process executed under the future SoS.

(4) **Response Variables:** Time. The start time, $T_{si}$, equates to the C2 Center accepting the $i^{th}$ CFF/ASR. The end time, $T_{ei}$, represents the time the C2 authority has evaluated the $i^{th}$ CFF/ASR for action.

(5) **Scope and Methodology:** The start time for this measurement will occur when a CFF/ASR is accepted ($T_{si}$); its corresponding end time will occur when the C2 authority has evaluated the CFF/ASR for action ($T_{ei}$). Since the focus of this MOSA is "Speed of Decision", the deconfliction time in the C2 process is not included. Each CFF/ASR will be assigned a number to differentiate it from other CFFs/ASRs (represented by the counter, i).

(6) **Expected Response:** Upon completion of each test run, there should be "n" data points collected, one for each CFF/ASR initiated. The variability amongst the current and future times is expected to be most influenced by whether the weapon system selected in response to the CFF/ASR conflicts with the current "owner" of the airspace the weapon will have to traverse. For example, if the response to the ASR involves aircraft (a JCAS mission), and the airspace the aircraft must fly is currently "owned" by

the ground forces performing a JFIRES mission, time must be allocated to ensure the airspace gets de-conflicted.

### 2.3 Risks

There are risks associated with the development of the test concept and evaluation strategy.

- An insufficient amount of data will be collected to allow for statistical evaluation of the measures.
- There may be little to no correlation between task performance and mission accomplishment.
- Inconsistencies may exist in data collection across different task threads.

# CHAPTER 3 – ANALYSIS AND EVALUATION

## 3.0 Overview

This chapter provides detailed information on the analysis and evaluation of the joint capability. This includes steps to process the test data using various statistical and analytical tools and techniques. Details are also provided to evaluate the joint capability for mission and task level performance.

## 3.1 Test Data Processing and Analysis

This section provides detailed information on the data sampling requirements, data processing, and analysis techniques.

### 3.1.1 Data Sample Requirements

#### 3.1.1.1 Test Run Summary Data

Data summarized for each test run (i.e., count of threat systems ineffective) will only provide one data point for each test run. To provide adequate sample data to conduct statistical tests across test trials, a large number of test runs would be required. Even if this was possible, it is questionable if the data would follow a normal distribution. The propensity to detect and neutralize (or kill) threat systems is greater earlier in a test run due to larger number of effective threat systems at the start of a test run and also due to a full strength of Blue assets. As time progresses, Blue may lose assets as well as the threat systems getting depleted. This would infer a tendency to detect and neutralize threat systems in a gamma or exponential distribution. Thus the mean value of threat systems ineffective for a set of test runs in each test trial would probably be skewed to the right. Based on this, a standard ANOVA test would not be acceptable, and therefore would require a non-parametric test to compare the population distributions for each test trial. The Kruskal-Wallis H Test[2] would be adequate to determine if the test trial distributions for average threat systems ineffective at test run completion are identical versus them differing in location. The assumptions for this test are the samples randomly and independently drawn, and there are five or more measurements in each sample. The second assumption requires five runs for each test trial as a minimum. Given the test design currently plans for four runs, based on resource and time restrictions, and there are four trials instead of the minimum three required by the test, this test may still be used with the knowledge it will not have as much power in determining differences in distribution locations (differences in the means).

#### 3.1.1.2 Test Run Systems Data

An adequate number of threat systems are required for each test trial in order to statistically evaluate differences in systems data (i.e. mean ineffectiveness time) for each treatment (test trial). An estimate of sample size will allow for determining the number of threat systems which must exist within each test trial. Making some estimates in

[2] Mendenhall, W et al, *Mathematical Statistics with Applications*, Fourth Edition, page 699

variability of the expected data and using a 90% confidence level, estimating a sample size for a different in means of 1 minute can be calculated as[3]:

$$t_{\alpha/2} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad = \quad 1$$

where s = sample standard deviation; n = sample size

Assuming a 10 minute range in data values, s can be estimated as one-fourth of the range (s = 10/4 = 2.5). The value of $t_{\alpha/2}$ for moderate values of $n_i$ is approximately 1.7. Thus:

$$t_{\alpha/2} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad = \quad 1 \qquad (1.7)(2.5)\sqrt{\frac{2}{n_i}} \quad = \quad 1 \qquad n_i \approx 36$$

Assuming four test runs for each test trial, the minimum number of threat systems in a trial would need to be 9 in order to provide some statistical conclusions as to differences in the mean threat system ineffectiveness time.

### 3.1.2 Analysis Techniques

#### 3.1.2.1 Distribution Analysis

The distribution of data across test runs could be plotted on a box plot to evaluate possible outliers. Figure 1 shows an example of sample indirect fires request time data displayed in a box plot. What is seen in the example is a distribution skewed to the upper values and one data point that appears to be an outlier at 39 minutes. In the actual test data, this could suggest the SoS configuration for that test run did not provided good results and may be shown to be statistically different. In this case we would be interested in proving the outlier is outside a given confidence interval. Assuming a 90% confidence level and using a t-distribution for a small sample size, we can calculate for the sample data below a confidence interval of:

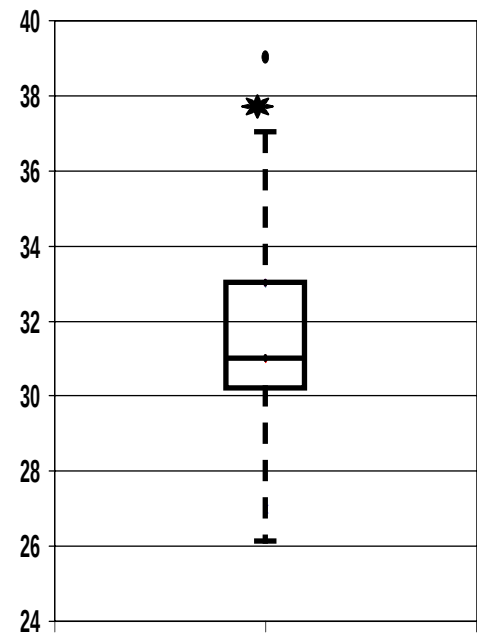



**Figure 3-1. Sample Box Plot**

$$\bar{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) \quad = \quad 31.5 \pm 1.833 * \left(\frac{3.44}{\sqrt{10}}\right) \quad = \quad 31.5 \pm 2.0$$

Therefore in this example, the outlier value of 39 falls outside the confidence interval and is statistically considered not equal to the sample mean. This assessment may lead to

[3] Mendenhall, W et al, *Mathematical Statistics with Applications*, Fourth Edition, page 622

further investigation regarding why the outlier occurred and whether the SoS configuration was a primary factor.

### 3.1.2.2 Non-Parametric Test - Kruskal-Wallis H Test

The Kruskal-Wallis H Test is a non-parametric test to compare k populations. The test may be described as:
$H_o$: The *k* populations are identical
$H_a$: At least two of the populations differ in location

Test Statistic: $$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

where: $n_i$ = Number of measurements in sample from population *i*
$R_i$ = Rank sum for sample *i*, where the rank of each measurement is computed according to its relative size in the overall set of $n = n_1 + n_2 + \ldots n_k$ observations formed by combining the data from all *k* samples.

Rejection Region: Reject $H_o \; if \; H > \chi^2_\alpha$ with (k-1) degrees of freedom

Assumptions: The *k* samples are randomly and independently drawn. There are five or more measurements in each sample.

### 3.1.2.3 Two-Sample Hypothesis Test of the Mean (Student's t-Test / Z Distribution)

The Student's t-test / Z Distribution are parametric tests to compare the means of two sets of data. Assumptions required to execute these tests are: the two population variances are not significantly different, and if the samples are small ($n_1 + n_2 < 30$, thus the t-distribution is selected), the parent populations are normally distributed. For either the t or Z-tests, one of the following hypotheses will be selected:

| Two-Tail | Upper-Tail | Lower-Tail |
|---|---|---|
| $H_o$: $\mu_1 = \mu_2$ | $H_o$: $\mu_1 \leq \mu_2$ | $H_o$: $\mu_1 \geq \mu_2$ |
| $H_a$: $\mu_1 \neq \mu_2$ | $H_o$: $\mu_1 > \mu_2$ | $H_o$: $\mu_1 < \mu_2$ |

Regardless of the hypothesis selected, the test statistic, $t_0$ ($Z_0$), is calculated as:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2)}{S_p\sqrt{1/n_1 + 1/n_2}}, \text{ where } S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where: $\bar{x}_i, s_i^2, n_i$ represent the sample means, variances and sample sizes, respectively, and $S_p$ represents the pooled standard deviation of the two samples.

Rejection Region: For one-tail tests, Reject $H_o$ if $t_0 > t_{(\alpha,\nu)}$, where $\nu = n_1 + n_2 - 2$ degrees of freedom; for two-tail test, Reject $H_o$ if $t_0 > t_{(\alpha/2,\nu)}$, where $\nu = n_1 + n_2 - 2$ degrees of freedom; if $n_1 + n_2 \geq 30$, use $Z_0$ and $Z_\alpha$ in place of $t_0$ and $t_{(\alpha,\nu)}$, respectively.

### 3.1.2.4 Analysis of Variance (ANOVA)

Analysis of Variance is a parametric test designed for use when comparing the means of more than two sets of data (it can be used for comparing two sets, but the t / Z-tests are simpler with respect to the amount of calculations required to make a conclusion about the hypothesis).

In general, the hypothesis for this test is:

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_n$$
$$H_a: \mu_1 \neq \mu_2 \neq \ldots \neq \mu_n$$

NOTE: If the alternate hypothesis is concluded ($H_a$), the interpretation is "the $\mu_i$ are not all equal". This conclusion does not state which means are statistically different (see the Tukey Post Hoc comparison test in section 3.1.2.5 for the answer).

The test statistic for the ANOVA is $F_0$. $F_0$ represents the ratio between the amount of variability contained within the factors being tested (mean squared between, or MSB), divided by the amount of variability inherent within the test itself (mean squared error, MSE). If the means being compared are all statistically the same, the ratio of the two variabilities will be approximately equal. The equations for each estimate of variance are:

$$\text{MSB} = \frac{\sum_{1}^{L} n_j \left(\bar{y}_j - \bar{y}\right)^2}{L-1}$$

$$\text{MSE} = \frac{\sum_{1}^{L} \left(n_j - 1\right) s_j^2}{\sum_{1}^{L} \left(n_j - 1\right)}$$

where, L = the number of factors to be compared
$n_j$ = sample size for the $j^{th}$ factor
$s_j^2$ = variance for the $j^{th}$ factor
$\bar{y}_j$ = average of the $j^{th}$ factor
$\bar{y}$ = grand average of all the factors

Once the two estimates for variance are calculated, the test statistic is:

$$F_0 = MSB/MSE$$

Rejection Region: If the test statistic is greater than the critical value ($F_{\nu_1,\nu_2}$, taken from a standard "F-table" from any statistics text-book), Reject $H_0$, and conclude at least one of the means being compared is statistically different from the others. Mathematically this is represented as:

$$\text{Rej } H_0 \text{ if: } F_0 > F_{\nu_1,\nu_2}$$

$$\text{where: } \nu_1 = L - 1 \text{ and } \nu_2 = \sum_1^L (n_j - 1)$$

### 3.1.2.5 Tukey Post Hoc Comparison Test

The Tukey Post Hoc comparison test is a procedure which allows for testing all the "pairs" of means present in an ANOVA test. Use of this test is pertinent when the ANOVA test statistic, $F_0$, was significant. The procedure is as follows:

First, compute the critical Tukey distance:

$$d_T = q_T \sqrt{\frac{MSE}{n_j}}$$

where $q_T$ is found from a statistical table, MSE is taken from the ANOVA test, and all $n_j$ are assumed to be equal. When the sample sizes are not equal, a technique to determine a representative $n_j$ is called the "harmonic" method. The generic equation for the "harmonic" method is:

$$\text{sample size} = \frac{j}{\left(\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_j}\right)}$$

where j represents the total number of samples, and each *n* represents the size of each sample.

Second, calculate the absolute difference between each of the sample means from the ANOVA test (it may be beneficial to create a table – see Table 3-8 for an example). The Tukey test compares the critical Tukey distance, $d_T$, against the absolute distance between each of the sample means. Any "pair" of differences greater than $d_T$ can be concluded as significant at the chosen $\alpha$-level. In other words, if the difference of any "pair" of means exceeds $d_T$, those two means are statistically different from each other.

Table 3.8 – Example of Absolute differences of Sample Means

| | $\bar{y}_1$ | $\bar{y}_2$ | $\cdots$ | $\bar{y}_n$ |
|---|---|---|---|---|
| $\bar{y}_1$ | 0 | | | |
| $\bar{y}_2$ | $\lvert \bar{y}_2 - \bar{y}_1 \rvert$ | 0 | | |
| : | | | | |
| $\bar{y}_n$ | $\lvert \bar{y}_n - \bar{y}_1 \rvert$ | $\lvert \bar{y}_n - \bar{y}_2 \rvert$ | | 0 |

#### 3.1.2.6 Half Normal Probability Plot

The Half Normal Probability Plot

### 3.1.3 Task Thread Identification

As mentioned in chapter 1, the THE TEST EVENT THE TEST ORGANIZATION 08 Test Event is focusing on 6 JFIRES / JCAS mission task threads occurring simultaneously and/or sequentially as the battle unfolds:

<u>JFIRES Variations</u>

- USA MVR Observer to USMC HIMARS
- USMC MVR Observer to USA NLOS-LS
- USAF UAS to USA NLOS-LS
- USAF UAS/USA FSE to USAF Weaponized UAS

<u>JCAS Variations</u>

- USAF JTAC to Fixed Wing (USAF/USN/USMC)
- Marine JTAC to Fixed Wing (USAF/USN/USMC)

These six mission task threads can be further broken out to identify each possible combination of Services. This also allows for a "numbering scheme" to be assigned to each thread – necessary for "flagging" of the thread's occurrence within a trial run, and consolidation / simplification of its identity. The "numbering scheme" for each thread is contained in table 3-1.

**Table 3-1. Numbering Scheme for All Thread Combinations**

| Thread "Number" | Thread |
|---|---|
| 1 | USA MVR to USMC HIMARS |
| 2 | USMC MVR to USA NLOS-LS |
| 3 | USAF UAS to USA NLOS-LS |
| 5A | USAF UAS to USAF Weaponized UAS |
| 5B | USA FSE to USAF Weaponized UAS |
| 6A | USAF JTAC to USAF Fixed Wing |
| 6B | USAF JTAC to USN Fixed Wing |
| 6C | USAF JTAC to USMC Fixed Wing |
| 7A | USMC JTAC to USAF Fixed Wing |
| 7B | USMC JTAC to USN Fixed Wing |
| 7C | USMC JTAC to USMC Fixed Wing |

## 3.2 Joint Mission Effectiveness Evaluation

The joint capability may be evaluated for performance at the mission and task level. The evaluation of mission performance is focused on SoS impacts to joint mission effectiveness. SoS configuration test factors will be considered in the process of determining mission effectiveness. The evaluation of task performance is focused at the SoS ability to accomplish required tasks. SoS characteristics are considered as well as threat and environmental conditions in the evaluation process.

### 3.2.1 Evaluation of Mission Performance

The evaluation of mission performance is focused on the impact of the SoS in joint mission effectiveness. The evaluation will try to answer the question, "Which SoS configuration provides the best response to improvement in joint mission effectiveness?" MMOEs will provide significant data on evaluating joint mission effectiveness, but the evaluation must consider how SoS factors also impact task accomplishment and SoS attributes.

Response surface methodology (RSM) explores the relationships between several explanatory variables and one or more response variables. Box and Wilson[4] first introduced the method in the early 1950's. The main idea of RSM is to use carefully designed empirical data sets to obtain an optimal response. RSM is a useful alternative to classical optimization and mathematical programming techniques in the presence of experimental design data and the need for an econometric model describing economic systems. Box and Wilson (1951) showed a mathematical framework for an economically-constrained RSM model. Furthermore, they concluded that constrained RSM optimization of the implicit response function had been completely ignored by researchers in general, and economists in particular. In the last five decades RSM methodology has evolved to cover a large number and a wide range of applications, especially as computational capabilities have advanced.

[4] Box, G.E. and P.K.B. Wilson. On the experimental attainment of optimum conditions, 1951. Journal of the Royal Statistical Society 13(1), 1-45.

The leverage plots in figure 3-2 are an example of some of the analysis that can be done on the test data. The plots below show how adding the Forward Observer's (FO) ability to pass information (a task) influences the number of threat casualties (another task). A leverage plot allows you to see which effect is significant and which task performance affects the overall mission accomplishment. Knowing which factors contribute to the overall mission will support the analysis of the SoS and whether or not the newer system allows the related tasks (detect, decide, deconflict, deliver and assess) to be more efficiently/effectively completed.

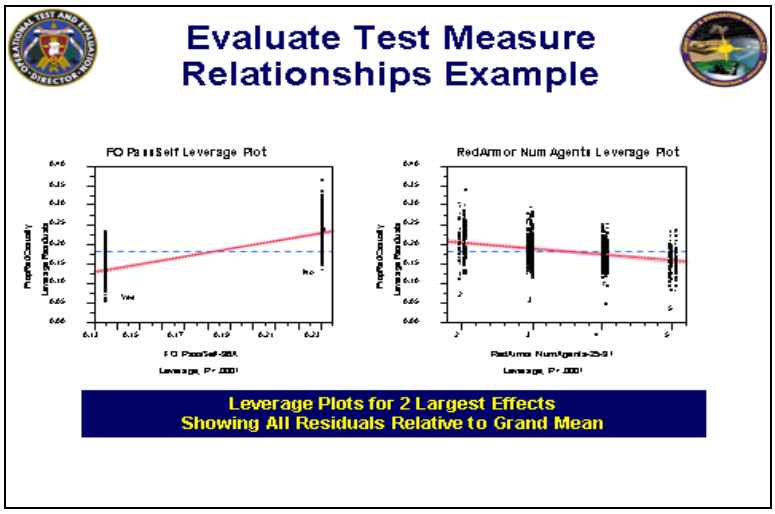



**Figure 3-2. Example Leverage Plots**

System performance to mission outcome will be assessed using operations research and innovative M&S techniques addressing the complex nature of this analysis. These techniques will model and analyze interactions and correlations between multiple levels of performance measures (shown in the diagram above measures section) to produce test results that measure the JAGS contribution to Battlespace Dynamic Deconfliction mission outcome. Classification and regression trees (CART), Pareto Plots, contour and interaction plots, response surface methodologies (RSM), and decision analysis techniques for examining objective response factor as well as other analysis tools will be utilized, as applicable, to determine which factors had the most significant impact on joint mission effectiveness. An example of CART completed after doing some basic modeling and simulation with a similar scenario to THE TEST EVENT is shown in figure 3-3. This CART analysis shows the ability of the Forward Observer to pass information to other squad members had more impact on the number of threat casualties than the number of threat forces in the scenario. Similar analysis will be completed for the test event. This analysis will help to produce the joint mission effectiveness evaluation of the JAGS capability being tested.

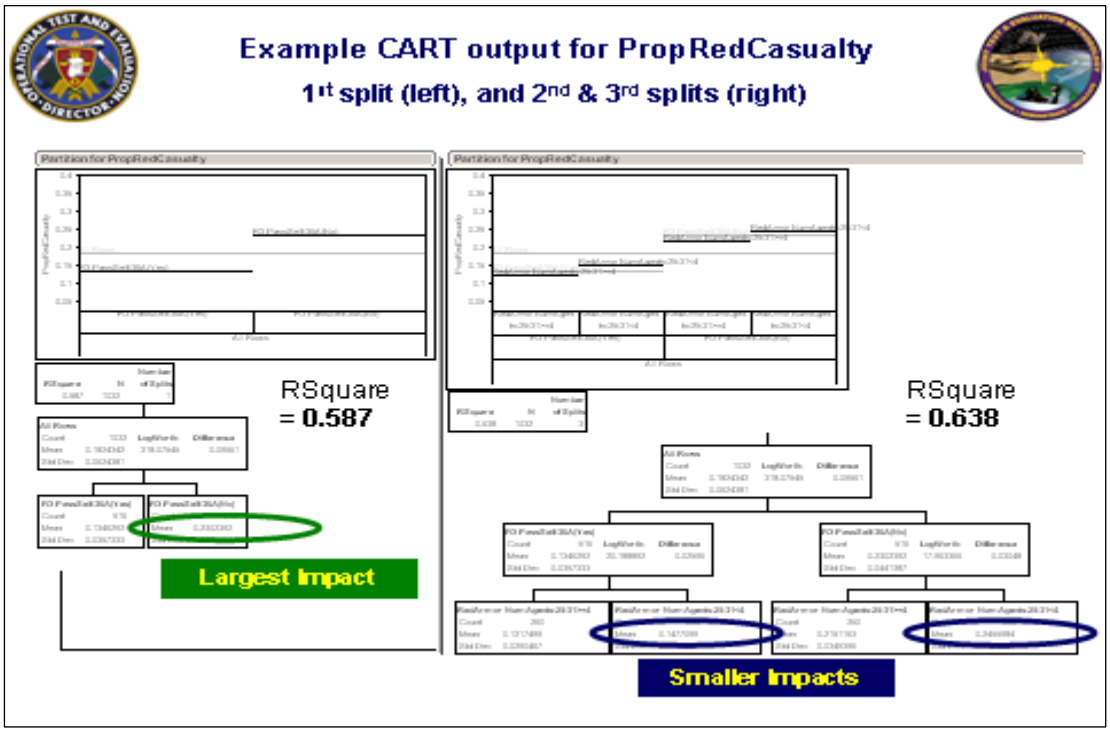



**Figure 3-3. CART Example**

Once the test data has been analyzed, the joint mission evaluator will use test results to evaluate the overall JMe and the contribution an individual system made to the accomplishment of the joint mission. Some processes included in this step are: integration of exploratory analysis results, synthesis of SoS, task, and mission effectiveness evaluations, identification of significant findings, and making recommendations. Using statistical results which have analyzed the SoS across test trials and types of measures, conclusions are reached as to the system/SoS contributions to overall mission performance in the joint mission environment. Due to the complex nature of SoS interactions while testing in a joint environment parameter space, analysts need to explicitly state conclusions in terms of assumptions made throughout the test event. These assumptions should be related to test treatment factor variations and controls during the test. The SoS and JMe evaluation will incorporate the results of the analysis and apply additional reasoning (possibly subject matter expert input for weighting) to arrive at conclusions and recommendations of JMe.[5]

From all of the analysis conducted, findings having significant implications concerning the JAGS SoS, or individual systems will be captured and organized. These findings will be the foundation for the recommendations made in the JCE. These findings will be captured and put into context given the assumptions made throughout the test preparation and execution.

[5] https://learn.dau.mil/html/scorm12/course/course TST 101: Evaluation and Analysis

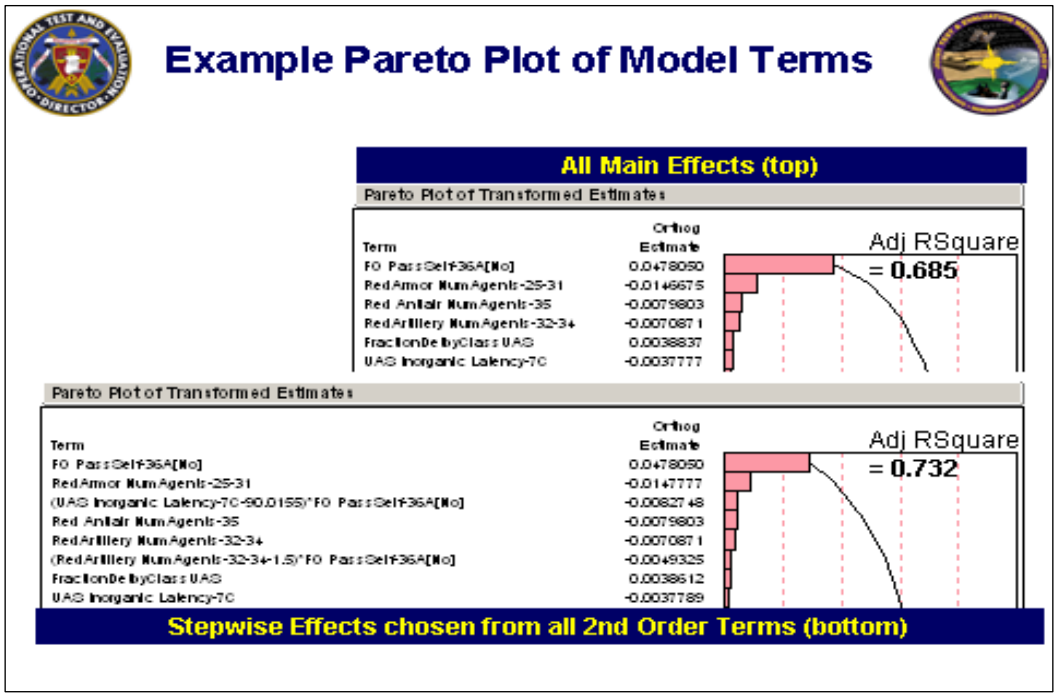



**Figure 3-4. Example Pareto Plot**

The Pareto Plot example in figure 3-4 shows some analysis on the effects of independent variables in order of their significance from top to bottom, so the first factor should be the most influential. Thus using the Pareto Plot for the test event factors will allow for analysis of prioritized factor importance such as FO Situational Awareness was one of the most important factors and the number of Red Armor was not as important, but did have influence. This would be a significant finding. In addition to reporting this finding, a comparison will be made to see if the findings are consistent with subject matter expertise (i.e., we want to ensure the actor performance was consistent with experience of SMEs) before making major recommendations.

Once SoS, task, and JMe evaluations have been developed, joint capability evaluation findings will be synthesized into a summary performance evaluation and the joint capability evaluation (JCE) will be drafted. The JCE will contain both test findings and recommendations. It will indicate impacts to test products and recommendations for possible impacts on the war fighter community. If possible, these findings will be prioritized based on factors such as contribution to mission desired effects, fulfillment of test goals and considerations for risk.

Specific analysis of individual measures will provide the basis for the overall JMe evaluation. These are discussed as follows:

#### 3.2.1.1 MMOE 1

This measure calls for determining the threat systems ineffectiveness at the end of each test run. This would be a summary value indicating the count of threat systems rendered

ineffective during the test. Since the total number of threat systems starting in the JOA is assumed to remain the same across test runs and the length of time for each test run is assumed to be the same, it would be simple to just use the count of ineffective threat systems $X_i$ for each run, *i,* as the measure. However, since a threshold value would typically be a percentage of threat systems ineffective, it is easy to calculate the percentage for each test run as:

$$X_i / X_t$$

where $X_i$ is the count of ineffective threat systems and $X_t$ is the total number of threat systems.

Table 3-2 provides an example of data needed for MMOE-1. Given the test event is planned for four runs of each treatment (trial), a total of sixteen test runs will be conducted for a set period of time. The table shows an example data set of 20 threat systems and the time that each was rendered ineffective. If a value is 0, then the system was never rendered ineffective.

**Table 3-2. Sample data on Threat Systems Ineffectiveness over Time**

| | Trial/Run # | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threat System | T1R1 | T1R2 | T1R3 | T1R4 | T2R1 | T2R2 | T2R3 | T2R4 | T3R1 | T3R2 | T3R3 | T3R4 | T4R1 | T4R2 | T4R3 | T4R4 |
| 1 | 174 | 0 | 130 | 8 | 51 | 0 | 10 | 46 | 0 | 136 | 141 | 59 | 14 | 76 | 133 | 0 |
| 2 | 51 | 81 | 154 | 125 | 163 | 27 | 148 | 103 | 0 | 112 | 88 | 45 | 56 | 47 | 148 | 158 |
| 3 | 142 | 17 | 73 | 0 | 0 | 0 | 103 | 154 | 50 | 0 | 34 | 0 | 107 | 100 | 82 | 44 |
| 4 | 0 | 126 | 0 | 122 | 157 | 0 | 0 | 0 | 0 | 83 | 0 | 167 | 34 | 58 | 93 | 162 |
| 5 | 33 | 92 | 0 | 0 | 0 | 38 | 0 | 156 | 35 | 99 | 0 | 156 | 118 | 152 | 61 | 0 |
| 6 | 76 | 126 | 117 | 82 | 10 | 86 | 0 | 36 | 113 | 147 | 0 | 159 | 139 | 78 | 128 | 75 |
| 7 | 0 | 86 | 129 | 133 | 0 | 120 | 26 | 102 | 129 | 87 | 131 | 150 | 0 | 13 | 100 | 109 |
| 8 | 7 | 14 | 39 | 142 | 138 | 48 | 0 | 16 | 102 | 92 | 120 | 42 | 178 | 41 | 0 | 37 |
| 9 | 0 | 164 | 75 | 76 | 0 | 0 | 128 | 84 | 112 | 106 | 99 | 0 | 53 | 106 | 161 | 0 |
| 10 | 96 | 63 | 147 | 99 | 0 | 89 | 0 | 0 | 28 | 1 | 80 | 52 | 125 | 0 | 128 | 148 |
| 11 | 0 | 0 | 113 | 56 | 59 | 0 | 79 | 9 | 171 | 0 | 105 | 168 | 96 | 55 | 121 | 22 |
| 12 | 157 | 88 | 38 | 38 | 112 | 162 | 162 | 1 | 169 | 15 | 177 | 25 | 67 | 0 | 0 | 0 |
| 13 | 0 | 52 | 167 | 135 | 67 | 56 | 45 | 39 | 0 | 113 | 46 | 92 | 0 | 0 | 82 | 138 |
| 14 | 110 | 159 | 60 | 0 | 38 | 64 | 45 | 145 | 133 | 109 | 36 | 10 | 133 | 110 | 16 | 34 |
| 15 | 65 | 133 | 159 | 50 | 76 | 0 | 160 | 145 | 29 | 78 | 0 | 0 | 149 | 123 | 34 | 0 |
| 16 | 59 | 58 | 0 | 40 | 124 | 100 | 131 | 0 | 9 | 0 | 10 | 90 | 30 | 6 | 0 | 5 |
| 17 | 42 | 13 | 137 | 0 | 7 | 13 | 0 | 100 | 0 | 0 | 57 | 0 | 179 | 0 | 0 | 38 |
| 18 | 94 | 143 | 142 | 104 | 176 | 0 | 0 | 160 | 17 | 68 | 79 | 0 | 0 | 65 | 70 | 8 |
| 19 | 159 | 117 | 16 | 115 | 63 | 159 | 0 | 0 | 0 | 42 | 0 | 171 | 176 | 129 | 90 | 0 |
| 20 | 6 | 140 | 83 | 67 | 0 | 97 | 34 | 117 | 0 | 79 | 36 | 52 | 94 | 129 | 0 | 154 |

A summary of the results can be seen in Table 3-3 as the count of threat units ineffective at the end of each test run. Based on the sample data below, analysis using the Kruskal-Wallis H test concludes to reject $H_o$, indicating at least two populations differ in location.

**Table 3-3. Sample Count of Threat Units Ineffective**

| Count Threat Units Ineffective | | | | | |
|---|---|---|---|---|---|
| | Percentage | | | | |
| **Trial #** | **Run 1** | **Run 2** | **Run 3** | **Run 4** | **Avg** |
| **1** | 16 | 15 | 14 | 17 | 15.5 |
| **2** | 15 | 15 | 13 | 16 | 14.8 |
| **3** | 17 | 17 | 13 | 16 | 15.8 |
| **4** | 17 | 12 | 17 | 14 | 15.0 |

A plot of this can be seen in Figure 3-5, where the sample data indicates trials one and four have the highest rank with the greatest numbers of threat systems ineffective across the four runs. The non-parametric test of the sample data concluded there is a statistical difference in the location of the populations, meaning trials one and four are much higher in threat units ineffective than trials two and three.

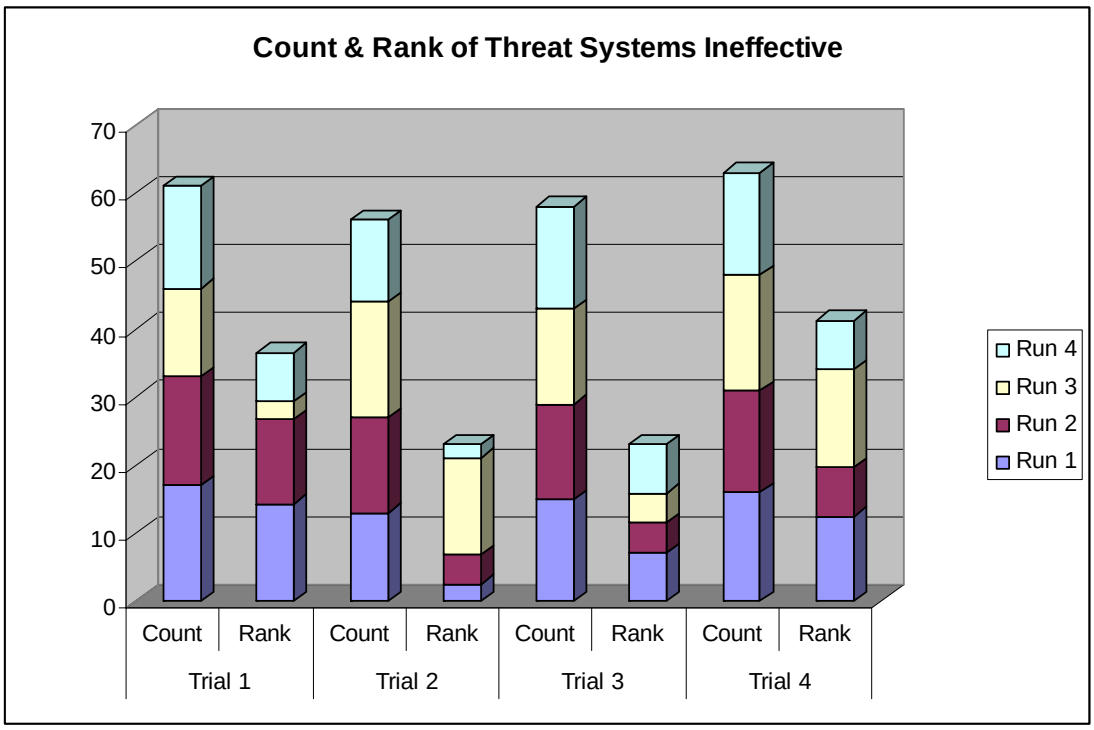



**Figure 3-5. MMOE 1 Data Evaluation**

### 3.2.1.3 TMOPs 1 - 4

TMOPs 1 through 4 measure the performance of the notional joint task, "joint battlespace dynamic deconfliction command and control". The measures are based on a time dimension, which hypothesizes the most effective SoS will be a future SoS configuration able to command and control the THE TEST EVENT process in the shortest time without loss of accuracy. As stated in chapter 2, the material and non-material factors selected to emulate the future SoS will take less time from acceptance of a CFF/ASR to the time it takes the C2 authority to approve or deny it, than those factors emulating the current SoS over the same process. For each instantiation *i* of a CFF/ASR in a test run, data will be collected for the start and stop time of the THE TEST EVENT C2 process. Data will be collected for each test run and separated across the four test trials to form population sample for each SoS configuration. Table 3-5 shows the format for some example data.

**Table 3-5. Sample Indirect Fires Data for One Run**

| | Deconfliction # | Type JCAS: 0 JFIRES: 1 | Task Thread 1 - 7 | C2 Start Time (min) | C2 End Time (min) | C2 Time (min) |
|---|---|---|---|---|---|---|
| TRIAL 1 - Run 1 | 1 | 1 | 8 | 4 | 10 | 6 |
| | 2 | 0 | 6 | 15 | 35 | 20 |
| | 3 | 0 | 3 | 27 | 42 | 15 |
| | 4 | 1 | 3 | 34 | 43 | 9 |
| | 5 | 0 | 2 | 49 | 56 | 7 |
| | 6 | 0 | 6 | 68 | 82 | 14 |
| | 7 | 1 | 8 | 79 | 95 | 16 |
| | 8 | 0 | 3 | 91 | 110 | 19 |
| | 9 | 1 | 8 | 115 | 135 | 20 |
| | 10 | 0 | 2 | 126 | 138 | 12 |
| | 11 | 1 | 5 | 138 | 147 | 9 |
| | 12 | 1 | 7 | 143 | 154 | 11 |
| | 13 | 0 | 2 | 157 | 166 | 9 |
| | 14 | 1 | 5 | 159 | 172 | 13 |
| | 15 | 0 | 2 | 164 | 175 | 11 |
| | 16 | 0 | 5 | 166 | 179 | 13 |

Given a sufficient number of indirect fires requiring de-confliction are processed in each of the test runs, the test run data can be consolidated across each trial to provide data sets sufficient in size to assume normality based on the central limit theory. The data can be summarized for each test trial as shown in table 3-6.

**Table 3-6. Trial Summary Data**

| Summary Analysis | | | | |
|---|---|---|---|---|
| **Trial** | **Count of THE TEST EVENT** | **Avg C2 Time** | **S.D. C2 Time** | **Skewness** |
| **1** | 64 | 11.25 | 4.567 | 0.2624 |
| **2** | 64 | 13.438 | 4.8399 | 0.3295 |
| **3** | 64 | 13.234 | 4.8032 | 0.2041 |
| **4** | 64 | 13.391 | 5.132 | 0.0869 |

A factorial Analysis of Variance (ANOVA) can be used to study the effects of the two factors (treatments) on the dependent variable. In this case, it is a 2x2 factorial ANOVA design, where there are two independent variables and each variable has two levels or distinct values. Table 3-7 shows the ANOVA results for the sample data. In this case, the ANOVA resulted in rejecting $H_o$ (the averages are different across the two factors (four test trials)). It is apparent trial 1 had a lower average time when compared to the other three test trials. To statistically show a difference requires use of the Tukey post-hoc test; a set of pairwise comparisons using the formula:

$$t_s = \frac{M_i - M_j}{\sqrt{\frac{MSE}{n_h}}}$$

where: $M_i$ = Mean of trial i
$M_j$ = Mean of trial j
MSE = Mean Square Error
$n_h$ = the harmonic mean of trials i and j

**Table 3-7. Sample ANOVA Calculations**

| | | | | | | |
|---|---|---|---|---|---|---|
| **ANOVA** | | | | | | |
| **CM =** | | | 42127.56 | | | |
| **Total SS =** | | | 6116.44 | | | |
| **SST =** | | | 213.97 | | | |
| **SSE =** | | | 5902.47 | | | |
| **MST =** | | | 71.32 | | | |
| **MSE =** | | | 23.42 | | | |
| **F =** | | | 3.05 | | | |
| | | | | | | |
| **v1=** | 3 | **v2 =** | 252 | | | |
| **F Inv =** | | | 2.64 | | | |
| **Ho = Avgs the Same** | | | | Reject Ho? | | Yes |
| **Ha = Avgs are Different** | | | | | | |

A statistical analysis of the sample data provides the results shown in table 3-8. If the value of the pairwise comparison is greater than the value Q (two-tailed t-distribution at $\alpha = 0.05$), then you can reject the null hypothesis and conclude the paired means are statistically different. In this sample set of data, we must conclude the differences are not statistically different.

**Table 3-8. Tukey Results**

| **Tukey HSD** | | **Q= 3.63** |
|---|---|---|
| **Trial #** | **3** | **4** |
| **1** | 3.280 | 3.538 |
| **2** | 0.336 | 0.077 |

### 3.2.1.8 MOSA 1

Measure of System of Systems Attribute (MOSA) One measures the speed of decision in the THE TEST EVENT process. Since the SoS configuration will impact the decision process, it is important to evaluate MOSA 1 to ensure adequate attribute performance. Evaluate:

- **Mission Level Eval: Across trials – evaluates different SoS factors**

This measure will determine if there is a statistical difference between how long it takes from when a Indirect Fires request enters the C2 queue, $T_{si}$, to the time it takes the C2 authority to start the deconfliction process, $T_{ei}$ ; mathematically $T_{MOSA1i} = T_{ei} - T_{si}$ . The hypothesis for this comparison is:

$\underline{H_a}$: $\mu_x \neq \mu_y$ : The mean time it takes from acceptance of Indirect Fires requests to the time it takes the C2 authority to start the deconfliction process for SoS configuration *x*, will be statistically different than the mean time it takes from acceptance-to-deconfliction of Indirect Fires requests for SoS configuration *y*.

$\underline{H_o}$: $\mu_x = \mu_y$ : The mean times are statistically equivalent.

Table 3-12 contains the summary data for a notional data set expected to be collected across four trials.

**Table 3-12 – Notional Summary Data for MOSA-1**

| Summary Analysis | | | |
|---|---|---|---|
| **Trial** | **Count of JBD2** | **Avg Decision Time** | **S.D. Decision Time** |
| **1** | 64 | 8.625 | 2.763 |
| **2** | 64 | 9.297 | 2.665 |
| **3** | 64 | 9.063 | 2.788 |
| **4** | 64 | 8.375 | 2.616 |

A factorial Analysis of Variance (ANOVA) can be used to study the effects of the two factors (treatments) on the dependent variable. In this case it is a 2x2 factorial ANOVA design, where there are two independent variables and each variable has two levels or distinct values. Table 3-13 shows the ANOVA results for the sample data. In this case the ANOVA failed to reject $H_o$ (the averages cannot be proved to be statistically different across the two factors (four test trials)). If the test had rejected $H_o$, then the use of the Tukey post-hoc test would have been necessary to conduct pairwise comparisons for the SoS configurations.

**Table 3-13. Sample ANOVA Calculations**

| **ANOVA** | | | | | | |
|---|---|---|---|---|---|---|
| **CM =** | | | 20004.57 | | | |
| **Total SS =** | | | 1882.43 | | | |
| **SST =** | | | 33.32 | | | |
| **SSE =** | | | 1849.11 | | | |
| **MST =** | | | 11.11 | | | |
| **MSE =** | | | 7.34 | | | |
| **F =** | | | 1.51 | | | |
| | | | | | | |
| **v1=** | 3 | **v2 =** | 252 | | | |
| **F Inv =** | | | 2.64 | | | |
| **Ho = Avgs the Same** | | | | Reject Ho? | | No |
| **Ha = Avgs are Different** | | | | | | |

In this sample set of data, we must conclude the differences are statistically the same.

### 3.2.2 Evaluation of Task Performance

The evaluation of task performance is focused on the impact of the SoS in performing the joint tasks. The evaluation will try to answer the question, "Did the SoS perform the task at the specified standards?" TMOPs will provide the data necessary to respond to this evaluation.

### 3.2.2.1 TMOP 1

The time to C2 THE TEST EVENT indirect fires:

- **Task Level Eval: Across task types – JFIRES, JCAS**

The measure calls for determining how long it takes from the acceptance of a CFF / ASR, $T_{si}$, to the time it takes the C2 authority to approve or deny it, $T_{ei}$. The requirement levied upon this task is all of the times collected ($T_{ei}$ - $T_{si}$), will be less than or equal to the Desired Blue Time, $T_D$ (mathematically, $T_D \geq T_{ei} - T_{si}$). Additionally, there is a desire to determine if there is a difference in the average time it takes to C2 the JFIRES tasks against the average time it takes to C2 the JCAS tasks. The hypothesis for this comparison is:

- $\underline{H_a}$: $\mu_{CAS} \neq \mu_{FIRES}$ : The mean time it takes from acceptance of a CFF (a JFIRES task) to the time it takes the C2 authority to approve or deny it will be statistically different than the mean time it takes from acceptance-to-decision of an ASR (a JCAS task; where $\mu_{CAS}$ represents the average time it takes to C2 a JCAS task and $\mu_{FIRES}$ represents the average time it takes to C2 a JFIRES task).
- $\underline{H_o}$: $\mu_{CAS} = \mu_{FIRES}$ : The mean times are statistically equivalent.

Table 3-15 contains a brief example of a notional data set expected to be collected during one trial.

**Table 3-15. Notional Data Set for TMOP 1**

| Counter, i | Start Time, $T_{si}$ | End Time, $T_{ei}$ | $T_{ei}$ - $T_{si}$ | Request Type |
|---|---|---|---|---|
| 1 | 03:18 | 08:42 | 05:24 | CFF |
| 2 | 05:26 | 09:37 | 04:11 | CFF |
| 3 | 07:00 | 12:19 | 05:19 | ASR |
| : | : | : | : | : |
| n | | | | |

Assessing whether each Indirect Fire time is less than or equal to the Desired Blue Time, $T_D$, is straight forward: compare $T_D$ against the values obtained in the fourth column of Table 3-15 ($T_{ei}$ - $T_{si}$). If any of the values in table 3-15 are greater than $T_D$, as much information should be retained to ascertain the cause (if any).

Assessment of the hypothesis involves calculating the sample means and standard deviations for each "Request Type", shown in table 3-16, then using the t-distribution to compare if there is a statistical difference between the means.

**Table 3-16. Sample Means and Standard Deviations by Request Type**

| Request Type | Sample Mean, $\bar{x}$ | Sample Standard Deviation, s |
|---|---|---|
| CFF (JFIRES) | 5:11 | 2:55 |
| ASR (JCAS) | 6:23 | 4:02 |

The equation to compute the test statistic, $t_0$, (representing a two-sample hypothesis test of the mean) is:

$$t_0 = \frac{(\bar{x}_{ASR} - \bar{x}_{CFF})}{S_p\sqrt{1/n_{ASR} + 1/n_{CFF}}}, \text{ where } S_p = \sqrt{\frac{(n_{ASR}-1)s_{ASR}^2 + (n_{CFF}-1)s_{CFF}^2}{n_{ASR}+n_{CFF}-2}}$$

Substituting the notional data into these equations yields a $t_0$ of:

$$t_0 = \frac{(6.383 - 5.183)}{3.233\ \sqrt{1/4 + 1/10}} = 0.627$$

The critical value for this test is 1.782 (a two-tailed test with 12 degrees of freedom and an alpha of 0.1). Since the test statistic, $t_0$, is less than the critical value, we fail to reject the null hypothesis – therefore, the means are statistically equivalent. In the context of TMOP 1, there is no statistical difference between the mean times to C2 the JFIRES and JCAS tasks.

### 3.2.3 Evaluation of SoS Attributes

The evaluation of SoS attributes is focused on the characteristics of the SoS and its impact on performing joint tasks and joint mission effectiveness. The evaluation will try to answer the question, "Did the SoS perform to its specified standards?" Measures of SoS attributes will provide the data necessary to respond to this evaluation.

#### 3.2.3.1 MOSA 1

The speed of decision in the THE TEST EVENT process.

- **Task Level Eval: Across task types – JFIRES, JCAS**

This measure will determine if there is a statistical difference between how long it takes from when a CFF request enters the C2 queue, $T_{si}$, to the time it takes the C2 authority to approve or deny it, $T_{ei}$ (not including the time taken to deconflict the airspace, $T_{di}$), versus how long it takes an ASR to complete the same process (mathematically, $T_{MOSA1i} = T_{ei} - T_{si} - T_{di}$). The hypothesis for this comparison is:

$\underline{H_a}$: $\mu_{CAS} \neq \mu_{FIRES}$ : The mean time it takes from acceptance of a CFF (a JFIRES task) to the time it takes the C2 authority to approve or deny it, minus the deconfliction time, will be statistically different than the mean time it takes from acceptance-to-decision-minus-deconfliction of an ASR (a JCAS task; where $\mu_{CAS}$ represents the average time it takes to C2 a JCAS task and $\mu_{FIRES}$ represents the average time it takes to C2 a JFIRES task).

$\underline{H_o}$: $\mu_{CAS} = \mu_{FIRES}$ : The mean times are statistically equivalent.

Table 3-31 contains a brief example of a notional data set expected to be collected during one trial.

**Table 3-31 – Notional Data Set for MOSA-1**

| Counter, i | Start Time, $T_{si}$ | End Time, $T_{ei}$ | Deconfliction Time, $T_{di}$ | $T_{ei}$ - $T_{si}$ - $T_{di}$ | Request Type |
|---|---|---|---|---|---|
| 1 | 03:18 | 08:42 | 02:00 | 03:24 | CFF |
| 2 | 05:26 | 09:37 | 01:00 | 03:11 | CFF |
| 3 | 07:00 | 12:19 | 01:30 | 03:49 | ASR |
| : | : | : | : | : | : |
| n | | | | | |

Assessment of the hypothesis involves calculating the sample means and standard deviations for each "Request Type", then using the t-distribution to compare if there's a statistical difference between the means.

**Table 3-32 – Notional Analysis for MOSA-1**

| Request Type | Sample Mean, $\bar{x}$ | Sample Standard Deviation, s |
|---|---|---|
| CFF (JFIRES) | 5:11 | 2:55 |
| ASR (JCAS) | 6:23 | 4:02 |

The equation to compute the test statistic, $t_0$, (representing a two-sample hypothesis test of the mean) is:

$$t_0 = \frac{\left(\bar{x}_{ASR} - \bar{x}_{CFF}\right)}{S_p\sqrt{1/n_{ASR} + 1/n_{CFF}}}, \text{ where } S_p = \sqrt{\frac{(n_{ASR}-1)s_{ASR}^2 + (n_{CFF}-1)s_{CFF}^2}{n_{ASR}+n_{CFF}-2}}$$

Substituting the notional data into these equations yields a $t_0$ of:

$$t_0 = \frac{(6.383 - 5.183)}{3.233\ \sqrt{1/4 + 1/10}} = 0.627$$

The critical value for this test is 1.782 (a two-tailed test with 12 degrees of freedom and an alpha of 0.1). Since the test statistic, $t_0$, is less than the critical value, we fail to reject the null hypothesis – therefore, the means are statistically equivalent. In the context of MOSA 1, there is no statistical difference between the mean "Speed of Decision" times to C2 the JFIRES and JCAS tasks.

- **Task Level Eval: Across task threads**

The measure calls for determining how long it takes from the acceptance of a CFF / ASR, $T_{si}$, to the time it takes the C2 authority to approve or deny it, $T_{ei}$. This measure will segregate each request by the task thread it represents (versus the above measure, which segregated the requests by their type (CFF or ASR)). Since there are six threads being represented (four JFIRES and two JCAS), the hypothesis will focus on if there is a difference in the average time it takes to C2 each of them. The hypothesis for this comparison is:

$\underline{H_a}$: $\mu_{T1} \neq \mu_{T2} \neq \mu_{T3} \neq \ldots \mu_{T6}$ : At least one of the mean times it takes from acceptance of a CFF / ASR (identified by the corresponding thread it represents) to the time it takes the C2 authority to approve or deny it, minus the deconfliction time, will be statistically different from one of the other mean times (where, $\mu_{T1}$ represents the average time it takes to C2 the first thread, $\mu_{T2}$ represents the average time it takes to C2 the second thread, etc).

$\underline{H_o}$: $\mu_{T1} = \mu_{T2} = \mu_{T3} = \ldots \mu_{T6}$ : All the mean times are statistically equivalent.

Table 3-33 contains a brief example of a notional data set expected to be collected during one trial. NOTE: Due to potentially small sample sizes, data from "like factors" may have to be pooled. "Like factors" refers to the four runs scheduled to be executed having the factors with the same setting.

**Table 3-33 – Notional Data Set for MOSA-1**

| Counter, i | Start Time, $T_{si}$ | End Time, $T_{ei}$ | $T_{ei} - T_{si} - T_{di}$ | Thread Type |
|---|---|---|---|---|
| 1 | 03:18 | 08:42 | 05:24 | 3 |
| 2 | 05:26 | 09:37 | 04:11 | 4 |
| 3 | 07:00 | 12:19 | 05:19 | 7 |
| : | : | : | : | : |
| n | | | | |

Assessment of the hypothesis involves calculating the sample means and standard deviations for each thread, then using an ANOVA to determine if there's a statistical difference between the means.

**Table 3-34 – Notional Analysis for MOSA-1**

| Thread | Sample Mean, $\bar{y}$ | Sample Standard Deviation, s | Sample Size, n |
|---|---|---|---|
| **1** | 5.11 | 2.55 | 4 |
| **2** | 6.23 | 4.02 | 4 |
| **3** | 5.78 | 1.68 | 4 |
| **5** | 7.45 | 2.79 | 4 |
| **6** | 6.97 | 1.25 | 4 |
| **7** | 4.97 | 2.22 | 4 |

The ANOVA equations necessary to compute the test statistic, $F_0$, are:

$$\text{MSB} = \frac{\sum_{1}^{L} n_j \left(\bar{y}_j - \bar{y}\right)^2}{L-1} = 9.534$$

$$\text{MSE} = \frac{\sum_{1}^{L} \left(n_j - 1\right) s_j^2}{\sum_{1}^{L} \left(n_j - 1\right)} = 8.328$$

$$F_0 = \text{MSB/MSE} = 1.145$$

The critical value for this test, $F_{CRIT}$ is 1.98 (at alpha = 0.1 with 7 and 24 degrees of freedom). Since the test statistic, $F_0$, is less than the critical value, we Fail to Reject the null hypothesis, meaning all of the sample means are statistically equivalent.

## 3.3 Data Management

### 3.3.1 Overview

Management of data collection, validation, and processing is critical to successful data analysis and evaluation. Many challenges exist with a distributed test that must be planned for and mitigated. Data will need to be collected across all test sites for each and every test run. Data must be consistent in type, format, and units of measure.

### 3.3.2 Data Collection Requirements

The specific data requirements and parameters are included in the Integrated Data Requirements List (IDRL) in Annex B. However, the following data management guidelines are necessary to the analysis of data:

- All data needs to be managed and labeled to identify the appropriate test trial and test run from which it was collected.
- The start and stop time of each test run must be recorded to ensure a constant test run time interval.
- Data on each threat system in a test run is required. Specifically, each threat system will need to be identified, along with an effectiveness trait that will be a Boolean parameter of effective or ineffective. Each threat system will start a test run in an effective state. An ineffectiveness time data field must record a time the threat system becomes ineffective.
- Data on each indirect fires request in a test run is required. Each IF will need to be numbered to identify it throughout the THE TEST EVENT process. Most data fields for the IF request are time related and must be recorded as the THE TEST EVENT process is conducted. Additionally, the task thread for which the IF request is a part of must be recorded to identify key components of the THE TEST EVENT process and JCAS/JFIRES tasks.

- Deconfliction disapprovals must be recorded along with the reason for disapproval. A list of possible reasons may be constructed in advance of the test to facilitate data entry.
- The time dimension may be provided in minutes and seconds measured from the start of the test trial, however it is desirable to have the time data be converted to minutes with 2 decimal points (i.e., convert seconds to $100^{th}$ of a minute).

### 3.3.3 Data Validation Requirements

The validation must be conducted to ensure completeness and reliability of the data. Some specific validation tests are:

- A review of data to ensure it is complete for all the threat systems (i.e., if 50 threat systems in a test run, then 50 records should be provided).
- An ineffectiveness time should be recorded for every threat system rendered ineffective. There should also not be a time recorded if the thread system was not rendered ineffective.
- All should also be recorded with the test trial number and test run number.

### 3.3.4 Data Processing

Quick look analysis will be required on a daily basis. To accommodate this requirement, data must be submitted daily, processed, validated, and queried to conduct some initial analysis and make some preliminary insights. Resources must be in place during the test event to ensure this capability.

*This page intentionally left blank*

# CHAPTER 4 – REPORTING REQUIREMENTS

## 4.0 Overview

Programs designated for OSD T&E oversight are required by DOD Instruction 5000.2 to provide formal, detailed, reports of results, conclusions, and recommendations. Although THE TEST EVENT does not fall under OSD T&E oversight, it will complete a final report IAW the appropriate reporting requirements. This report will: summarize the purpose of the test as well as an overview of the methodology followed; identify test and evaluation limitations, constraints, as well as assumptions that were made; include the effect of these limitations on SoS performance and system performance. Most importantly the final report will include significant findings, conclusions, and recommendations. There will be several reports that need to be completed which are outlined below:

## 4.1 Reports

### 4.1.1 Test Incident and Corrective Action Report

If an incident occurs affecting the analysis of the data or has a major impact in delaying or stopping the test event, a Test Incident and Corrective Action Report (TIR) will be completed. This report will describe the incident, how it occurred and any corrective actions taken as well as their status. At the end of the test event all TIRs will be reviewed for possible corrective action and will be incorporated into the final report as applicable.

### 4.1.2 Joint Capability Evaluation

Once SoS, task, and JMe evaluations have been developed, joint capability evaluation findings will be synthesized into a summary performance evaluation and the joint capability evaluation (JCE) will be drafted. The test data will be analyzed and the evaluators will use test results to determine the overall JMe and the contribution an individual system/SoS makes to the accomplishment of the joint mission. Specifically, the evaluators will be focusing on answering the CJIs developed for the test (CJIs are detailed in chapter 2 of this plan). The analysts will need to explicitly state conclusions and recommendations in terms of the CJIs developed. The JCE outline is shown as annex F of this plan. The JCE will contain test findings and recommendations as well as indicate impacts to test products and possible implications to the war fighter community. The JCE findings will be prioritized based on factors such as contribution to mission desired effects, fulfillment of test goals and considerations for risk. The JCE will include an audit trail of deviations from the planned testing event and their impact on the evaluation, some of which may be captured in the test incident and corrective action reports. Finally, the JCE report will include details such as test conditions and test results to include, detailed displays of data from the tests, and testers' observations.

### 4.1.3 System Evaluation Reports

Individual System Evaluation Report (SER) documents may be completed on a case by case basis for this test event. These document independent evaluation findings and recommendations of system operational effectiveness, suitability, and survivability. The SER for individual systems should address the critical operational issues and additional evaluation focus areas based on all available credible data and the evaluator's analytic treatment of the data. If available and as appropriate the

individual system evaluation reports will be incorporated as annexes to the capability evaluation report.

## 4.2 Reporting Timelines

The final report (Joint Capability Evaluation) for this test event is due 90 days after the completion of the event.

## 4.3 Capability Evaluation

The capability evaluation is composed of three different levels of measures and their relationships.

- SoS Performance Relations to Task / JMe: This evaluation gives an assessment of the systems/SoS capabilities to perform command and control as required. For this test event these assessments will primarily be focused on the MOSAs. The MOSAs focus on SoS characteristics such as speed of command and force agility. These characteristics of JAGS as well as the systems incorporated into JAGS may provide the justification, in operational terms, to assist decision-makers in their determination of future acquisition and implementation of the systems under test/JAGS SoS.
- Task Performance Relations to SoS / JMe: Evaluation at the mission level will not assess the effectiveness of the JAGS to perform as required, but rather it will assess effectiveness of mission-level outcomes concerning capability areas the system is being employed. Mission level assessment includes conclusions and recommendations concerning effectiveness of the JAGS task performance to achieve desired mission outcomes. For example the JAGS performance on TMOPs will most likely have an impact on the mission accomplishment.
- Evaluate JMe Relations to Task / SoS: Evaluations are performed concerning system level contributions to mission level outcomes, which provide decision support information concerning the overall JMe of the system. This evaluation includes looking at the second-order interactions and effects of the JAGS on the battlespace. Upon completion of the evaluation of system performance to mission outcome, the JAGS impact (positive or negative) on successful joint mission execution can be determined.

### 4.3.1 Goal-Objectives Resolution

#### 4.3.1.1 Test Goals

The test goals clarify and refine the system capabilities of the system from the perspective of joint testing. Available T&E inputs were used from the DOD acquisition system as a base to narrow the test to a specific set of joint system capabilities. The goal of this test is to evaluate the JAGS solution to the capability gap, in the execution of Battlespace Dynamic Deconfliction for the conduct of CAS and JFIRES. This SoS C2 capability will analyze two variables. The first variable (factor) will be the battlespace management capability. The second variable (factor) will be the C2 processes utilized by the forces.

The test and evaluation strategy was developed to define the overall test approach for the C2 SoS. It will ensure all necessary data points are collected in the proper format so a thorough analysis of the capability can be conducted. This particular test will not assess all of the key performance characteristics but instead will focus on the procedural processes and battle management. The evaluation strategy and data analysis portion of this test event were described in detail in chapter three of this plan. The test data from the test event will be analyzed to determine the overall JMe and the contribution an individual system/SoS makes to the accomplishment of the joint mission (i.e., how the JAGS system will fulfill C2). The TMOPs, MMOEs, and MOSAs will be analyzed

in respect to mission accomplishment. A method that will be explored is making Task Measure of Performance factors and then see how they contribute to the mission effectiveness. This method would tie task performance to joint mission effectiveness. Other methods such as weighting the different factors and conducting sensitivity analysis on the factors will also be included. This analysis should answer the test goal and answer the question of whether or not the JAGS system will fulfill a C2 capability gap.

#### 4.3.1.2 Proposed Evaluation Test Objectives

The test has two evaluation objectives:

1. Determine which SoS configuration provides the best response to improvement in joint mission effectiveness.
2. Determine if the SoS performed the task at the desired standards.

The first objective is explained in detail in chapter 3.2. The objective will be achieved primarily through the use of the MMOEs, however TMOPs and MOSAs will be considered. The evaluation of objective one will focus at the mission level and will determine which configuration of the JAGS SoS has the greatest influence on joint mission effectiveness. The different configurations of the JAGS for this test event are outlined in section 1.4.2 of this test plan. The second objective is to determine if the JAGS can perform the task at the designated levels or better. This objective focuses on task performance; therefore the TMOPs and measures of SoS attributes data will be utilized when determining whether or not objective two was met.

### 4.3.2 Issues Resolution

The Test Event has three test issues that will be addressed during the test event. Defined and described in chapter two of this plan, these test issues are:

- Assess the ability to perform immediate battlespace deconfliction C2 across a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA.
- Assess the ability to perform joint close air support across (JCAS) a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA
- Assess the ability to perform joint fires across a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA

An issue mapping to MMOEs, TMOPs, and MOSAs is shown in table 4-1. This shows how the measures, which were defined in chapter three of this plan, map to the relevant issues of the test. Thus, if the analysis plan is followed it should be relative easy to answer the issues outlined above.

# Table 4-1. Issue Mapping

| Issue | MMOE/TMOP/MOSA | Summary Description |
|---|---|---|
| 1. Assess the ability to perform immediate battlespace deconfliction C2 across a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA. | MMOE 1: Threat Systems Combat Ineffectiveness | Threat systems combat ineffectiveness measures a count of threat systems that become ineffective during a test run. Quantifying threat systems ineffectiveness leads to a mission desired effect of threat forces destroyed or neutralized in the JOA that result in a mission end state of expanded lodgment and control of key infrastructure in the conduct of a joint forcible entry mission. |
| | MMOE 2: Threat Systems Combat Ineffectiveness Rate | Threat systems combat ineffectiveness rate measures the number of threat systems ineffective over time. Quantifying threat systems ineffectiveness rate leads to a mission desired effect of threat forces destroyed or neutralized in the JOA resulting in a mission end state of expanded lodgment and control of key infrastructure in the conduct of a joint forcible entry mission |
| | MOSA 1: Speed of Decision | This measure will ascertain how long it takes from when a call-for-fire (CFF) request / air support request (ASR) enters the C2 queue to the time the C2 authority has evaluated it (does not including the time taken to deconflict the airspace). |
| | MOSA 2: Speed of Deconfliction | This measure will ascertain how long it takes to conduct the airspace deconfliction task within the C2 process. |
| | MOSA 3: Force Simultaneity | This measure will determine how many CFFs/ASRs are being executed within the C2 node at the same time. |
| 2. Assess the ability to perform joint close air support across (JCAS) a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA | TMOP 1: C2 JBD2 Indirect Fire (JCAS & JFIRES) Tasks | This measure will ascertain how long it takes from the issuance of a call-for-fire (CFF) request / air support request (ASR) to the time it takes the C2 authority to approve or deny it. |
| | TMOP 2: C2 JBD2 Indirect Fire (JCAS & JFIRES) by Task Thread | This measure will ascertain how long it takes for call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service and platform initiating it, and the Service and weapon system assigned as its response. |
| | TMOP 3: Deconfliction of Indirect Fire (JCAS & JFIRES) Missions when Airspace is Controlled by Requesting Service | This measure will ascertain how long it takes for a call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service initiating it, the Service in control of the desired airspace when initiated, and the Service assigned as its response. |
| | TMOP 4: Deconfliction of Indirect Fire (JCAS & JFIRES) Missions when Airspace is Controlled by a Service Other than the Requesting Service | This measure will ascertain how long it takes for a call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service initiating it, the Service in control of the desired airspace when initiated, and the Service assigned as its response. |
| | TMOP 5: Number of Indirect Fire (JCAS & JFIRES) Missions Denied (with Reason) | This measure will track the number of call-for-fire (CFF) requests / air support requests (ASR) for Indirect Fires denied by the C2 authority along with the rationale as to why it was denied. |
| | TMOP 6: Time to get Ordnance on Target after Initiating Air to Surface Attack Request (JCAS) | This measure will ascertain how long it takes from the acceptance of an air support request (ASR) to the time it takes an approved JCAS mission to deliver its ordnance on target. |
| | TMOP 7: Percent of Targets Attacked IAW JCAS | This measure will determine the effectiveness of JAGS and its ability to perform JCAS missions by calculating the percentage of targets attacked IAW JCAS. |
| 3. Assess the ability to perform joint fires across a full range of military operations by JAGS in order to destroy or neutralize threat forces in the JOA | TMOP 1: C2 JBD2 Indirect Fire (JCAS & JFIRES) Tasks | This measure will ascertain how long it takes from the issuance of a call-for-fire (CFF) request / air support request (ASR) to the time it takes the C2 authority to approve or deny it. |
| | TMOP 2: C2 JBD2 Indirect Fire (JCAS & JFIRES) by Task Thread | This measure will ascertain how long it takes for call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service and platform initiating it, and the Service and weapon system assigned as its response. |
| | TMOP 3: Deconfliction of Indirect Fire (JCAS & JFIRES) Missions when Airspace is Controlled by Requesting Service | This measure will ascertain how long it takes for a call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service initiating it, the Service in control of the desired airspace when initiated, and the Service assigned as its response. |
| | TMOP 4: Deconfliction of Indirect Fire (JCAS & JFIRES) Missions when Airspace is Controlled by a Service Other than the Requesting Service | This measure will ascertain how long it takes for a call-for-fire (CFF) request / air support request (ASR) to be approved or denied based upon the Service initiating it, the Service in control of the desired airspace when initiated, and the Service assigned as its response. |
| | TMOP 5: Number of Indirect Fire (JCAS & JFIRES) Missions Denied (with Reason) | This measure will track the number of call-for-fire (CFF) requests / air support requests (ASR) for Indirect Fires denied by the C2 authority along with the rationale as to why it was denied. |
| | TMOP 8: Percent of Targets Attacked IAW Requests for Fires (JFIRES) | This measure will determine the effectiveness of JAGS and its ability to perform JFIRES missions by calculating the percentage of targets attacked IAW JFIRES. |
| | TMOP 9: Time to get Ordnance on Target after Initiating Task (JFIRES) | This measure will ascertain how long it takes from acceptance of a call-for-fire (CFF) request to the time it takes an approved JFIRES mission to deliver its ordnance on target |

# Annex A: Integrated Data Requirements List (IDRL)

| New Number | CJI/COI | Old Number | Scale Units | Measure | UJTL | Standard | Data Element | Data Fields |
|---|---|---|---|---|---|---|---|---|
| **MMOE-1** | | MMOE-1 | Percent | Threat Systems Combat Ineffectiveness | | TBD | | |
| | | | Number | | | | Count threat units in JOA at test run start time | For each test run: (Test Trial #, Test run #, Threat Unit Count at Start) |
| | | | Number | | | | Count threat units ineffective in JOA at test run end time | For each test run: (Test Trial #, Test run #, Count Threat Units Ineffective at end time) |
| **MMOE-2** | | MMOE-3 | Number | Threat Systems Combat Ineffectiveness Rate | | TBD | | |
| | | | Number | | | | Count threat units in JOA at test run start time | For each test run: (Test Trial #, Test run #, Threat Unit Count at Start) |
| | | | Time | | | | Start time for each test run | For each test run: (Test Trial #, Test run #, Start time) |
| | | | Time | | | | End time for each test run | For each test run: (Test Trial #, Test run #, End time) |
| | | | Time | | | | Time of each threat system became ineffective | For each threat unit in a test run (Unit ID, Effective {Yes/No}, Ineffective time) |
| | **CJI** | **Assess the ability to perform immediate battlespace deconfliction C2 by JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA** | | | | | | |
| **TMOP-1** | | TMOP-5 | Time | To C2 JBD2 Indirect Fires (JCAS. JFIRES) missions | TA 5.x Execute Immediate Deconfliction C2 | TBD | | |
| | | | Time | | | | Start time when Indirect Fires (IF) request enters C2 (Decide & Deconflict - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time) |
| | | | Time | | | | End time when Indirect Fires (IF) request exits C2 (Decide & Deconflcit - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time) |
| **TMOP-2** | | TMOP-6 | Time | To C2 JBD2 Indirect Fires (JCAS. JFIRES) missions by task thread | TA 5.x Execute Immediate Deconfliction C2 | TBD | | |
| | | | Time | | | | Start time when Indirect Fires (IF) request enters C2 (Decide & Deconflict - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Test Thread #, Decide start time) |
| | | | Time | | | | End time when Indirect Fires (IF) request exits C2 (Decide & Deconflcit - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Task Thread #, Deconfliction end time) |
| **TMOP-3** | | TMOP-11 | Time | To C2 JBD2 Indirect Fires (JFIRES & JCAS) missions when airspace is controlled by requesting Service | TA 5.x Execute Immediate Deconfliction C2 | TBD | | |
| | | | Time | | | | Start time when Indirect Fires (IF) request enters C2 (Decide & Deconflict - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time, Requesting Service, Airspace Owner) |
| | | | Time | | | | End time when Indirect Fires (IF) request exits C2 (Decide & Deconflcit - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time, Requesting Service, Airspace Owner) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TMOP-4** | | TMOP-12 | Time | To C2 JBD2 Indirect Fires (JFIRES & JCAS) missions when airspace is controlled by another Service different from requesting Service | TA 5.x Execute Immediate Deconfliction C2 | | | |
| | | | Time | | | | Start time when Indirect Fires (IF) request enters C2 (C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time, Requesting Service, Airspace Owner) |
| | | | Time | | | | End time when Indirect Fires (IF) request exits C2 (Approved/Disapproved box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time, Requesting Service, Airspace Owner) |
| | | | | | | | | |
| **TMOP-5** | | TMOP-13 | Number | Of Indirect Fires (JFIRES & JCAS) missions denied and reason for denial (higher priority mission, irresolvable airspace conflict, etc) | TA 5.x Execute Immediate Deconfliction C2 | | | |
| | | | | | | | Count of IF (JFIRES/JCAS) missions denied | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Status {Pending/Approved/Disapproved}) |
| | | | | | | | Reason for IF (JFIRES/JCAS) missions denial | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Status {Pending/Approved/Disapproved}, Disapproval reason) |
| | | | | | | | | |
| | | | | | | | | |
| | **CJI** | **Assess the ability to perform joint close air support by JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA** | | | | | | |
| | | | | | | | | |
| **TMOP-6** | | TMOP-23 | Time | To get ordnance on target after initiating air-to-surface attack request | TA 3.2.2 Conduct Joint Close Air Support | TBD | | |
| | | | Time | | | | Start time when IF for JCAS request is initiated (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JCAS,CCA}, Threat Detection start time) |
| | | | Time | | | | End time when IF for JCAS request is approved, weapons delivered, and BDA completed (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JCAS,CCA}, BDA Completion time) |
| | | | | | | | | |
| **TMOP-7** | | TMOP-24 | Percent | Of requested missions executed | TA 3.2.2 Conduct Joint Close Air Support | TBD | | |
| | | | Number | | | | Count of IF for JCAS missions initiated | For each IF request in a test run (IF ID #, Type {JCAS,CCA}, Status {Pending, Approved, Disapproved}) |
| | | | Number | | | | Count of IF for JCAS missions approved | For each IF request in a test run (IF ID #, Type {JCAS,CCA}, Status {Approved}) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **CJI** | **Assess the ability to perform joint fires by JAGS across a full range of military operations in order to destroy or neutralize threat forces in the JOA** | | | | | | |
| | | | | | | | | |
| **TMOP-8** | | TMOP-26 | Percent | Targets attacked IAW requests for fires | TA 3.2.1 Conduct Joint Fires | TBD | | |
| | | | Number | | | | Count of IF for JFIRES missions initiated | For each IF request in a test run (IF ID #, Type {JFIRES}, Status {Pending, Approved, Disapproved}) |
| | | | Number | | | | Count of IF for JFIRES missions approved | For each IF request in a test run (IF ID #, Type {JFIRES}, Status {Approved}) |
| | | | | | | | | |
| **TMOP-9** | | TMOP-27 | Time | To get ordnance on target after initiating task (Immediate Call For Fire) | TA 3.2.1 Conduct Joint Fires | TBD | | |
| | | | Time | | | | Start time when IF for JFIRES request is initiated (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES}, Threat Detection start time) |
| | | | Time | | | | End time when IF for JFIRES request is approved, weapons delivered, and BDA completed (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES}, BDA Completion time) |
| | | | | | | | | |
| | | | | | | | | |
| | | **M-SoS-1** | | **Speed of Command** | | | | |
| **MOSA-1** | | Sub-M-SoS-1a | Time | Speed of Decision (C2 JBD2 time minus the Deconfliction time) See Mission Task Flow Diagram 3 | | TBD | | |
| | | | Time | | | | Start time when IF request accepted (Accepted C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time) |
| | | | Time | | | | End time when IF request exits request evaluated (Evaluate C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide end time) |
| **MOSA-2** | | Sub-M-SoS-1b | Time | Speed of Deconfliction (Airspace cleared or assigned box) | | TBD | | |
| | | | Time | | | | Start time when IF request enters airspace cleared/assigned box (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction start time) |
| | | | Time | | | | End time when IF request exits approved/disapproved box (see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time) |
| | | | | **Force Agility** | | | | |
| **MOSA-3** | | Sub-M-SoS-3a | Incidents | Force Simultaneity (Simultaneous JCAS/JFIRES C2) | | | | |
| | | | Time | | | | Start time when IF request accepted (Accepted C2 box - see Mission Task Flow Diagram) | For each IF request in a test run (IF ID #, Type {JFIRES,JCAS,CCA}, Decide start time) |
| | | | Time | | | | End time when IF request exits C2 (Approved/Disapproved box - see Mission Task Flow Diagram) | For each IF request in a test run (ID #, Type {JFIRES,JCAS,CCA}, Deconfliction end time) |
| | | | | | | | Count of IF (JFIRES/JCAS/CCA) missions submitted | For each IF request in a test run (ID #, Type {JFIRES,JCAS,CCA}) |

| Data Element | Data Fields | Possible Values | Amplifying information | Data Structure | Data Format | Collection Rate | Test Run |
|---|---|---|---|---|---|---|---|
| **Task Thread** | Task Thread ID Number | 1 to 8 | Identifies the task thread | Integer | 1,2,3, … | Before test event | All |
| **Test Trial** | Test Trial ID # | 1 to 8 | Identifies the Test Trial | Integer | 1,2,3, … | Before test event | All |
| **Test Run** | | | | | | | |
| | Test Trial ID # | 1 to 8 | Identifies the Test Trial | Integer | 1,2,3, … | Each Test Run | All |
| | Test Run ID # | 1, 2 | Identifies the Test Run for a Test Trial | Integer | 1,2,3, … | Each Test Run | All |
| | Test Run Start time | 0 | Test run start at time 1 | minutes | xxx.xx | Each Test run | All |
| | Test Run End time | 180 | Assumes each trial lasts 180 minutes | minutes | xxx.xx | Each Test run | All |
| | Threat Unit Count at Start Time | | Count of threat units in the JOA at test run start time | Integer | 1,2,3, … | Each Test run | All |
| | Threat Unit Count at End Time | | Count of threat units in the JOA at test run end time | Integer | 1,2,3, … | Each Test run | All |
| **Threat Units** | | | | | | | |
| | Unit ID Number | 1,2,3, … | Identifies the threat unit | Integer | 1,2,3, … | Each threat unit | All |
| | Effective | Yes/No | All threat units starts test run as effective (yes) | Text | E, I | Status change on threat unit | All |
| | Ineffective time | | Time threat unit becomes ineffective. Blank if does not become ineffective | minutes | xxx.xx | Status change on threat unit | All |
| | Status | Active, Killed, Neutralized | Starts out "active". Neutralized is when ineffective but not killed (damaged) | Text | A, K, N | Status change on threat unit | All |

**Indirect Fires Request**

| | | | | | | |
|---|---|---|---|---|---|---|
| ID Number | 1,2,3, … | Sequential ID number for each indirect fires (JCAS/JFIRES request | Integer | 1,2,3, … | At creation of request | All |
| Type | JCAS, JFIRES, CCA | Type of IF | Text | JC, JF, CC | At creation of request | All |
| Threat Unit ID # | 1,2,3, … | Identifies the threat unit that is prosecuted in the IF | Integer | 1,2,3, … | At creation of request | All |
| Decision Start time | | Time request enters Decision process | minutes | xxx.xx | At creation of request | All |
| Decision End Time | | Time request exits Decision process | minutes | xxx.xx | At end of request | All |
| Status | Pending, Approved, Disapproved | Starts as Pending | Text | P, A, D | As status changes | All |
| Task Thread ID | 1 to 8 | Identifies which task thread the request comes from | Integer | 1, 2, 3, … | At creation of request | All |
| Requesting Service | USA, USMC, USAF, USN | Which Service initiates request | Text | A, C, F, N | At creation of request | All |
| Airspace Owner | USA, USMC, USAF, USN | Which Service owns the airspace at time of request | Text | A, C, F, N | At creation of request | All |
| Disapproval Reason | | Only when Request Status is disapproved | Text | String | At Disapproval time | All |
| Deconfliction Start Time | | Time request enters Deconfliction | minutes | xxx.xx | As deconfliction occurs | All |
| Deconfliction End Time | | Time request exits Deconfliction | minutes | xxx.xx | As deconfliction occurs | All |
| Threat Detection Start Time | | Time threat is detected and identified | minutes | xxx.xx | At start of detection box | All |
| BDA Completion Time | | Time BDA is completed on threat | minutes | xxx.xx | At end of BDA box | All |

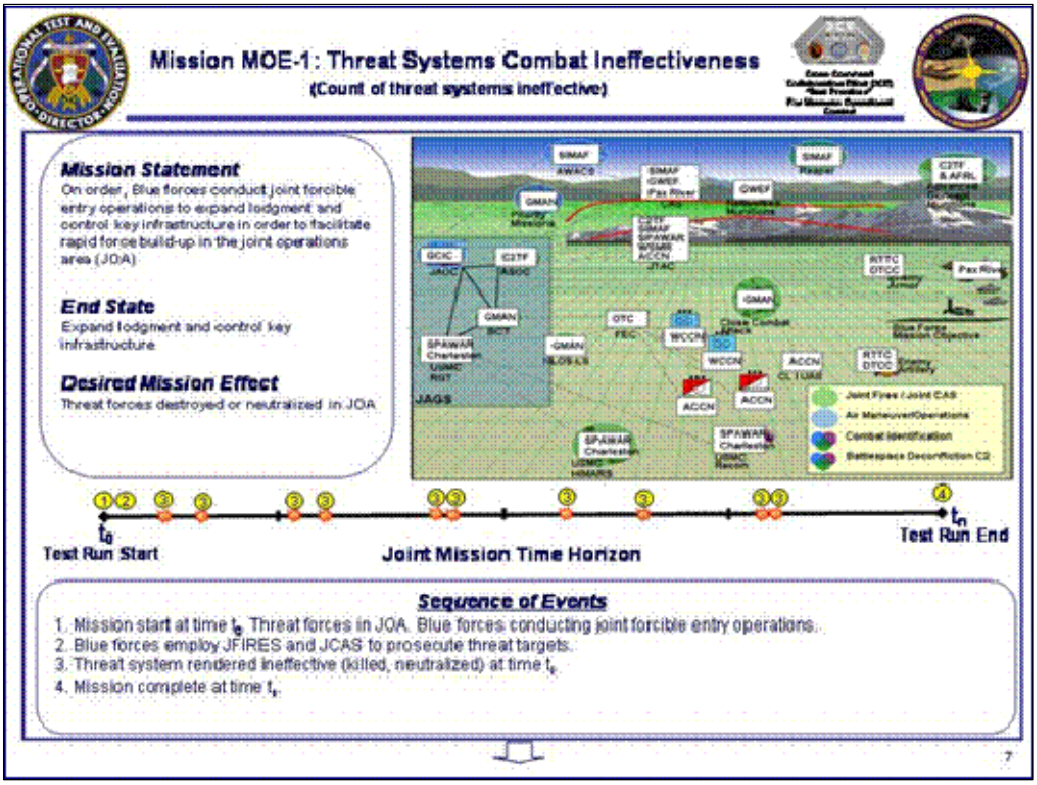
*This page intentionally left blank*

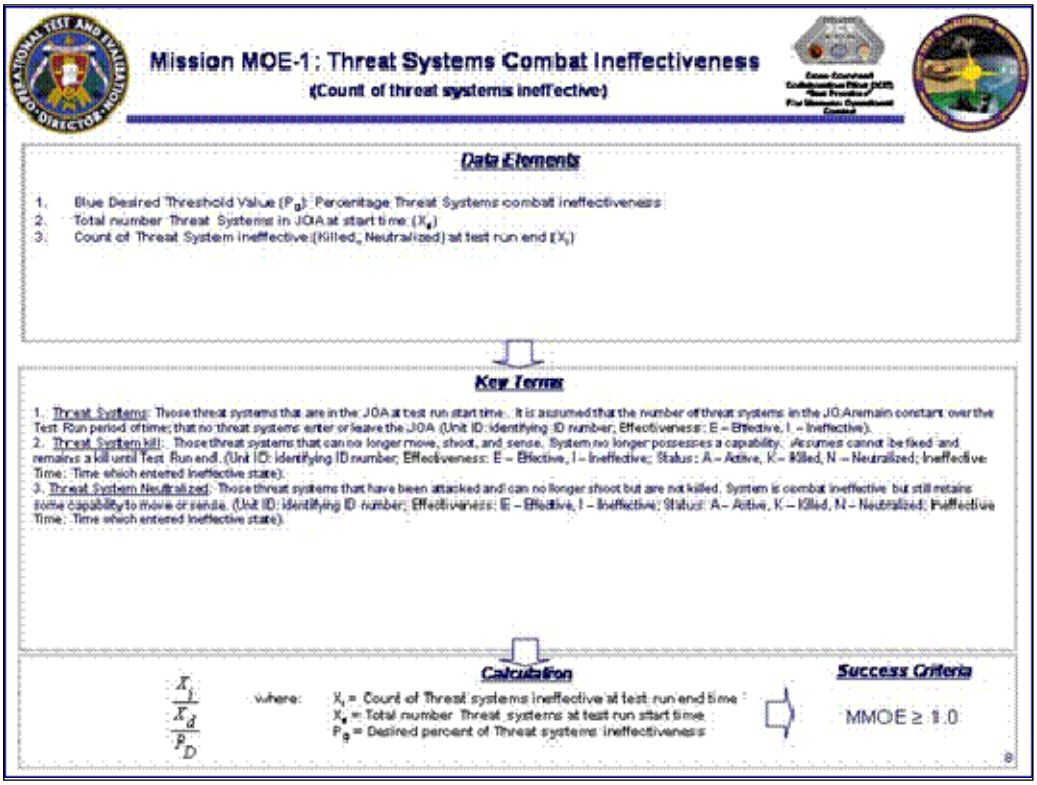
# Annex B: Crosswalk Matrix

| Critical Joint Issue | SoS Config | | | | Standards | Conditions |
|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | C4 | | |
| **Assess the ability to perform Battlespace Dynamic Deconfliction C2 by Current JAGS under a full range of military operations in order to achieve destroyed or neutralized threat forces in JOA** | X | X | | | S1 | T1E1 |
| **Assess the ability to perform Battlespace Dynamic Deconfliction C2 by Future JAGS under a full range of military operations in order to achieve destroyed or neutralized threat forces in JOA** | | | X | X | S1 | T1E1 |
| **Additional SUT configuration factor** | | | | | | |
| **Command and Control Processes** | | | | | | |
| • "Procedural" Processes | X | | X | | | |
| • "Expedited" Processes | | X | | X | | |

*This page intentionally left blank*

# Annex C: Evaluation Views



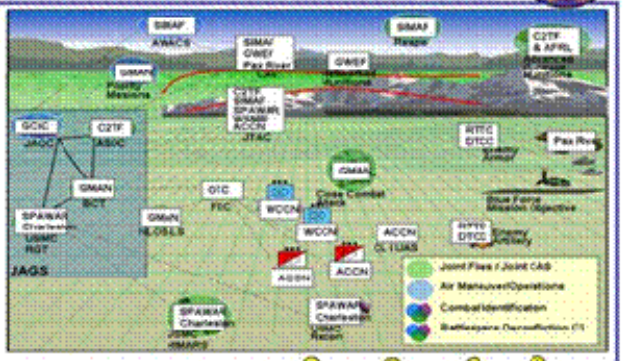
## Task MOP 1: Time to C2 JBD2 Indirect Fires (JFIRES & JCAS) Missions

### Mission Statement

On order, Blue forces conduct joint forcible entry operations to expand lodgment and control key infrastructure in order to facilitate rapid force build-up in the joint operations area (JOA).

### UJTL

TA 5.x Execute Immediate Deconfliction C2



### Sequence of Events

1. Mission start at time $t_0$. Threat forces in JOA. Blue forces conducting joint forcible entry operations.
2. C2 start time occurs when Indirect Fires request enters JBD2 "deride process" at time $t_{si}$.
3. C2 end time occurs when Indirect Fires request exits the JBD2 "deconflict process" at time $t_{ei}$.

11

## Task MOP 1: Time to C2 JBD2 Indirect Fires (JFIRES & JCAS) Missions

### Data Elements

1. Indirect Fires ID: IF request $i$ ($IF_i$) is instantiated and processed through JBD2
2. Decision start time ($t_{si}$): Time C2 starts and is when Indirect Fires request $IF_i$ is accepted
3. Deconfliction end time ($t_{ei}$): Time C2 ends and is when Indirect Fires request $IF_i$ is approved/disapproved
4. Indirect Fires type (JCAS, JFIRES, CCA)
5. Blue desired threshold value ($T_D$) for C2

### Key Terms

1. Indirect Fires: An indirect fire (IF) request is initiated as a concise message prepared by the observer. It contains all information needed by C2 to determine the method of target attack. It is a request for fire, not an order. It must be sent quickly but clearly enough that it can be understood, recorded, and read back, without error, by the recorder. For the test event we will assume the start time is when the observer tells C2 he has seen a target so the C2 can start the IF while the target location is being determined. (FM6-30)
2. Indirect Fire Accepted: When the indirect fire is accepted into C2 at time, $t_{si}$
3. Indirect Fire Disapproval/ Approval: When the indirect fire is approved or disapproved in the C2 at time, $t_{ei}$
4. JFIRES: Joint fires are defined as fires produced during the employment of forces from two or more components in coordinated action toward a common objective. Joint Publication 3-0.
5. JCAS: "DOD Dictionary of Military and Associated Terms," defines CAS as "air action by fixed- and rotary-wing aircraft against hostile targets which are in close proximity to friendly forces and which require detailed integration of each air mission with the fire and movement of those forces." Joint CAS is CAS conducted through joint air operations (as described in Joint Pub 3-56.1, "Command and Control for Joint Air Operations") or in the case of rotary-wing aircraft, through the establishment of a command relationship between components (Chapter III, "Requesting and Tasking," paragraph 5 Joint Pub 1-02).
6. CCA: CCA is a technique to control attack helicopters (AH-64 or OH-58D) in a close air support role, usually very near to friendly troops in contact. It is a method to get the pilot to see the target that you are seeing. It is similar to the Call for Fire for artillery or mortars. The primary difference is that you have to talk to the pilots, who also need to see the target. [Reference: FM 3-04.111 Utility and Cargo Helicopter Operations December 2007]

### Calculation

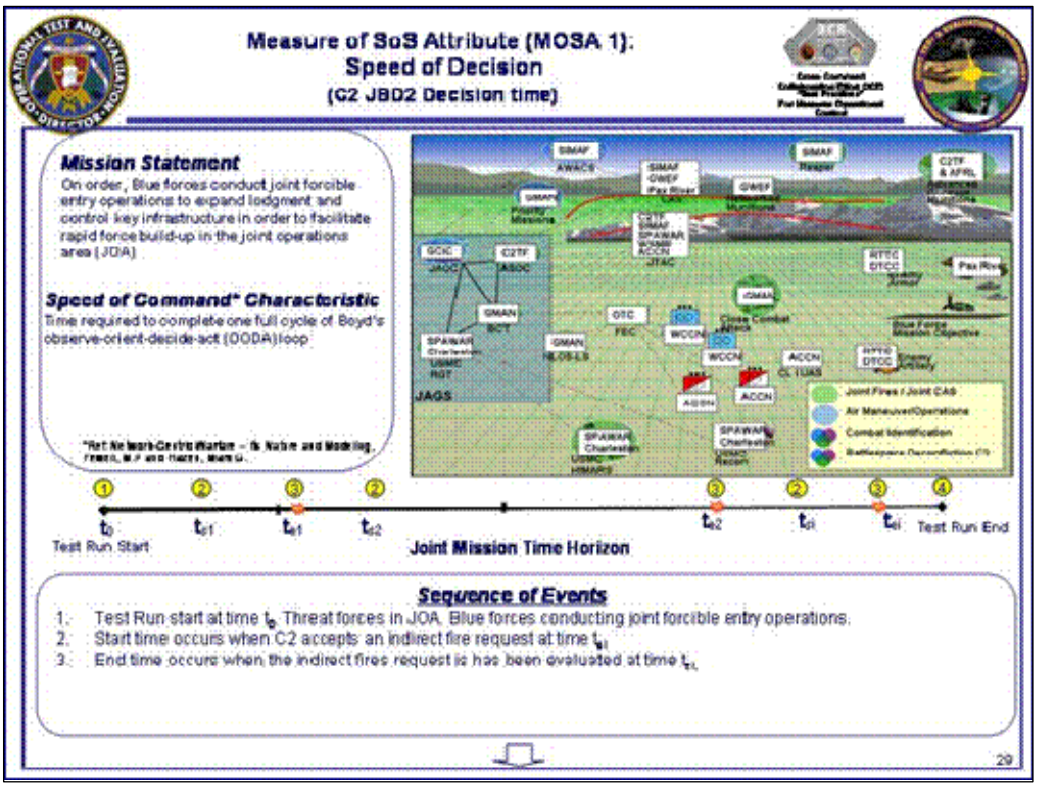
$T_{ei} - T_{si}$

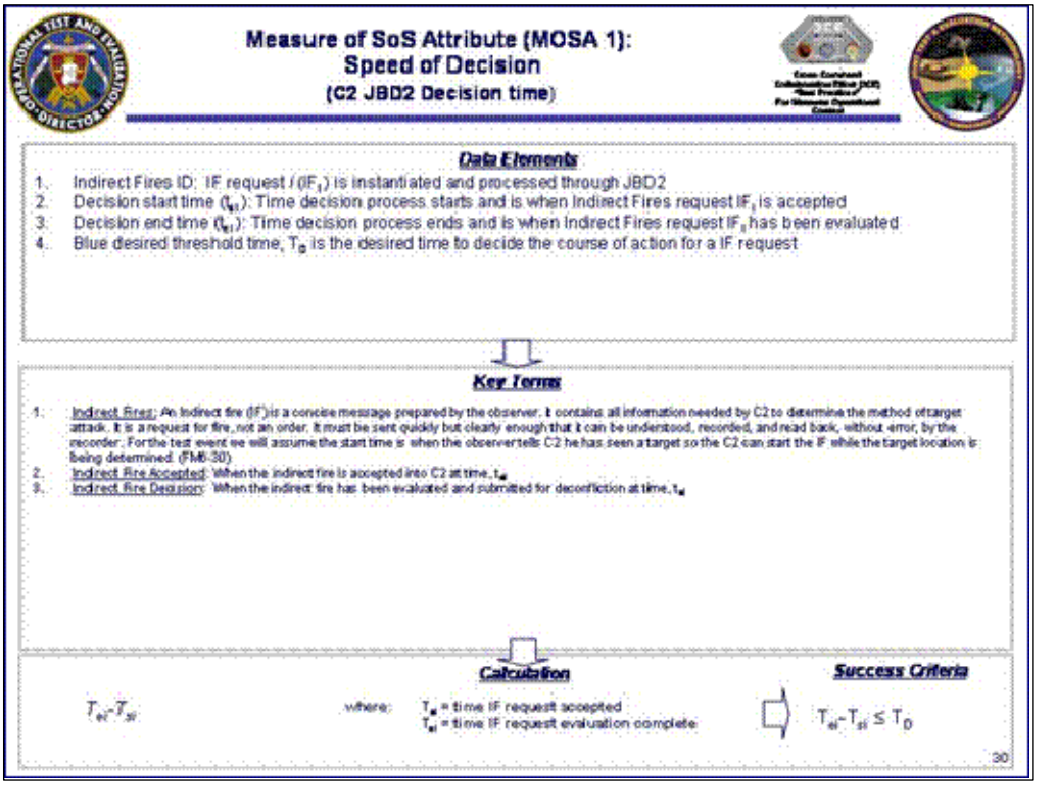
where: $T_{si}$ = time IF is accepted into C2
$T_{ei}$ = time IF is approved/disapproved

### Success Criteria

$T_{ei} - T_{si} \leq T_D$

12


Measure of SoS Attribute (MOSA 1):
Speed of Decision
(C2 JBD2 Decision time)
Mission Statement
On order, Blue forces conduct joint forcible entry operations to expand lodgment and control key infrastructure in order to facilitate rapid force build-up in the joint operations area (JOA).
Speed of Command* Characteristic
Time required to complete one full cycle of Boyd's observe-orient-decide-act (OODA) loop
JAGS
Test Run Start
Joint Mission Time Horizon
Test Run End
Sequence of Events
1. Test Run start at time $t_0$. Threat forces in JOA. Blue forces conducting joint forcible entry operations.
2. Start time occurs when C2 accepts an indirect fire request at time $t_{s1}$
3. End time occurs when the indirect fires request is has been evaluated at time $t_{e1}$
29


Measure of SoS Attribute (MOSA 1):
Speed of Decision
(C2 JBD2 Decision time)
Data Elements
1. Indirect Fires ID: IF request i ($IF_i$) is instantiated and processed through JBD2
2. Decision start time ($t_{si}$): Time decision process starts and is when Indirect Fires request $IF_i$ is accepted
3. Decision end time ($t_{ei}$): Time decision process ends and is when Indirect Fires request $IF_i$ has been evaluated
4. Blue desired threshold time, $T_D$ is the desired time to decide the course of action for a IF request
Key Terms
1. Indirect Fires: An indirect fire (IF) is a concise message prepared by the observer. It contains all information needed by C2 to determine the method of target attack. It is a request for fire, not an order. It must be sent quickly but clearly enough that it can be understood, recorded, and read back, without error, by the recorder. For the test event we will assume the start time is when the observer tells C2 he has seen a target so the C2 can start the IF while the target location is being determined. (FM6-30)
2. Indirect Fire Accepted: When the indirect fire is accepted into C2 at time, $t_{si}$
3. Indirect Fire Decision: When the indirect fire has been evaluated and submitted for deconfliction at time, $t_{ei}$
Calculation
$T_{ei}$-$T_{si}$
where: $T_{si}$ = time IF request accepted
$T_{ei}$ = time IF request evaluation complete
Success Criteria
$T_{ei}$-$T_{si}$ ≤ $T_D$
30

*This page intentionally left blank*

# Annex D: Acronyms and Abbreviations

| | |
|---|---|
| 46TS | 46th Test Squadron |
| AAR | After Action Review |
| ABCS | Army Battle Command System |
| ACETEF | Air Combat Environment Test and Evaluation Facility |
| ACCN | Army Command and Control Network |
| ADP | Automated Data Processing |
| AFATDS | Advanced Field Artillery Tactical Data System |
| AFC2ISRC | Air Force Command and Control, Intelligence, Surveillance, and Reconnaissance Center |
| AIS | Automated Information System |
| AO | Action Officer |
| ASOC | Air Support Operations Center |
| A/V | Audio Visual |
| C2 | Command and Control |
| C4I | Command, Control, Communications, Computers, and Intelligence |
| CD | Compact Disc |
| CDC | Concept Development Conference |
| CDD | Capability Development Document |
| CDR | Critical Design Review |
| CEM | Capability Evaluation Metamodel |
| CJI | Critical Joint Issue |
| CM | Capability Manager |
| CONOPS | Concept of Operations |
| CTM | Capability Test Methodology |
| CTO | Combined Test Organization |
| DAC | Data Analysis Center |
| DAP | Data Analysis Plan |
| DCAT | Data Collection and Analysis Team |
| DCT | Data Collection Team |
| DM | Data Manager |

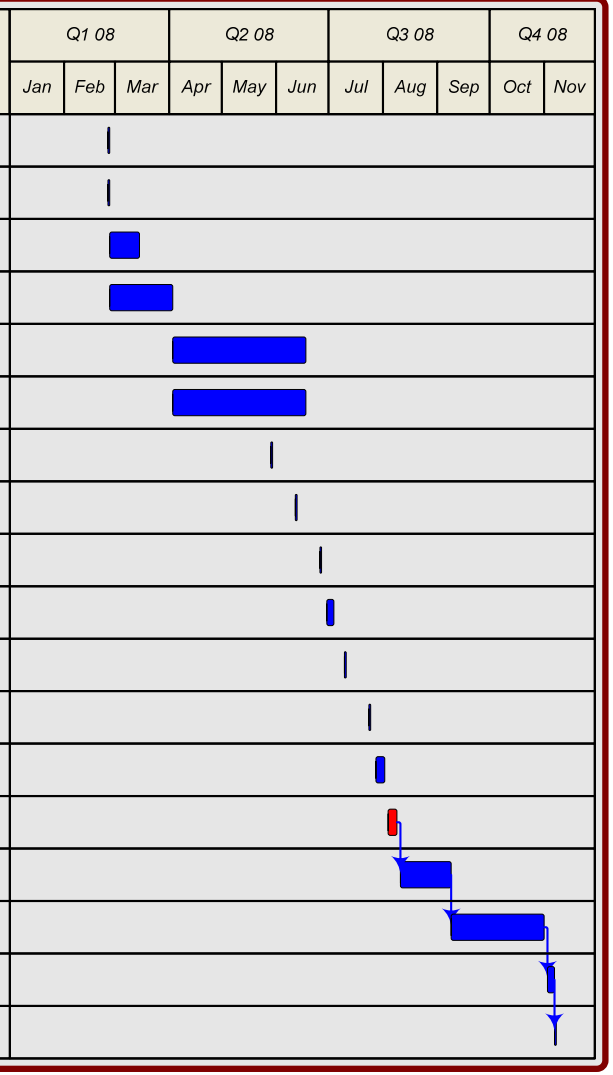
| | |
|---|---|
| DOD | Department of Defense |
| DODAF | Department of Defense Architecture Framework |
| DODD | DOD Directive |
| DOE | Design of Experiment |
| DOT&E | Director, Operational Test and Evaluation |
| DOTMLPF | Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities |
| DPS | Defense Planning Scenarios |
| DRCT | Distributed Range Coordination Team |
| DTP | Detailed Test Plan |
| DV | Dependent Variable |
| EM | Event Manager |
| FCS | Future Combat System |
| FOUO | For Official Use Only |
| FPC | Final Planning Conference |
| FY | Fiscal Year |
| GWEF | Guided Weapons Evaluation Facility |
| IAW | In Accordance With |
| ICD | Initial Capabilities Document |
| IDRL | Integrated Data Requirements List |
| IPC | Initial Planning Conference |
| IRCC | Inter-Range Control Center |
| JAMUS | Joint Aviation, Missile, and Unmanned Systems |
| JBD2 | Joint Battlespace Dynamic Deconfliction |
| JCA | Joint Capability Area |
| JCE | Joint Capability Evaluation |
| JCIDS | Joint Capabilities Integration and Development System |
| JMe | Joint Mission Effectiveness |
| JME | Joint Mission Environment |
| JMETC | Joint Mission Environment Test Capability |
| JOC-T | Joint Operational Context for Test |
| JOpsC | Joint Operations Concept |
| JPO | JT&E Program Office |

| | |
|---|---|
| JROC | Joint Requirements Oversight Council |
| JSIC | Joint Systems Integration Command |
| JT&E | Joint Test and Evaluation |
| JTD | Joint Test Director |
| JTF | Joint Task Force |
| JTEM | Joint Test and Evaluation Methodology |
| KPP | Key Performance Parameter |
| LVC | Live, Virtual, Constructive |
| LVC-DE | Live, Virtual, Constructive Distributed Environment |
| M&P | Methods and Processes |
| M&S | Modeling and Simulation |
| MMOE | Mission Measure of Effectiveness |
| MOA | Memorandum of Agreement |
| MPC | Mid-Planning Conference |
| MS | Milestone |
| OSD | Office of the Secretary of Defense |
| PDR | Preliminary Design Review |
| PID | Program Introduction Document |
| PM | Program Manager |
| POA | Plan of Action |
| POC | Point of Contact |
| PTP | Program Test Plan |
| RDD | Requirements Deep Dive |
| RTTC | Redstone Technical Test Center |
| SCG | Security Classification Guide |
| SCI | Sensitive Compartmented Information |
| SDD | System Design Document |
| SME | Subject Matter Expert |
| SOC | Statement of Capability; Statement of Capabilities |
| SoS | System of Systems; Systems of Systems |
| SPAWAR | Space and Naval Warfare |
| SPG | Strategic Planning Guidance |
| SSC | Space and Warfare (SPAWAR) Systems Center |

| | |
|---|---|
| SSO | Special Security Office |
| STE | Secure Telephone Equipment |
| STU-III | Secure Telephone Unit, Third Model |
| SUT | System Under Test |
| T&E | Test and Evaluation |
| TACP-CASS | Tactical Air Control Party - Close Air Support System |
| TAG | Technical Advisory Group |
| TBD | To Be Determined |
| TBMCS | Theater Battle Management Core Systems |
| TC | Transformation Center |
| TCC | Telephone Conference Call |
| TDAWG | Test Design and Analysis Working Group |
| TDY | Temporary Duty |
| TE2 | Test Event 2 |
| TEMP | Test and Evaluation Master Plan |
| TENA | Test and Training Enabling Architecture |
| TES | Test and Evaluation Strategy |
| TIR | Test Incident Report |
| TM | Test Manager |
| TMOP | Task Measure of Performance |
| TRR | Test Readiness Review |
| TTP | Tactics, Techniques, and Procedures |
| UJTL | Universal Joint Task List |
| USJFCOM | United States Joint Forces Command |
| USMC | US Marine Corps |
| USMTF | Uniform Services Messages Text Format |
| v | Version |
| V&V | Verification and Validation |
| VIP | Very Important Person |
| VV&A | Verification, Validation, and Accreditation |
| WARCAP | Warfighter Capability Demonstration Center |

# Annex E: Analysis Schedule

| ID | Task Name | Start | Finish | Duration |
|---|---|---|---|---|
| 1 | TDAWG Kick Off Meeting | 2/26/2008 | 2/26/2008 | 1d |
| 2 | JBD2 Test Design Concept | 2/26/2008 | 2/26/2008 | 1d |
| 3 | Define JBD2 Data elements | 2/27/2008 | 3/14/2008 | 13d |
| 4 | Develop JBD2 Evaluation Plans | 2/27/2008 | 4/2/2008 | 26d |
| 5 | Develop Test Trial Execution Plan | 4/3/2008 | 6/17/2008 | 54d |
| 6 | Develop Data Collection Plan | 4/3/2008 | 6/17/2008 | 54d |
| 7 | TDAWG Meeting | 5/29/2008 | 5/29/2008 | 1d |
| 8 | TDAWG Meeting | 6/12/2008 | 6/12/2008 | 1d |
| 9 | TDAWG Meeting | 6/26/2008 | 6/26/2008 | 1d |
| 10 | Conduct Analysis Spiral #1 | 6/30/2008 | 7/3/2008 | 4d |
| 11 | TDAWG Meeting | 7/10/2008 | 7/10/2008 | 1d |
| 12 | TDAWG Meeting | 7/24/2008 | 7/24/2008 | 1d |
| 13 | Conduct JBD2/Analysis Dry Runs | 7/28/2008 | 8/1/2008 | 5d |
| 14 | Test Execution | 8/4/2008 | 8/8/2008 | 5d |
| 15 | Conduct Data analysis activities | 8/11/2008 | 9/8/2008 | 21d |
| 16 | Write Final Report | 9/9/2008 | 10/31/2008 | 39d |
| 17 | Conduct Technical Review & Editing | 11/3/2008 | 11/6/2008 | 4d |
| 18 | Final Report Due | 11/7/2008 | 11/7/2008 | 1d |



*This page intentionally left blank*

**ANNEX K – SAMPLE JOINT CAPABILITY EVALUATION REPORT (JCER)**

# Joint Test and Evaluation

# Joint Capability Evaluation
# Month Year

DISTRIBUTION STATEMENT: Distribution limited to U.S. Government agencies and their supporting contractors. Other requests for this document shall be referred to

# I. EXECUTIVE SUMMARY

1. Introduction – *Brief overview of project, lead and supporting Services and Commands and description of operational capability impacted by the problem addressed by this report.*

2. Test Purpose
   A. Test & Evaluation Problem – *Problem statement in operational context*
      (1) Joint Mission Level
      (2) System of Systems (SoS) Level
      (3) System Under Test (SUT) Level
   B. Test & Evaluation Issues – *Issues from Program Test Plan (PTP) and any significant sub-issues of interest to Decision Authority*
      (1) Joint Mission Level
      (2) System of Systems (SoS) Level
      (3) System Under Test (SUT) Level
   C. Test & Evaluation Objectives - *Statements of the principle test objectives addressed*
      (1) Joint Mission Level
      (2) System of Systems (SoS) Level
      (3) System Under Test (SUT) Level

3. Test Concept Overview
   A. *Indicate if previous or baseline testing was accomplished, what that testing provided via findings and contrast with this additional test.*
   B. *List organizations that participated in the test (overall, tabular format).*
   C. *List test articles (Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance, and Targeting (C4ISR) architectures, changes to tactics, techniques, and procedures (TTP), processes, evaluation of training , etc.) tested and compare pre-test capability to support warfighter with post-test capabilities (enhanced capability).*
   D. *Information should be limited to what is needed to put findings, conclusions and recommendations in context.*

4. Significant Findings, Conclusions and Recommendations – *Should answer the principle issues identified in 2.B. listing test finding, conclusion and recommendation in bulletized format. Rule of thumb is that if there is no recommendation for a Flag/General Officer level to implement, do not put the finding in the executive summary.*

5. Critical Test Product Issues – *Name of product, description, which organization is accepting test product, comments, problems.*

6. Summary – *Address the key items the reader needs to take away. In some cases this may be the only paragraph that is read.*

## II. INTRODUCTION

1. Introduction – *Identify lead Service(s) and primary sponsoring and participating organizations.*

2. Joint Test and Evaluation (JTE) Background - *Brief description of operational capability impacted by problem addressed by project, brief discussion of the problem statement, and overview of principle issues answered.*

3. Authority – *Provide authority mechanism, issue date and summary of purpose as stated by the authorizing mechanism.*

4. Test Articles – *Briefly describe the test articles that were tested and evaluated (changes to TTP, SoS architectures, concept of operations (CONOPS), process model, test methods, etc.).*

## III. TEST CONCEPT OVERVIEW

1. Introduction – *Short description of test concept and method development in relation to test issues.*

2. Problem and Test Issues – *Highlight issues and sub-issues answered in this report.*

3. Test Objectives – *Specific statement of intent. Emphasize test objectives that directly impact the problem.*

4. Test Scenarios and Vignettes – *Briefly describe the vignette for each test event executed. Provide a general scenario overview and description of contribution to overall test plan/outline.*

5. Test Constraints and Limitations – *List all applicable constraints, limitations and related mitigation measures. If there is no mitigation measure for a limitation or constraint, indicate the impact of the limitation or constraint on the test event and on test findings and conclusions.*

6. Supporting Organizations – *List of organizations involved in supporting test planning, test execution, reporting and test product transition.*

7. Schedule/Chronology – *Indicate major test event planning, execution and post-test milestones (includes general officer steering committee (GOSC) meetings, Quick Look report release dates and test event report release dates). [Gantt Chart]*

8. System Descriptions – *Describe the system under test (SUT), system of systems (SoS) components, and overall joint mission environment (JME). Identify relevant SoS interfaces, architectures, operational process models, SoS component characteristics (live, virtual or constructive), joint mission profile, operational mission profiles and other similar information (figures, photos, Department of Defense Architecture Framework [DoDAF] models, etc.) that will provide a meaningful context for understanding the importance of test findings, why specific conclusions were reached and the importance of implementing the recommendations.*

## IV. TEST FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

1. Introduction – *Summary description of data types, format, quality, and fidelity in representing operational environment. Describe measures to ensure test rigor and operational realism. Discuss use and relationships of SUT to SoS, relevant interface descriptions, and measures (key performance parameters (KPP), measures of effectiveness (MOE), and measures of performance (MOP)) as they relate to joint mission effectiveness (JMe).*

2. Summary of Significant Findings – *Bullet format presenting items of significant importance to improving or providing operational capability within JME and contribution to joint mission effectiveness (JMe). (4-5 items)*

3. Test Scenarios/Vignettes – *General description of the test scenario and associated vignettes from the test plan. If a particular scenario/vignette was not used in conduct of a test event or a different scenario/vignette from those in the test plan was used in the conduct of a test event, a detailed discussion addressing these anomalous actions should be provided.*

4. Test Events Relating to System Under Test Performance Evaluation

   A. Test Event 4.1
      (1) Test Objective – *Object of the test as stated in the test plan. Indicate that a test item at this level may also be used in the evaluation of lower level effectiveness, (e.g., the configuration of Test 4.1 is identical to Test Event 5.2 and 6.4.).*
      (2) Test Procedure – *Should refer to test plan to indicate the procedures used or, if different procedures were used, then a full description of the procedure used (how equipment was operated, how data was gathered, types of data, formats, etc.) must be provided.*
      (3) Data Analysis – *Describe how data was analyzed to include significant assumptions and mathematical relationships. Definitions of success, failure or incomplete test; material failures and failure categories; and "up" and "down" times should be provided. Any and all formulas used for calculations should be presented if different from those indicated by the test plan.*
      (4) Test Results/Findings –*Indicate the results in clear and unambiguous terms, i.e., satisfactory, unsatisfactory, and inconclusive. Qualitative results must directly answer the test objective. Summarize the data (database) rather than presenting a mass of raw data (e.g., 120 test runs were conducted. The mean detection range was X, maximum and minimum observed ranges were Y and Z. All test runs were made within an environment of active jamming.)*
         (a) Summarize the significant findings of the test event.
         (b) Provide an overview of the findings within a characterization grouping, i.e., accuracy problems, and any operational elements affected, i.e., navigation.
         (c) Issue Item 4.1.1 – *Test finding should be a single concise statement, avoiding the use of "can" or "may," (e.g., Navigation accuracy is degraded under certain conditions).*

c.1 Sub-issue Item 4.1.1.1 – *As needed.*
c.2 Sub-Issue Item 4.1.1.2 – *As needed.*
i. Discussion – *Address factors that support findings and conclusion statements. Consider impact on warfighting capability, system capabilities, organizations that will be impacted, doctrine, organizational, training, material, leadership and education, personnel, and facilities (DOTMLPF) factors, TTP, etc.*
ii. Deficiencies – *Discuss any deficiencies or anomalies that were noted during testing that are determined to have an impact on joint mission effectiveness (JMe).*
iii. Additional Considerations – *Discuss any additional considerations that may relate to SUT operational employment within a SoS construct and impacts on JMe.*
iv. Conclusions – *Should be directly related to findings. There may be more than one finding related to a Conclusion.*
v. Recommendations – *Findings and conclusions must directly support and justify recommendations. There may be multiple recommendations resulting from a single finding or a single recommendation resulting from multiple findings. If further testing or evaluation is needed; describe the specific aspects that need to be tested and why.*

(d) Issue Item 4.1.2
i. Discussion
ii. Deficiencies
iii. Additional Considerations
iv. Conclusions
v. Recommendations

5. Test Events Relating to System of System Effectiveness Evaluation

A. Test Event 5.1
(1) Test Objective
(2) Test Procedure
(3) Data Analysis
(4) Test Results/Findings
(a) Summary
(b) Overview
(c) Issue Item 1
i. Discussion
ii. Deficiencies
iii. Additional Considerations
iv. Conclusions
v. Recommendations
(d) Issue Item 2

6. Test Events Relating to Joint Mission Effectiveness Evaluation

A. Test Event 6.1
B. Test Event 6.2

7. Summary – *Summarize recommendations, overall impacts on capabilities and JMe, organization that has action, benefits of implementation and estimate of costs (DOTMLPF) and schedule.*

   *Include a matrix of Test Events to Evaluation Level to provide an overview of common test events that may impact more than one level of evaluation.*

| **Test Set** | **System Under Test (SUT) Level Effectivity** | **System of System (SoS) Level Effectivity** | **Joint Mission Level Effectivity** |
|---|---|---|---|
| *4.1, 5.2, 6.4* | *X* | *X* | *X* |
| *4.2* | | | *X* |
| *5.1* | | *X* | |
| *6.1* | *X* | | |
| *6.2* | *X* | | |
| *6.3* | *X* | | |

## V. TEST PRODUCTS

1. Test Product Description – *List the affected JME test products in tabular format (DOTMLPF issues, processes, C4ISR architectures, TTP, warfighting operations, joint capabilities, etc.); relevant test findings and recommendations; the benefits of adopting the related recommendations; and organization that should be assigned responsibility (funding, maintenance, ownership) for adoption and institutionalization of test recommendations.*

2. Test Product Transition – *As far as is possible discuss the benefits of test product change/transition to the warfighting community, emphasizing a specific part of that community as applicable. Discuss the coordination activities needed to transition the test product, financial factors (fielding, maintenance, training, etc.) and identify stakeholders/agencies that should be involved.*

# ANNEX L - CTM LEXICON

In order to provide conceptual consistency and an underlying business rule structure for the Capability Test Methodology (CTM), Joint Test and Evaluation Methodology (JTEM) is employing an ontology approach. An ontology can be defined as "an explicit formal specification of how to represent the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them." In keeping with this definition, the ontology supporting the CTM evaluation thread incorporates a *CTM Lexicon* to provide underlying conceptual definitions for the CTM. The *CTM Lexicon* is a cross-domain dictionary of CTM-relevant DoD terminology and definitions. Authoritative DoD sources are used, where possible, for JTEM terms and definitions. Modifications to current terminology or additional terms not currently defined in authoritative sources are noted as CTM version 3.0. This lexicon is one of the test products developed during the course of JTEM.

*This page intentionally left blank.*

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Abstraction | | | |
| | ▪ The act of identifying the essential characteristics of a thing that distinguish it from all other kinds of things. Abstraction involves looking for similarities across sets of things by focusing on their essential common characteristics. An abstraction always involves the perspective and purpose of the viewer; different purposes result in different abstractions for the same things. All modeling involves abstraction, often at many levels for various purposes. | DoDAF 1.5, Volume II<br>Annex C C-1 | 04/2007 |
| | ▪ A kind of dependency that relates two elements that represent the same concept at different abstraction levels. | DoDAF 1.5, Volume II<br>Annex C C-1 | 04/2007 |
| Accreditation | | | |
| | ▪ The official certification that a model or simulation is acceptable for use for a specific purpose. | DOD 5000.59-M<br>Definitions P.2.1.7. 87 | 01/1998 |
| Accuracy | | | |
| | ▪ The degree to which a parameter or variable or set of parameters or variables within a model or simulation conform exactly to reality or to some chosen standard or referent. See resolution, fidelity, precision. | Recommended Practices Guide<br>Fidelity RPG Special Topic --- 8 | 09/2000 |
| Adaptable/Tailorable | | | |
| | ▪ An adaptable/tailorable joint force is versatile in handling threat missions with equal success; scalable in applying appropriate mass and weight of effort; agile in shifting between different types of missions without loss of momentum; responsive to changing conditions and environments; and whose leaders are intellectually empowered by a background of experience and education. Adaptability ensures that the joint force can rapidly shift from one operation to another across the range of military operations, and adjust operations based on changing conditions. An adaptive mindset and flexible force capabilities are essential for success in countering the full spectrum of anticipated threats and challenges and enhance the joint force ability to respond with unmatched speed of decision and action.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution 4.E.5. 22 | 08/2005 |
| Agile | | | |
| | ▪ An agile joint force has the ability to move quickly and seamlessly to defuse (or help defuse) a crisis situation or effectively operate inside the decision loop of even the most capable adversary. Agility is about timeliness--thinking, planning, communicating, and acting in a manner that allows effective and efficient adaptation to an unfolding situation. Agility permits JFCs to exploit fleeting opportunities, protect friendly vulnerabilities, and adapt rapidly to changes in the operational environment--a characteristic essential to a force that is expected to succeed across the range of military operations.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution 4.E.10. 23 | 08/2005 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| All View (AV) | | | | | |
| | ▪ There are some overarching aspects of an architecture that relate to all three views. These overarching aspects are captured in the AV products. The AV products provide information pertinent to the entire architecture but do not represent a distinct view of the architecture. AV products set the scope and context of the architecture. The scope includes the subject area and time frame for the architecture. The setting in which the architecture exists comprises the interrelated conditions that compose the context for the architecture. These conditions include doctrine; tactics, techniques, and procedures; relevant goals and vision statements; Concepts of Operations (CONOPS); scenarios; and environmental conditions.<br>– AV-1 -- Overview and Summary Information:<br>Scope, purpose, intended users, environment depicted, analytical findings.<br>– AV-2 -- Integrated Dictionary:<br>Architecture data repository with definitions of all terms used in all products. | DoDAF 1.5, Volume I<br>1 | <br>1.4.5 | <br>1-9 | 04/2007 |
| Analysis | | | | | |
| | ▪ An examination of a concept using quantitative and qualitative measures to assess potential capabilities. It produces metrics that are applied to assumptions and risks and to formulate recommendations and support decisions. | CJCSI 3010.02B<br>Glossary | <br>GL | <br>GL-3 | 01/2006 |
| Analysis of Alternatives (AoA) | | | | | |
| | ▪ AoA are an important element of the defense acquisition process. An AoA is an analytical comparison of the operational effectiveness, suitability, and Life-Cycle cost of alternatives that satisfy established capability needs. Initially, the AoA process typically explores numerous conceptual solutions with the goal of identifying the most promising options, thereby guiding the Materiel Solution Analysis phase [previously, "Concept Refinement phase"] (see section 3.3.3). Subsequently, at Milestone B (which usually represents the first major funding commitment to the acquisition program), the AoA is used to justify the rationale for formal initiation of the acquisition program. An AoA normally is not required at Milestone C unless significant changes to threats, costs, or technology have occurred, or the analysis is otherwise deemed necessary by the Milestone Decision Authority. For a joint program, the lead DOD Component normally is responsible for the preparation of a single comprehensive analysis. | Defense Acquisition University Guidebook<br>3 | <br>3.3 | <br>--- | 12/2008 |
| | ▪ The evaluation of the performance, operational effectiveness, operational suitability, and estimated costs of alternative systems to meet a mission capability. The AoA assesses the advantages and disadvantages of alternatives being considered to satisfy capabilities, including the sensitivity of each alternative to possible changes in key assumptions or variables. The AoA is one of the key inputs to defining the system capabilities in the capability development document. | CJCSM 3170.01C<br>Definitions | <br>GL | <br>GL-5 | 05/2007 |
| Analysis of Alternatives Plan (AoAP) | | | | | |
| | ▪ Approved by the Milestone Decision Authority in conjunction with the Concept Decision. It details the approach to be followed in conducting the AoA during the Materiel Solution Analysis phase [previously, "Concept Refinement phase"]. See Analysis of Alternatives. | DAU Glossary of Defense Acquisition Acronyms and Terms, 12th Edition<br>Glossary | <br>Appendix B | <br>B-9 | 12/2008 |
| Analysis Plan (AP) | | | | | |
| | ▪ A capability level plan that denotes a detailed examination and application of disciplined techniques to evaluate joint mission effectiveness, system of systems performance, and joint task accomplishments. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Analytic Agenda | | | |
| | ▪ A timeline for the development of defense planning scenarios, multi-Service force deployment documents, and analytical baselines for use in strategic analyses; based upon scenario priorities identified by the Under Secretary of Defense for Policy. | CJCSI 3010.02B<br>Glossary GL GL-3 | 01/2006 |
| | ▪ The Analytic Agenda is a Department-wide cooperative agreement to make major, joint analysis efforts more effective, and responsive. It seeks to align analytical efforts with strategic decision milestones and the budget process. The Analytic Agenda includes, but is not limited to, Defense Planning Scenarios (DPS), Multi-service Force Deployment (MSFD) documents, and Analytical Baselines (AB). | J8 Force Structure Resources and Assessment; http://www.jcs.mil/j8/ddfm.html<br>Studies and Analysis Management Division --- --- | 02/2008 |
| Analytical Baseline | | | |
| | ▪ Referred to as "baseline" in the text of this Instruction. A package comprising a scenario, concept of operations, and integrated data used by the DOD Components as a foundation for strategic analyses. Analytical baselines shall be produced and reviewed in an open, collaborative, and transparent environment. | DODI 8260.01<br>Definitions E1.1. 6 | 01/2007 |
| Assumption | | | |
| | ▪ A supposition on the current situation or a presupposition on the future course of events, either or both assumed to be true in the absence of positive proof, necessary to enable the commander in the process of planning to complete an estimate of the situation and make a decision on the course of action. | JP 1-02<br>Appendix A-1 49 | 03/2007 |
| Attribute | | | |
| | ▪ A quantitative or qualitative characteristic of a system of systems that is expressed in terms of joint Doctrine, Organization, Training, Material, Leadership and Education, Personnel, and Facilities (DOTMLPF). | CTM v3.0<br>--- --- --- | 04/2009 |
| Capability | | | |
| | ▪ The ability to execute a specified course of action. (A capability may or may not be accompanied by an intention.) | JP 1-02<br>Appendix A-1 77 | 03/2007 |
| | ▪ The ability to achieve a desired effect under specified standards and conditions through combinations of means and ways across the doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) to perform a set of tasks to execute a specified course of action. It is defined by an operational user and expressed in broad operational terms in the format of an initial capabilities document or a joint DOTMLPF change recommendation. In the case of materiel proposals/documents, the definition will progressively evolve to DOTMLPF performance attributes identified in the capability development document and the capability production document. | CJCSI 3170.01G<br>Part II -- Definitions Glossary GL-3 | 03/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Capability Development Document (CDD) | | | | | |
| | ▪ A document that captures the information necessary to develop a proposed program(s), normally using an evolutionary acquisition strategy. The CDD outlines an affordable increment of militarily useful, logistically supportable, and technically mature capability. The CDD may define multiple increments if there is sufficient definition of the performance attributes (key performance parameters, key system attributes, and other attributes) to allow approval of multiple increments. | CJCSI 3170.01F<br>Glossary | <br>GL | <br>GL-5 | 05/2007 |
| Capability Evaluation Metamodel (CEM) | | | | | |
| | ▪ A conceptual model to relate key capability test and evaluation concepts. The CEM provides the "rules" for conducting Joint Mission effectiveness (JMe) assessments of capability relational structures defined in a Joint Operational Context for Test (JOC-T) and are approximated by a Joint Mission Environment (JME).<br>– Capability Test Methodology (CTM) process steps produce the following CEM structures: JOC-T; Capability Evaluation Strategy; Capability Test Design; Joint Mission Environment (JME); Test Event; and Joint Capability Evaluation (JCE).<br>– The Joint Operational Context for Test (JOC-T) is the joint operational context for the Capability Evaluation Strategy. This Capability Evaluation Strategy contains design of experiment (DOE) factors and measures which are filtered to produce various Capability Test Designs focused on one or more Critical Capability Issues (CCI).<br>– The Capability Test Design is instantiated in a test event using a JME, built from live, virtual, constructive (LVC) test technologies. Testers use the JME to execute the Capability Test Design in a test event, which provides response data for a joint capability evaluation (JCE).<br>– JCEs are conducted based on analysis structures in the Capability Test Design. Such JCEs provide SoS recommendations for DOD acquisition and other capability development managers. The CEM is based on the definition of capability and its relationships from the Joint Capabilities Integration and Development System (JCIDS) in CJCSI 3170.01F. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Capability Gap | | | | | |
| | ▪ The inability to achieve a desired effect under specified standards and conditions through combinations of means and ways to perform a set of tasks. The gap may be the result of no existing capability, lack of proficiency or sufficiency in existing capability, or the need to recapitalize an existing capability. | CJCSI 3170.01F<br>Glossary | <br>GL | <br>GL-5 | 05/2007 |
| Capability Lifecycle | | | | | |
| | ▪ Capability generation lifecycle including business practice, information flow, and their associated attributes, directed toward the efficient, synchronized delivery of required system of systems capabilities. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Capability Manager/Capability Portfolio Manager (CM/CPM) | | | | | |
| | ▪ Manages selected groupings of capabilities using integrated strategic planning, integrated architectures, measures of performance, risk management techniques, transition plans, and portfolio investment strategies. Portfolio management influences the Joint Capability Integration and Development System (JCIDS), the Planning Programming, Budgeting, and Execution process (PPBE), and the Defense Acquisition System, through the appropriate policy instructions. It delivers integrated capabilities, improves interoperability, identifies and captures efficiencies, reduces capability redundancies and gaps, and increases joint operational effectiveness. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Capability Production Document (CPD) | | | | | |
| | ▪ A document that addresses the production elements specific to a single increment of an acquisition program. | CJCSM 3170.01C<br>Glossary | GL | GL-7 | 05/2007 |
| Capability Test Methodology (CTM) | | | | | |
| | ▪ The Capability Test Methodology (CTM) is an integral part of the Defense Acquisition System, providing methods and processes that guide the design and execution of system-of-systems tests in the joint mission environment to produce high quality capability assessments and evaluations supporting Department of Defense development and investment decisions. CTM can involve developmental or operational testing during multiple phases of the acquisition lifecycle, including Materiel Solution Analysis, Technology Development, and Engineering and Manufacturing Development phases. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| Capacity | | | | | |
| | ▪ The number of instances of an object or detail that are simultaneously represented by a model or simulation; cardinality. | Recommended Practices Guide<br>Fidelity RPG Special Topic | --- | 8 | 09/2000 |
| Capstone Concept for Joint Operations (CCJO) | | | | | |
| | ▪ The Capstone Concept for Joint Operations describes in broad terms [the CJS's] vision for how the joint force circa 2016-2028 will operate in response to a wide variety of security challenges. It proposes that future joint force commanders will combine and subsequently adapt some combination of four basic categories of military activity -- combat, security, engagement, and relief and<br>reconstruction -- in accordance with the unique requirements of each operational situation. The concept is informed by current strategic guidance, but because it looks to the future, it is intended to be adaptable, as it must be, to changes in that guidance. | CCJO 3.0<br>Foreword | FW | iii | 01/2009 |
| Characteristic | | | | | |
| | ▪ A desirable trait, quality, or property that distinguishes how the future joint force should conduct military operations. | CJCSI 3010.02B<br>Glossary | GL | GL-3 | 01/2006 |
| Command-linked Tasks | | | | | |
| | ▪ Discrete activities or actions designated by a joint force commander or identified by the lead federal agency that must be performed by commands and combat support agencies outside the command or directive authority of the joint force, if the joint force is to successfully perform its missions. Command-linked tasks are selected by the supported command or lead federal agency and are normally scheduled for training, evaluated, and assessed by the organization providing the support. | CJCSM 3500.04E<br>Glossary | GL-4 | 61 | 08/2008 |
| Component | | | | | |
| | ▪ A modular, deployable, and replaceable part of a system that encapsulates implementation and exposes a set of interfaces. A component is typically specified by one or more classifiers (e.g. implementation classes) that reside on it, and may be implemented by one or more artifacts (e.g., binary, executable, or script files). | DoDAF 1.5, Volume II<br>Annex | C | C-2 | 04/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Concept of Operations (CONOPS) | | | | | |
| | ▪ A verbal or graphic statement that clearly and concisely expresses what the joint force commander intends to accomplish and how it will be done using available resources. The concept is designed to give an overall picture of the operation. Also called commander's concept or CONOPS. | JP 1-02<br>Appendix | <br>A-1 | <br>112 | 03/2007 |
| Condition | | | | | |
| | ▪ Those variables of an operational environment or situation in which a unit, system, or individual is expected to operate and may affect performance.<br>(See Joint Mission-Essential Task) | JP 1-02<br>Appendix | <br>A-1 | <br>112 | 03/2007 |
| | ▪ Variable of the operational environment, including a scenario that affects task performance. | CJCSM 3500.04E<br>Glossary | <br>GL | <br>GL-4 | 08/2008 |
| Constructive Model or Simulation | | | | | |
| | ▪ Models and simulations that involve simulated people operating simulated systems. Real people stimulate (make inputs) to such simulations, but are not involved in determining the outcomes. | DOD 5000.59-P<br>Definitions | <br>36c | <br>A-6 | 10/1995 |
| Credibility | | | | | |
| | ▪ The criteria that the model, simulation, or federation of models and simulations needs to meet to be acceptable for its intended use. | IEEE Std 1516.4-2007<br>Definitions | <br>3.1 | <br>4 | 12/2007 |
| Criterion | | | | | |
| | ▪ The minimum acceptable level of performance associated with a particular measure of task performance. It is often expressed as hours, days, percent, occurrences, minutes, miles, or some other command stated measure. | CJCSI 3500.01D<br>Glossary | <br>GL | <br>GL-4 | 05/2007 |
| Critical Capability Issue (CCI) | | | | | |
| | ▪ A Critical Capability Issue is an analytical statement used to assess performance pertaining to capabilities which support joint missions. The essential elements of a CCI include a capability's essential tasks, mission desired effects, blue system of systems (SoS) across DOTMLPF, and conditions involving threat and environmental factors. These essential elements are contained in the Capability Crosswalk.<br>– It is important to state how the test issue contributes to achieving the mission end state outcomes in terms of mission desired effects. The CCIs should address the SoS capability to perform joint operational tasks and/or the SoS, system, or service attribute performance. CCIs are of primary importance to the decision authority in reaching a decision to allow the system of systems to advance into the next phase of development. An example CCI format which captures the essential elements would be: Assess the ability to perform Task X under Conditions A by SoS Configuration Y to achieve Desired Effects Z. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Critical Operational Issue (COI) | | | | | |
| | ▪ Critical Operational Issues are the operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined in operational test and evaluation to evaluate/assess the system's capability to perform its mission. | Memorandum of Agreement on Multi-Service Operational Test And Evaluation<br>Introduction | <br>Purpose | <br>1 | 08/2004 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Critical Technical Parameter (CTP) | | | |
| | ▪ The critical technical parameters of the system (including software maturity and performance measures) that will be evaluated (or reconfirmed if previously evaluated) during the remaining phases of developmental testing. Critical technical parameters are measurable critical system characteristics that, when achieved, allow the attainment of desired operational performance capabilities. They are not user requirements. Rather, they are technical measures derived from desired user capabilities. Failure to achieve a critical technical parameter should be considered a reliable indicator that the system is behind in the planned development schedule or will likely not achieve an operational requirement. Limit the list of critical technical parameters to those that support critical capability issues. The system specification is usually a good reference for the identification of critical technical parameters. | Defense Acquisition University Guidebook<br>9 9.10.1 --- | 12/2008 |
| Data Analysis Plan (DAP) | | | |
| | ▪ The Data Analysis Plan (DAP) is a document that provides detailed procedures for the collection, reduction, collation, and analysis of data gathered to support determination of a system's/SoS's operational effectiveness and suitability. The DAP aligns with the test plan in terms of contribution to a successful test and is a planning tool to ensure procedures are in place for assessing data collection upon completion of test execution. The DAP is designed to provide the specifics for the analysis of operational effectiveness and suitability of an SoS. The DAP should be completed before the test event begins to ensure the needs of various system/SoS customers and that the resources are available to complete the capability analysis. The DAP should include the purpose of the data analysis, data sources (including a description and any limitations), key variables to be used, and the capability analysis methods. The capability manager should review the plan to ensure that the proposed capability analysis will answer relevant questions. Data analysis experts should review the plan to ensure that appropriate data and methods will be used, and the DAP should be approved by the capability manager. | CTM v3.0<br>--- --- --- | 04/2009 |
| Data Elements List Table (DELT) | | | |
| | ▪ The data elements list table (DELT ) begins the correlation between the issues and sub-issues to the measures, and data elements. The DELT will be refined and transformed into the IDRL once the units of measurement, sample size needed, data source, data media, data format, data structure, instrumentation, test variables, and individual test event have been specified. The DELT also forms the foundation for the Data Analysis Plan (DAP). | CTM v3.0<br>--- --- --- | 04/2009 |
| Data Management Plan (DMP) | | | |
| | ▪ The purpose of the Data Management Plan (DMP) is to provide detailed procedures for the collection, reduction, collation, storage, and disposition of data gathered to support determination of a system's operational effectiveness and suitability. The DMP is both a planning tool to ensure procedures are in place for data collection, and a data management tool for tracking and assessing data collection during test execution. | CTM v3.0<br>--- --- --- | 04/2009 |
| Defense Planning Scenario (DPS) | | | |
| | ▪ DPSs, written 8-20 years into the future, are used in CBA. These scenarios have classified CONOPS that provide a high level of specificity and defined parameters to aid in robust analysis of capabilities and a comparison of alternate solutions. | CJCSI 3010.02B<br>Enclosure A 7.b.1.a. A-5 | 01/2006 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Department of Defense Architecture Framework (DoDAF) | | | |
| | ▪ The DoDAF is a three-volume set that inclusively covers the concept of the architecture framework, development of architecture descriptions, and management of architecture data.<br>– Volume I introduces the DoDAF framework and addresses the development, use, governance, and maintenance of architecture data.<br>– Volume II outlines the essential aspects of architecture development and applies the net-centric concepts to the DoDAF products.<br>– Volume III introduces the architecture data management strategy and describes the pre-release CADM v1.5, which includes the data elements and business rules for the relationships that enable consistent data representation across architectures. | DoDAF 1.5, Volume I<br>Executive Summary ES ES-2 | 04/2007 |
| Developmental Test and Evaluation (DT&E) | | | |
| | ▪ Test and evaluation conducted to evaluate design approaches, validate analytical models, quantify contract technical performance and manufacturing quality measure progress in system engineering design and development, minimize design risks, predict integrated system operational performance (effectiveness and suitability) in the intended environment, and identify system problems (or deficiencies) to allow for early and timely resolution or correction. Decision-makers use DT&E results to minimize design risk, whereas OT&E evaluates military utility, and system effectiveness and suitability. DT&E usually includes contractor testing (AFPD 99-1). | AFOTEC OT&E Guide, 5th edition.<br>Glossary Attch B B-9 | 06/2007 |
| Distributed Range Coordination Team (DRCT) | | | |
| | ▪ Team representing required additions to each development and operational test organization to provide expertise for tests in joint environments and to absorb the increased scope of such testing (reference: Testing in a Joint Environment Roadmap, paragraph 2.2.9). Roles and responsibilities for team members may include providing single points of contact for program managers and lead ranges to work with multiple distributed test organizations; providing top-level facilitation for activities spanning various functional and organizational elements across distributed test organizations; and making sure distributed planning, integration, execution, and analysis activities are regularly and frequently coordinated with all participants. | CTM v3.0<br>--- --- --- | 04/2009 |
| DOTMLPF | | | |
| | ▪ Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities. (See Joint Doctrine, Joint Organization, Joint Training, Joint Materiel, Joint Leadership and Education, Joint Personnel, and Joint Facilities). | CJCSI 3170.01F<br>Glossary GL GL-11 | 05/2007 |
| Effect | | | |
| | ▪ The physical or behavioral state of a system that results from an action, a set of actions, or another effect. | JP 1-02<br>Appendix A-1 176 | 03/2007 |
| | ▪ A change to a condition, behavior, or degree of freedom. | JP 1-02<br>Appendix A-1 176 | 03/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| | ▪ The result, outcome, or consequence of an action. | JP 1-02<br>Appendix | A-1 | 176 | 03/2007 |
| Element | | | | | |
| | ▪ An atomic constituent of a model. | DoDAF 1.5, Volume II<br>Annex | C | C-3 | 04/2007 |
| End State | | | | | |
| | ▪ The set of required conditions that defines achievement of the commander's objectives. | JP 1-02<br>Appendix | A-1 | 183 | 03/2007 |
| | ▪ The set of conditions, behaviors, and freedoms that defines achievement of the commander's mission. | CJCSI 3010.02B<br>Glossary | GL | GL-4 | 01/2006 |
| Enduring/Persistent | | | | | |
| | ▪ This has both a mental and physical aspect. The mental aspect can be expressed as will, while the physical aspect can be expressed as the staying power of the joint force--in both cases, sustaining ours while breaking the adversaries. This characteristic is especially important given the interaction between the anticipated environment, joint force Operations, and unanticipated events in any complex and adaptive system. It demands that the joint force possess the depth and capacity to sustain operations over time, regardless of the situation or adversary.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | 4.E.6. | 22 | 08/2005 |
| Engineering and Manufacturing Development (EMD) | | | | | |
| | ▪ The purpose of the EMD Phase is to develop a system or an increment of capability; complete full system integration (technology risk reduction occurs during Technology Development); develop an affordable and executable manufacturing process; ensure operational supportability with particular attention to minimizing the logistics footprint; implement human systems integration (HSI); design for producibility; ensure affordability; protect CPI by implementing appropriate techniques such as anti-tamper; and demonstrate system integration, interoperability, safety, and utility. The CDD, Acquisition Strategy, SEP, and Test and Evaluation Master Plan (TEMP) shall guide this effort. | DODI 5000.2<br>Enclosure 2 | Procedures | 20 | 12/2008 |
| Environment | | | | | |
| | ▪ Includes the air, water, land, plants, animals, and other living organisms, man-made structures, historical and cultural resources, and the interrelationships that exist among them and with people. | DAU Glossary of Defense Acquisition Acronyms and Terms, 12th Edition<br>Glossary | GL | B-56 | 12/2008 |
| | ▪ The aggregate of all external and internal conditions (such as temperature, humidity, radiation, magnetic and electric fields, shock vibration, etc.) either natural or man - made, or self - induced, that influences the form, performance, reliability or survival of an item. | DAU Glossary of Defense Acquisition Acronyms and Terms, 12th Edition<br>Glossary | GL | B-56 | 12/2008 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Environmental Condition | | | |
| | ▪ Those physical environment (land, sea, air, and space) condition variables of an operational environment or situation in which a unit, system, or individual is expected to operate. Those civil environment (political, cultural, and economic) condition variables of an operational environment or situation in which a unit, system, or individual is expected to operate and may affect performance. | CTM v3.0<br>--- --- --- | 04/2009 |
| Error | | | |
| | ▪ The difference between an observed, measured, or calculated value and a correct value. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| Essential Task | | | |
| | ▪ In the context of joint operation planning, a specified or implied task that an organization must perform to accomplish the mission. An essential task is typically included in the mission statement.<br>(See Task) | JP 1-02<br>Appendix A-1 187 | 03/2007 |
| | ▪ Tasks based on mission analysis and approved by the commander that are absolutely necessary, indispensable, or critical to the success of a mission. | CJCSM 3500.04E<br>Glossary GL 61 | 08/2008 |
| Evaluation Strategy | | | |
| | ▪ The evaluation strategy serves as the blueprint to assess a capability's joint mission effectiveness (JMe). Key elements of the evaluation strategy are Critical Capability Issues (CCI), evaluation independent factors, and evaluation dependent response measures.<br>– Evaluation independent factors include: joint mission(s) and task(s); threat and environmental conditions; and system of systems (SoS) configuration options across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) resources.<br>– Evaluation dependent response measures are structured in three levels: mission measures of effectiveness (MMOEs), task measures of performance (TMOPs), and system/SoS attributes. These measures should be described in terms of their nature (e.g., qualitative and quantitative), measurement units, and desired fidelity.<br>– Evaluation strategy elements can be refined and related using the Capability Crosswalk structure. Using the elements of the Joint Operational Context for Test (JOC-T), the evaluation strategy is created and refined to support Joint Mission Effectiveness (JMe) evaluation of a capability's joint task performance by a system of systems configuration (across DOTMLPF) under threat and environmental condition sets to achieve mission desired effects in a realistic joint environment. | CTM v3.0<br>--- --- --- | 04/2009 |
| Evaluation View (EV) | | | |
| | ▪ A proposed view in the Department of Defense Architecture Framework (DoDAF) designed to capture the evaluation framework for assessing joint mission effectiveness, joint task performance, and system of systems performance. The Evaluation View(s) would include mission desired effects, mission measures of effectiveness, task measures of performance, system of systems attributes, and performance measures, and all associated data. | CTM v3.0<br>--- --- --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Event Management Plan (EMP) | | | | | |
| | ▪ The Event Management Plan includes three sections. These three sections are the event schedule, the data management plan, and the coordinated event support. These sections are generally done by the test range facility in coordination with the customer and are specific to a test event and its iterations. The three sections outlined are the minimum items that should be included in the plan and coordinated before an event is run. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Event Manager (EM) | | | | | |
| | ▪ Responsible for planning, coordinating, and executing an LVC-DE event supporting Capability Manager and Program Manager(s) requirements for Capability Test & Evaluation. In addition to coordination of event support, the EM develops and manages an integrated schedule and a data management plan, both addressing requirements from the CM(s) and PM(s). The EM function is generally done by the lead test range facility for a specific test event and its iterations. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Expeditionary | | | | | |
| | ▪ An expeditionary joint force is organized, postured and capable of rapid and simultaneous deployment, employment, and sustainment. Implicit in this is a joint force that converges mission-tailored capabilities at the desired point of action from dispersed locations around the globe, regardless of anti-access or area-denial environments. As elusive and adaptive adversaries seek refuge in remote and inaccessible areas, the norm will be short-notice operations, austere operational environments, incomplete information and the requirement to fight on arrival throughout the battlespace and to dominate potential adversaries for the duration of a campaign. The future joint force will be immediately employable even in austere conditions and largely independent of existing infrastructure. As a situation evolves, these elements will be readily capable of transitioning to sustained operations, blending into new capability packages to execute follow-on or different operations, or dispersing until otherwise required. The term "expeditionary" also describes the joint force mindset that inculcates an expeditionary perspective into all aspects of force planning, training, and education. The future joint force will increasingly require a mechanism to enable global sourcing of military forces and capabilities; in order to leverage the most responsive, best positioned forces at the time of need.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | <br>4.E.4. | <br>21 | 08/2005 |
| Fast | | | | | |
| | ▪ Key to effectively controlling tempo is the ability to be faster than the adversary or situational events. The speed at which forces maneuver and engage, or decisions are made, or relief is provided, will largely determine operational successes or failures. Successfully overcoming future challenges may require speed of action across all domains. Acting fast is in itself a force multiplier and often a requisite for the effective application of military capabilities.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | <br>4.E.8. | <br>23 | 08/2005 |
| Federate | | | | | |
| | ▪ An application that may be or is currently coupled with other software applications under a Federation Object Model Document Data/Federation Execution Data (FDD/FED) and a runtime infrastructure (RTI). This may include federation managers, data collectors, real world ("live") systems (e.g., C4I systems, instrumented ranges, sensors), simulations, passive viewers, and other utilities. | IEEE Std 1516.4-2007<br>Definitions | <br>3.1 | <br>4 | 12/2007 |
| Federation | | | | | |
| | ▪ A named set of federate applications and a common Federation Object Model (FOM) that are used as a whole to achieve some specific objective. | IEEE Std 1516.4-2007<br>Definitions | <br>3.1 | <br>4 | 12/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Fidelity | | | |
| | ▪ The description of a model, simulation, or federation of models and simulations and its associated data representational capabilities (e.g. resolution, error, precision, and sensitivity). | IEEE Std 1516.4-2007<br>Special Terms 3.2 5 | 12/2007 |
| Fitness | | | |
| | ▪ Providing the capabilities needed or being suitable for some purpose, function, situation or application. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| Implied Task | | | |
| | ▪ A task that is not stated but necessary to do the mission. | CJCSM 3500.04E<br>Glossary GL 61 | 08/2008 |
| Increment | | | |
| | ▪ A militarily useful and supportable operational capability that can be effectively developed, produced or acquired, deployed, and sustained. Each increment of capability will have its own set of threshold and objective values set by the user. Spiral development is an instance of an incremental development strategy where the end state is unknown. Technology is developed to a desired maturity and injected into the delivery of an increment of capability. | CJCSM 3170.01C<br>Glossary GL GL-9 | 05/2007 |
| Initial Capabilities Document (ICD) | | | |
| | ▪ Documents the requirement for a materiel or non-materiel approach, or an approach that is a combination of materiel and non-materiel, to satisfy specific capability gap(s). It defines the capability gap(s) in terms of the functional area, the relevant range of military operations, desired effects, time and doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) and policy implications and constraints. The ICD summarizes the results of the DOTMLPF and policy analysis and the DOTMLPF approaches (materiel and non-materiel) that may deliver the required capability. The outcome of an ICD could be one or more joint DCRs or capability development documents. | CJCSM 3170.01C<br>Glossary GL GL-10 | 05/2007 |
| Integrated Capability Portfolio (ICP) | | | |
| | ▪ Executive Level Management of capability groupings that cover the entire DOD budget authority. | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>Definitions --- 9 | 04/2007 |
| Integrated Data Requirements List (IDRL) | | | |
| | ▪ Serving as the foundation for the Data Analysis Plan (DAP), the IDRL correlates the issues to the sub-issues, measures, data elements, units of measurement, sample size, data source, data media, data format, data structure, instrumentation, test variables, and test event. | CTM v3.0<br>--- --- --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Integrated Priority List (IPL) | | | |
| | ▪ A list of a combatant commander's highest priority requirements, prioritized across Service and functional lines, defining shortfalls in key programs that in the judgment of the combatant commander, adversely affect the capability of the combatant commander's forces to accomplish their assigned mission. The integrated priority list provides the combatant commanders' recommendations for programming funds in the planning, programming, and budgeting system process. | JP 1-02<br>Definitions --- 266 | 03/2007 |
| Interaction | | | |
| | ▪ A specification of how stimuli are sent between instances to perform a specific task. The interaction is defined in the context of a collaboration. | DoDAF 1.5, Volume II<br>Annex C C-3 | 04/2007 |
| Interagency Coordination | | | |
| | ▪ Within the context of Department of Defense involvement, the coordination that occurs between elements of the Department of Defense and engaged US government agencies, for the purpose of achieving an objective. | CJCSM 3500.04E<br>Glossary GL 62 | 08/2008 |
| Interoperable | | | |
| | ▪ Interoperability is a necessary prerequisite to integrated and interdependent joint operations. The future joint force will be able to share and exchange knowledge and services between units and commands at all levels. The interoperable joint force can act in an integrated and ultimately an interdependent way among joint force components and capabilities, facilitating more effective interoperability with interagency and multinational partners. Interoperability implies systems, capabilities and organizations working in harmony across all joint force elements; however, it involves more than systems and equipment. Interoperability includes a cultural change at all levels that extends through DOTMLPF.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution 4.E.3. 21 | 08/2005 |
| JCA Decomposition | | | |
| | ▪ JCAs are logically broken down from higher capability categories to further scope, bound, and clarify capability categories by providing greater granularity to facilitate detailed analysis or allow better mapping of resources to capabilities. The number of tiers/levels required to decompose a JCA down to its component capabilities is not a constant across the JCAs. This decomposition enhances JCA usefulness in DOD processes, (e.g., Integrated Priority List (IPL) submissions, Universal Joint Task List (UJTL) integration, roadmaps, and program and budget databases). | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>Definitions --- 9 | 04/2007 |
| JCA Lexicon | | | |
| | ▪ A collection of joint capability definitions that provide a common capabilities language for DOD in order to facilitate capabilities-based planning, analysis, and decision-making. (Modified from Joint Capability Area Management Plan). | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>Definitions --- 9 | 04/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| JCA Taxonomy | | | |
| | ▪ The structure or framework of joint capabilities, used in conjunction with the JCA lexicon, to facilitate capabilities-based planning, analysis, and decision-making. | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>Definitions --- 9 | 04/2007 |
| JFM Element | | | |
| | ▪ A member of a component. Elements contain the base attribute or operations, and provide the basis for instantiated entities in the JME. [Model Driven Architecture (MDA) References]. | CTM v3.0<br>--- --- --- | 04/2009 |
| Joint | | | |
| | ▪ Connotes activities, operations, organizations, etc., in which elements of two or more Military Departments participate. | JP 1-02<br>Appendix A-1 281 | 03/2007 |
| Joint Capabilities Document (JCD) | | | |
| | ▪ The JCD identifies a set of capabilities that support a defined mission area utilizing associated Joint Operations Concepts (JOpsC), concept of operations (CONOPs), or Unified Command Plan or other assigned missions. The capabilities are identified by analyzing what is required across all functional areas to accomplish the mission. The gaps or redundancies are then identified by comparing the capability needs to the capabilities provided by existing or planned systems. The JCD will be used as a baseline for one or more functional solution analyses leading to the appropriate initial capabilities documents or doctrine, organization, training, materiel, leadership and education, personnel, and facilities change recommendation documents, but cannot be used for the development of capability development or capability production documents. The JCD will be updated as changes are made to the supported JOpsC, CONOPs, or assigned missions. | CJCSM 3170.01C<br>Glossary GL GL-12 | 05/2007 |
| Joint Capabilities Evaluation (JCE) | | | |
| | ▪ The documented analysis of one or more capability test events used to support milestone A, B, or C acquisition decisions. | CTM v3.0<br>--- --- --- | 04/2009 |
| Joint Capabilities Integration and Development System (JCIDS) | | | |
| | ▪ The Joint Capabilities Integration and Development System (JCIDS) is a joint-concepts-centric capabilities identification process that allows joint forces to meet future military challenges. The Joint Capabilities Integration and Development System process assesses existing and proposed capabilities in light of their contribution to future joint concepts. Joint Capabilities Integration and Development System, supported by robust analytic processes, identifies capability gaps and potential solutions. | Defense Acquisition University Guidebook<br>1 1.3 --- | 12/2008 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Joint Capability Area (JCA) | | | |
| | ▪ Collections of like DOD activities functionally grouped to support capability analysis, strategy development, investment decision making, capability portfolio management, and capabilities-based force development and operational planning. | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>--- --- --- | 04/2007 |
| Joint Doctrine | | | |
| | ▪ Joint doctrine consists of fundamental principles that guide the employment of US military forces in coordinated action toward a common objective. Joint doctrine contained in joint publications also includes terms, tactics, techniques, and procedures.<br>(See DOTMLPF) | CJCSI 5120.02<br>Enclosure A 1.a A.1 | 11/2004 |
| Joint DOTMLPF Change Recommendation (DCR) | | | |
| | ▪ A recommendation for changes to existing joint resources when such changes are not associated with a new defense acquisition program. | CJCSI 3170.01F<br>Glossary GL GL-11 | 05/2007 |
| Joint Environment (JE) | | | |
| | ▪ Realistic test environment comprised of friendly forces and equipment, threats, and geophysical environments that are required to assess military capabilities that are 'born joint' as identified in JCIDS capability documents. | CTM v3.0; derived from "Testing in a Joint Environment (TIJE) Roadmap"<br>2.0 2.2.3 10 | 11/2004 |
| Joint Exercise | | | |
| | ▪ A joint military maneuver, simulated wartime operation, or other CJCS- or combatant commander-designated event involving planning, preparation, execution, and evaluation. A joint exercise involves forces of two or more Military Departments interacting with a combatant commander or subordinate joint force commander, involves joint forces and/or joint staffs, and is conducted using joint doctrine or joint tactics, techniques, and procedures. | CJCSM 3500.04E<br>Glossary GL-5 62 | 08/2008 |
| Joint Facilities | | | |
| | ▪ Real property consisting of one or more of the following: a building, a structure, a utility system, pavement, and underlying land. Key facilities are selected command installations and industrial facilities of primary importance to the support of military operations or military production programs. A key facilities list is prepared under the policy direction of the Joint Chiefs of Staff.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Glossary GL GL-13 | 05/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Joint Force Characteristics | | | | | |
| | ▪ The joint force must have certain key characteristics. These particular characteristics are considered important because they will guide how the joint force is developed, organized, trained and equipped and must be reflected in all subordinate concepts in the JOpsC family. Such a force is designed to be a dominant national asset, compelling in all situations, and lethal when required.<br>– Joint Force Characteristics include: Knowledge Empowered, Networked, Interoperable, Expeditionary, Adaptable/Tailorable, Enduring/Persistent, Precise, Fast, Resilient, Agile, and Lethal. | CCJO 2.0<br>4. Solution | <br>4.E. | <br>20 | 08/2005 |
| Joint Functional Concept (JFC) | | | | | |
| | ▪ Addresses broad enduring functions across the range of military operations (e.g., force application and battlespace awareness). | CCJO 2.0<br>2. Scope | <br>2.B. | <br>3 | 08/2005 |
| Joint Functions | | | | | |
| | ▪ Related capabilities and activities grouped together to help joint force commanders synchronize, integrate, and direct joint operations. Functions that are common to joint operations at all levels of war fall into six basic groups—command and control, intelligence, fires, movement and maneuver, protection, and sustainment. | JP3-0<br>Glossary | <br>GL | <br>GL-20 | 09/2006 |
| Joint Integrating Concept (JIC) | | | | | |
| | ▪ A JIC is an operational-level description of how a joint force commander, 8-20 years into the future, will perform a specific operation or function derived from a JOC and/or a JFC. JICs are narrowly scoped to identify, describe, and apply specific military capabilities, decomposing them into fundamental tasks, conditions, and standards. Further analysis and expansion of tasks, conditions, and standards is accomplished after JIC completion in order to effectively execute CBA. Additionally, a JIC contains illustrative vignettes to facilitate understanding of the concept. | CJCSI 3010.02B<br>Enclosure A | <br>A-3 | <br>13 | 01/2006 |
| Joint Leadership and Education | | | | | |
| | ▪ Professional development of the joint commander is the product of a learning continuum that comprises training, experience, education, and self-improvement. The role of Professional Military Education and Joint Professional Military Education is to provide the education needed to complement training, experience, and self-improvement to produce the most professionally competent individual possible.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-13 | 05/2007 |
| Joint Materiel | | | | | |
| | ▪ All items (including ships, tanks, self-propelled weapons, aircraft, etc., and related spares, repair parts, and support equipment, but excluding real property, installations, and utilities) necessary to equip, operate, maintain, and support [joint] military activities without distinction as to its application for administrative or combat purposes.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-13 | 05/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Joint Mission Effectiveness (JMe) | | | |
| | ▪ Joint Mission Effectiveness (JMe) is the evaluation of a capability's joint task performance by a system of systems configuration (across DOTMLPF) under threat and environmental condition sets to achieve mission desired effects in a joint operational context for test (JOC-T).<br>– Critical Capability Issues (CCI) are used to capture essential JMe elements in terms of an evaluation question (e.g., Can the Capability perform Task X by SoS Configuration Y under Condition Set A to achieve Mission Desired Effect Z?<br>– JMe follows a traditional scientific methodology of empirical-inductive reasoning to evaluate causal relationships between capabilities and increased warfighting effectiveness. The scientific method employs a basic experimental design process to determine if a proposed capability A causes the anticipated military effect B. This can be stated in terms of an experimental hypothesis, "If the proposed capability, then an improved mission effectiveness". These relational concepts are reflected in a Capability Evaluation Metamodel (CEM), an underlying conceptual model that supports the Capability Test Methodology to frame the evaluation of Joint Mission Effectiveness (JMe).<br>– A design of experiment (DOE) approach is used in the CEM to frame capability test designs in terms of independent variables (IVs), the causal condition A, and dependent variables (DVs), the effect B. CEM IVs are manipulated factors in the test whose presence or degree affects change in dependent variables.<br>– There are three IV treatment dimensions in a CEM test design: joint mission(s) and task(s); threat and environmental conditions; and system of systems (SoS) configuration options across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) resources. CEM DVs are response measures, whose changes are caused by the presence, or degree of IVs in the test. DVs are measured for increases and decreases in mission effectiveness across a set of test trials.<br>– There are three levels of response measures: mission measures of effectiveness (mission MOE), task measure of performance (task MOP), and system or SoS attribute performance (system/SoS attribute). | CTM v3.0<br>--- --- --- | 04/2009 |
| Joint Mission Environment (JME) | | | |
| | ▪ A subset of the joint operational environment composed of force and non-force entities; conditions, circumstances and influences within which forces employ capabilities to execute joint tasks to meet a specific mission objective. | TSSG Approved<br>--- --- --- | 06/2008 |
| Joint Mission Environment Foundation Model (JFM) | | | |
| | ▪ The purpose of the JME Foundation Model (JFM) is to provide an authoritative framework for applying a logical capabilities-based process that can be robustly applied for reasoning among Stakeholders in a wide range of situations and test capability applications. The JFM is a design template for the CTM system engineering M&P that can be used to guide the reuse and development of LVC-DE systems. The JFM is a theoretical construct that represents physical processes, with a set of logical and quantitative relationships between those components, and component interactions. The goal of the JFM is to provide a frame of reference for LVC-DE configuration design. | CTM v3.0<br>--- --- --- | 04/2009 |
| Joint Mission Essential Task (JMET) | | | |
| | ▪ A mission task selected by a joint force commander deemed essential to mission accomplishment and defined using the common language of the universal joint task list in terms of task, condition, and standard. Also called JMET. | CJCSI 3500.01D<br>Glossary GL GL-6 | 05/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Joint Mission Essential Task List (JMETL) | | | | | |
| | ▪ A mission task selected by a joint force commander deemed essential to mission accomplishment and defined using the common language of the universal joint task list in terms of task, condition, and standard. Also called JMET. | CJCSM 3500.04E<br>Glossary | GL | GL-5 | 08/2008 |
| Joint Operating Concept (JOC) | | | | | |
| | ▪ Individual joint operating concepts will address the joint contribution to dealing with each of these [five] challenges [1. win the Nation's wars; 2. deter potential adversaries; 3. develop cooperative security; 4. defend the homeland; and 5. respond to civil crises] in greater detail. | CCJO 3.0<br>3. National Security Challenges | --- | 7 | 01/2009 |
| Joint Operating Environment | | | | | |
| | ▪ The joint operating environment is the environment of land, sea, and/or airspace within which a joint force commander employs capabilities to execute assigned missions. It is the broad area of operations and key features of that area where a joint force commander is expected to operate. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| Joint Operational Context for Test (JOC-T) | | | | | |
| | ▪ The JOC-T is the appropriate combination of representative systems, forces, threats, and environmental conditions assembled for test in a Joint Mission Environment. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| | ▪ Alternately, it is the comprehensive description of the mission, forces, environment, and TTPs – and the dependencies among these – that must be addressed in the test environment. It includes a description of the resources, live, virtual, or constructive, that will be employed to create this environment for the purposes of testing. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| | ▪ The JOC-T describes capability concepts and relationships, as defined in JCIDS, including mission, task, condition, and system of systems (SoS). JOC-T mission aspects include the mission statement, mission desired effects, and mission end state. JOC-T task aspects include mission concept of operations (CONOPs), Blue force UJTL-based JMETs, Service tasks, and tactics, techniques, and procedures (TTP). JOC-T condition aspects include threat conditions (e.g., threat actions, threat order of battle, threat command and control structure, threat systems, threat force laydown), and environmental conditions (e.g., physical and civil environment). JOC-T Blue SoS aspects include joint capability area (JCA) operational functions and DOTMLPF materiel and non-materiel resource descriptions across DOTMLPF. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| Joint Operational Environment | | | | | |
| | ▪ Joint operational environment is defined as a composite of the conditions, circumstances and influences that affect the employment of capabilities and bear on the decisions of the commander. It includes:<br>– Physical areas and factors (of the air, land, sea, and space domains).<br>– The information environment.<br>– Adversary, friendly, and neutral systems relevant to a specific joint operation. | CTM v3.0<br>--- | --- | --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Joint Operations Concepts (JOpsC) | | | |
| | ▪ JOpsC is a family of joint future concepts consisting of a Capstone Concept for Joint Operations, Joint Operating Concepts, Joint Functional Concepts, and Joint Integrating Concepts. They are a visualization of future operations and describe how a commander, using military art and science, might employ capabilities necessary to meet successfully challenges 8 to 20 years in the future. Ideally, they will produce military capabilities that render previous ways of warfighting obsolete and may significantly change the measures of success in military operations overall. JOpsC presents a detailed description of "how" future operations may be conducted and provides the conceptual basis for joint experimentation and capabilities-based assessments (CBAs). The outcomes of experimentation and CBA will underpin investment decisions leading to the development of new military capabilities beyond the Future Years Defense Program. | CJCSM 3170.01C<br>Glossary GL GL-13 | 05/2007 |
| Joint Organization | | | |
| | ▪ A joint unit or element with varied functions enabled by a structure through which individuals cooperate systematically to accomplish a common mission and directly provide or support [joint] warfighting capabilities. Subordinate units/elements coordinate with other units/elements and, as a whole, enable the higher-level joint unit/element to accomplish its mission. This includes the joint manpower (military, civilian, and contractor support) required to operate, sustain, and reconstitute joint warfighting capabilities.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Enclosure H (2) H-2 | 05/2007 |
| Joint Personnel | | | |
| | ▪ The personnel component primarily ensures that qualified personnel exist to support joint capabilities. This is accomplished through synchronized efforts of joint force commanders and Service components to optimize personnel support to the joint force to ensure success of ongoing peacetime, contingency, and wartime operations.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Enclosure (6) H-3 | 05/2007 |
| Joint Professional Military Education (JPME) | | | |
| | ▪ A CJCS-approved body of objectives, outcomes, policies, procedures, and standards supporting educational requirements of joint officer management. | CJCSM 3500.04E<br>Glossary GL-5 62 | 08/2008 |
| Joint Training | | | |
| | ▪ Training, including mission rehearsals, of individuals, units, and staffs using joint doctrine or joint tactics, techniques, and procedures to prepare joint forces or joint staffs to respond to strategic, operational, or tactical requirements considered necessary by the combatant commanders to execute their assigned or anticipated missions.<br>(See DOTMLPF) | CJCSM 3170.01C<br>Glossary GL GL-12 | 05/2007 |
| Key Performance Parameter (KPP) | | | |
| | ▪ Those attributes or characteristics of a system that are considered critical or essential to the development of an effective military capability and those attributes that make a significant contribution to the characteristics of the future joint force as defined in the Capstone Concept for Joint Operations. KPPs must be testable to enable feedback from test and evaluation efforts to the requirements process. KPPs are validated by the Joint Requirement Oversight Council (JROC) for JROC Interest documents, and by the DOD Component for Joint Integration, Joint Information, or Independent documents. Capability development and capability production document KPPs are included verbatim in the acquisition program baseline. | CJCSM 3170.01C<br>Glossary GL GL-16 | 05/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Key System Attribute (KSA) | | | | | |
| | ▪ An attribute or characteristic considered crucial in support of achieving a balanced solution/approach to a key performance parameter (KPP) or some other key performance attribute deemed necessary by the sponsor. KSAs provide decision makers with an additional level of capability performance characteristics below the KPP level and require a sponsor 4-star, Defense agency commander, or Principal Staff Assistant to change. | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-16 | 05/2007 |
| Knowledge Empowered | | | | | |
| | ▪ The future joint force will emphasize better decisions made faster throughout all levels of command. The fundamentals of this knowledge empowerment are experienced and empowered decision makers benefiting from an enhanced understanding of the environment, potential adversaries and cultures, as well as enhanced collaborative decision-making processes. Although we will never eliminate the fog of war, an increased level of understanding should empower leaders throughout the joint force. This will enable them to anticipate and act as opportunities are presented, apply innovative solutions, mitigate risk, and increase the pace, coherence, and effectiveness of operations even in complex environments. A knowledge empowered force, capable of effective information sharing across all agencies and partners, will be able to make better decisions quicker, increasing joint force effectiveness.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | <br>4.E.1. | <br>21 | 08/2005 |
| Lethal | | | | | |
| | ▪ This is the ability to destroy an adversary and/or his systems in all conditions and environments when required. It includes the use of kinetic and/or non-kinetic means, while leveraging technological advances in greater precision and more devastating target effects at both longer-ranges and in close combat.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | <br>4.E.11. | <br>23 | 08/2005 |
| Live, Virtual, Constructive Distributed Environment (LVC-DE) | | | | | |
| | ▪ The enterprise capability necessary to accurately and realistically test systems and systems of systems, and/or train individuals, units, and organizations, performing tasks in a Joint Operational Context. It is achieved when all required joint systems, personnel, and equipment to execute the task in real-world operations are present or accurately replicated, realistically exercised or tested, and evaluated. The LVC-DE is defined using non-materiel aspects and materiel aspects across the enterprise's Doctrine (business practice), Organization, Training, Materiel, Leadership & Education, Personnel, and Facilities (DOTMLPF). | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Logical Design | | | | | |
| | ▪ The logical design provides a system-level viewpoint of the LVC system component types, descriptions of the roles these components serve, and how they are intended to work together. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| M&P Effectiveness | | | |
| | ▪ Within the context of the JTEM JT&E, effectiveness is how well JTEM-developed Capability Test Methodology (CTM) overall outputs satisfy end customer requirements, how well the outputs of every CTM process meet the input requirements of internal customers, and how well the inputs from CTM suppliers meet the requirements of the methodology. JTEM effectiveness specifically addresses (1) the extent to which the goals of the method and processes are attained for designing and executing system-of-systems tests in the JME; and (2) the extent to which the goals of the methods and processes are achieved for assessing performance pertaining to capabilities supporting joint missions. JTEM JT&E effectiveness metrics assess the following criteria areas:<br>– Usability (product template and instruction/guidebook usefulness, helpfulness, and simplicity).<br>– Consistency (alignment of product and processes structures within the M&P, and between relevant external DOD/service M&P).<br>– Workflow (leanness of process sequencing, product input/output mappings).<br>– Completeness (the sufficiency of JTEM M&P input/output products and processes to address customer needs).<br>– Adaptability (how well M&P adapts to different enterprises, changing environments, compressed deliverable timelines, etc.).<br>– Repeatability (the degree to which different groups of JTEM M&P users demonstrate similar actions and produce similar output products).<br>– Timeliness (the latency of performing JTEM processes or the amount of processes performed in a work period). | CTM v3.0<br>--- --- --- | 04/2009 |
| M&P Suitability | | | |
| | ▪ Within the context of JTEM JT&E, suitability is the degree to which JTEM M&P can be efficiently implemented and sustained in a Live, Virtual, Constructive (LVC) distributed range enterprise tasked with designing and executing system-of-systems tests in a JME. JTEM M&P suitability evaluation includes resource utilization, minimization, interoperability, and reuse across non-materiel and materiel criteria areas including:<br>– Doctrine translates to M&P and policy related to JTEM (relates to issues concerning compatibility with current T&E M&P/policy, interoperability with external DOD domain M&P/policy including acquisition, JCIDS, training, and experimentation).<br>– Organization (relates to migration/extensions from current T&E organizations and organization-policy change requirement issues).<br>– Training (relates to M&P and external DOD business practice training issues).<br>– Materiel (relates to M&P supporting materiel (hardware, software) and LVC distributed range materiel issues).<br>– Leadership and education (relates to M&P leadership/governance change requirement issues, M&P personnel educational foundation issues). | CTM v3.0<br>--- --- --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| | – Personnel (relates to M&P personnel availability issues and personnel-organization change issues).<br>– Facilities (relates to M&P supporting facility and LVC distributed range facility issues). | | | | |
| Materiel Solution | | | | | |
| | ▪ Correction of a deficiency, satisfaction of a capability gap, or incorporation of new technology that results in the development, acquisition, procurement, or fielding of a new item (including ships, tanks, self-propelled weapons, aircraft, etc., and related software, spares, repair parts, and support equipment, but excluding real property, installations, and utilities) necessary to equip, operate, maintain, and support military activities without disruption as to its application for administrative or combat purposes. In the case of family of systems and system of systems approaches, an individual materiel solution may not fully satisfy a necessary capability gap on its own. | CJCSI 3170.01F<br>Glossary | <br>GL | <br>GL-15 | 05/2007 |
| Materiel Solution Analysis Phase (MSAP) | | | | | |
| | ▪ The purpose of this phase is to assess potential materiel solutions and to satisfy the phase-specific entrance criteria for the next program milestone designated by the MDA. | DODI 5000.2<br>Enclosure 2 | <br>Procedures | <br>14 | 12/2008 |
| Means | | | | | |
| | ▪ Forces, units, equipment, and resources. | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>--- | <br>--- | <br>--- | 04/2007 |
| | ▪ Applied to Testing in a Joint Environment, Means are Organization, Materiel, and Facility Resources required to instantiate a Joint Mission Environment (JME) System of Systems (SoS). | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Measure | | | | | |
| | ▪ A parameter that provides the basis for describing varying levels of task accomplishment. | CJCSM 3500.04E<br>Glossary | <br>GL-5 | <br>62 | 08/2008 |
| Measure of Effectiveness (MOE) | | | | | |
| | ▪ Measures designed to correspond to accomplishment of mission objectives and achievement of desired effects. | CJCSI 3170.01F<br>Glossary | <br>GL | <br>GL-15 | 05/2007 |
| | ▪ A criterion used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect. | JP 1-02<br>Appendix | <br>A-1 | <br>333 | 03/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Measure of Performance (MOP) | | | | | |
| | ▪ A criterion used to assess friendly actions that is tied to measuring task accomplishment. Also called MOP. | JP 1-02<br>Appendix | <br>A-1 | <br>333 | 03/2007 |
| Meteorological and Oceanographic (METOC) | | | | | |
| | ▪ A term used to convey all meteorological (weather) and oceanographic (physical oceanography) factors as provided by Service components. These factors include the whole range of atmospheric and oceanographic phenomena, from the sub-bottom of the earth's oceans up to the space environment (space weather). | JP 1-02<br>Appendix | <br>A-1 | <br>336 | 03/2007 |
| Methods and Processes (M&P) | | | | | |
| | ▪ Within the context of JTEM JT&E, methods and processes involve defining the DOD Capability Test Methodology (CTM) business practice doctrine, including guidance on processes, information products, and their associated dynamics. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Military Training | | | | | |
| | ▪ The instruction of personnel to enhance their capacity to perform specific military functions and tasks. | CJCSM 3500.04E<br>Glossary | <br>GL-6 | <br>63 | 08/2008 |
| | ▪ The exercise of one or more military units conducted to enhance combat readiness. | CJCSM 3500.04E<br>Glossary | <br>GL-6 | <br>63 | 08/2008 |
| Mission | | | | | |
| | ▪ The task, together with the purpose, that clearly indicates the action to be taken and the reason therefore. | JP 1-02<br>Appendix | <br>A-1 | <br>349 | 03/2007 |
| | ▪ In common usage, especially when applied to lower military units, a duty assigned to an individual or unit; a task. | JP 1-02<br>Appendix | <br>A-1 | <br>349 | 03/2007 |
| | ▪ The dispatching of one or more aircraft to accomplish one particular task. | JP 1-02<br>Appendix | <br>A-1 | <br>349 | 03/2007 |
| Mission Measures of Effectiveness (MMOE) | | | | | |
| | ▪ Those measures that evaluate achievement of desired mission end state outcomes in terms of mission desired effects. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Model | | | | | |
| | ▪ A semantically complete abstraction of a system. | DoDAF 1.5, Volume II<br>Annex | <br>C | <br>C-4 | 04/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Multi-Service Force Deployment (MSFD) | | | | | |
| | ▪ Multi-Service Force Deployment (MSFD) products are DOD-approved theater campaign sets of ally and threat scenario data describing the full spectrum of conflict for future postulated scenarios outlined in the Defense Planning Scenarios. MSFDs consist of joint service coordinated, D-Day, H-Hour scenario depictions of forecasted and mobilized opposing air, land, sea and space forces in total battlespace environments.<br>– They will provide standard threat baselines used by the acquisition, operational test and evaluation, war gaming, and joint modeling communities to accomplish studies and analysis under simulated combat conditions and are also used for DoD policy studies (e.g., the Quadrennial Defense Review). The MSFD products consist of a CONOPS, a 200-page description of how the postulated conflict unfolds to include 1) database, and 2) Orders of Battle, strategy/tactics at the operational level, axes of attack, defensive dispositions, TOEs, force allocation to missions, optempo/sortie rates, readiness factors, munitions, and sustainment. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Networked | | | | | |
| | ▪ All joint force elements will be connected and synchronized in time and purpose to facilitate integrated and interdependent operations across the global battlespace. A networked joint force can extend the benefits of decentralization--initiative, adaptability, and increased tempo--without sacrificing the coordination or unity of effort emblematic of centralization. The joint force will capitalize on being networked by making user-defined information and expertise available anywhere within the network, and will exploit network connectivity among dispersed joint force elements to improve information sharing, collaboration, coordinated maneuver, and integrated situational awareness. Networks should extend to interagency and multinational partners, where possible, to support and enhance unified action.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution | <br>4.E.2. | <br>21 | 08/2005 |
| Node | | | | | |
| | ▪ A representation of an element of architecture that produces, consumes, or processes data. | DoDAF 1.5, Volume II<br>Annex B | <br>--- | <br>B-5 | 04/2007 |
| Non-Materiel Solution | | | | | |
| | ▪ Changes in doctrine, organization, training, materiel, leadership and education, personnel, facilities, or policy (including all human systems integration domains) to satisfy identified functional capabilities. The materiel portion is restricted to commercial or non-developmental items, which may be purchased commercially or by purchasing more systems from an existing materiel program. | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-18 | 05/2007 |
| Objective | | | | | |
| | ▪ The clearly defined, decisive, and attainable goal toward which every operation is directed. | JP 1-02<br>Appendix | <br>A-1 | <br>385 | 03/2007 |
| | ▪ The specific target of the action taken (for example, a definite terrain feature, the seizure or holding of which is essential to the commander's plan, or, an enemy force or capability without regard to terrain features). | JP 1-02<br>Appendix | <br>A-1 | <br>385 | 03/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Objective Value | | | | | |
| | ▪ The desired operational goal associated with a performance attribute beyond which any gain in utility does not warrant additional expenditure. The objective value is an operationally significant increment above the threshold. An objective value may be the same as the threshold when an operationally significant increment above the threshold is not significant or useful. | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-19 | 05/2007 |
| Operation | | | | | |
| | ▪ A military action or the carrying out of a strategic, operational, tactical, service, training, or administrative military mission. | JP 1-02<br>Appendix | <br>A-1 | <br>390 | 03/2007 |
| | ▪ The process of carrying on combat, including movement, supply, attack, defense, and maneuvers needed to gain the objectives of any battle or campaign. | JP 1-02<br>Appendix | <br>A-1 | <br>390 | 03/2007 |
| Operational Effectiveness | | | | | |
| | ▪ Measure of the overall ability to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, supportability, survivability, vulnerability, and threat. | CJCSI 3170.01F<br>Glossary | <br>GL | <br>GL-17 | 05/2007 |
| Operational Test and Evaluation (OT&E) | | | | | |
| | ▪ The field test, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for the purpose of determining the effectiveness and suitability of the weapons, equipment or munitions for use in combat by typical military users, and the evaluation of the results of such test. | AFOTEC OT&E Guide, 5th edition.<br>Glossary | <br>Attch B | <br>B-26 | 06/2007 |
| Operational Threat Environment | | | | | |
| | ▪ A generalized overview of the operational, physical and technological environment in which the system will have to function during its lifetime. Developments and trends that can be expected to affect mission capability during the system's life span should be included. Areas to be covered should include all generations of threat as outlined by US Strategic Command.<br>– Threats, first generation: Common hacker tools and techniques used in a non-sophisticated manner. Lone or possibly small groups of amateurs without large resources.<br>– Threats, second generation: Non state-sponsored computer network attack, espionage or data theft. Common tools used in a sophisticated manner. Individuals or small groups supported by resources of a business, criminal syndicate or other trans-national group, including terrorists.<br>– Threats, third generation: State-sponsored computer network attack or espionage. More sophisticated threat (than first and second) supported by institutional processes and significant resources. | CJCSI 6510.01E<br>Glossary | <br>GL | <br>GL-13 | 08/2007 |
| Operational View (OV) | | | | | |
| | ▪ The OV captures the operational nodes, the tasks or activities performed, and the information that must be exchanged to accomplish DOD missions. It conveys the types of information exchanged, the frequency of exchange, which tasks and activities are supported by the information exchanges, and the nature of information exchanges.<br>– OV-1 -- Operational Level Operational Concept Graphic:<br>High-level graphical/textual description of operational concept. | DoDAF 1.5, Volume I<br>1 | <br>1.4.2 | <br>1-8 | 04/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| | – OV-2 -- Operational Node Connectivity Description:<br>Operational nodes, connectivity, and information exchange need lines between nodes.<br>– OV-3 -- Operational Information Exchange Matrix:<br>Information exchanged between nodes and the relevant attributes of that exchange.<br>– OV-4 -- Organizational Relationships Chart:<br>Organizational, role, or other relationships among organizations.<br>– OV-5 -- Operational Activity Model:<br>Capabilities, operational activities, relationships among activities, inputs, and outputs; overlays can show cost, performing nodes, or other pertinent information.<br>– OV-6a -- Operational Rules Model:<br>One of three products used to describe operational activity—identifies business rules that constrain operation.<br>– OV-6b -- Operational State Transition Description:<br>One of three products used to describe operational activity—identifies business process responses to events.<br>– OV-6c -- Operational Event-Trace Description:<br>One of three products used to describe operational activity—traces actions in a scenario or sequence of events.<br>– OV-7 -- Logical Data Model:<br>Documentation of the system data requirements and structural business process rules of the Operational View. | | |
| Physical Design | | | |
| | ▪ The physical design identifies all of the services or components necessary to implement the logical design. | CTM v3.0<br>--- --- --- | 04/2009 |
| Precise | | | |
| | ▪ The ability to act directly upon key elements and processes demands precisely executed joint actions. Precision extends beyond surgical strikes to the exact application of all joint force capabilities to achieve greater success at less risk. Knowledge gained in all dimensions will enhance the capability of the JFC to understand a situation, determine the effects desired, select a course of action and the forces to execute it, accurately assess the effects of that action and reengage as necessary. Regardless of its application in combat or noncombat operations, the capability to engage precisely allows commanders to shape situations or battlespace in order to generate the desired effects while minimizing unintended effects and contributing to the most effective use of resources. The overall effect of precision is far-reaching with considerable payoff in terms of combat effectiveness.<br>(See Joint Force Characteristics) | CCJO 2.0<br>4. Solution 4.E.7. 22 | 08/2005 |
| Precision | | | |
| | ▪ The quality or state of being clearly depicted, definite, measured or calculated. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| | ▪ A quality associated with the spread of data obtained in repetitions of an experiment as measured by variance; the lower the variance, the higher the precision. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| | ▪ A measure of how meticulously or rigorously computational processes are described or performed by a model or simulation. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| Program Introduction (PI) | | | |
| | ▪ The Program Introduction (PI) is the test customer's initial requirements document to the lead support agency. Within the context of the capabilities test methodology, the PI should include the test concept, the test evaluation strategy, and the joint operational context for test. The PI is also referred to as the program introduction document or PID. | CTM v3.0<br>--- --- --- | 04/2009 |
| | ▪ The PI is the initial planning document submitted by a user to the support agency immediately on identification of the scope and duration of a program activity. The user should submit the PI using the best available information, enabling the support agency to initiate resource and technical planning. This information, while sometimes fragmentary and incomplete, is of substantial value to the support agency in determining the scope of the program. For many programs, the PI will eliminate further documentation except for conducting specific operations. | Universal Documentation System Handbook 501-97<br>https://wsmrc2vger.wsmr.army.mil/rcc/manuals/uds/501chaps.htm<br>--- --- --- | 11/1997 |
| Program Manager (PM) | | | |
| | ▪ The individual designated by the implementing command as having single-point management responsibility for an acquisition program. The program director may delegate specific program authority to system program office staff members as long as the authority is documented in management instructions or official correspondence. | AFOTEC OT&E Guide, 5th edition.<br>Glossary Attch B B-28 | 06/2007 |
| Referent | | | |
| | ▪ The best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation. | The Referent Study Final Source (by D.K. Pace)<br>Executive Summary --- ES-1 | 06/2004 |
| Resilient | | | |
| | ▪ To operate successfully, the future joint force must be able to protect and sustain its capabilities from the effects of adversaries or adverse conditions. It must also be able to withstand pressure or absorb punishment without permanently losing its focus, structure, momentum, or integrity. Resilience provides joint forces with the ability to sustain performance at high levels, despite losses, setbacks, or similar developments. The future joint force must be resilient to meet the demands of being successful across the ROMO [Range of Military Operations] in an uncertain future security environment. (See Joint Force Characteristics) | CCJO 2.0<br>4. Solution 4.E.9. 23 | 08/2005 |
| Resolution | | | |
| | ▪ The degree of detail used to represent aspects of the real world or a specified standard or referent by a model or simulation. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| | ▪ Separation or reduction of something into its constituent parts; granularity. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Risk | | | | | |
| | ▪ Probability and severity of loss linked to hazards. | JP 1-02<br>Appendix | A-1 | 465 | 03/2007 |
| Scenario | | | | | |
| | ▪ An account or synopsis of a projected course of action or events. For the purpose of this Instruction, the focus of scenarios is on strategic and operational levels of warfare. Scenarios include information such as threat and friendly politicomilitary contexts and/or backgrounds, assumptions, operational objectives, and other planning considerations. | DODI 8260.01<br>Definitions | E1.3. | 6 | 01/2007 |
| Sensitivity | | | | | |
| | ▪ The ability of a component, model or simulation to respond to a low level stimulus. | Recommended Practices Guide<br>Fidelity RPG Special Topic | --- | 8 | 09/2000 |
| Service | | | | | |
| | ▪ A distinct part of the functionality that is provided by a system on one side of an interface to a system on the other side of an interface. | DoDAF 1.5, Volume II<br>2. Architecture Basics | 2-5 | 2-14 | 04/2007 |
| Specified Task | | | | | |
| | ▪ In the context of joint operation planning, a task that is specifically assigned to an organization by its higher headquarters. | CJCSM 3500.04E<br>Glossary | GL-6 | 63 | 08/2008 |
| Standard | | | | | |
| | ▪ Quantitative or qualitative measures for specifying the levels of performance of a task. | CJCSM 3500.04E<br>Glossary | GL-6 | 63 | 08/2008 |
| Statement of Capability (SC) | | | | | |
| | ▪ The SC is the support agency's response to the PI. The SC is a basic agreement between the user and the support agency. Within the context of the capabilities test methodology, the SC should incorporate the test concept, the test evaluation strategy, and the joint operational context for test. The SC is also referred to as the statement of capabilities or SOC. | CTM v3.0<br>--- | --- | --- | 04/2009 |
| | ▪ The SC is the support agency's response to the PI. When properly signed, the SC is evidence that a program has been accepted for support by the support agency. Support conditions, qualifications, and resources, or other considerations are initially identified in this document which serves as a baseline reference for subsequent acceptance and commitment by the support agency. | Universal Documentation System Handbook 501-97<br>https://wsmrc2vger.wsmr.army.mil/rcc/manuals/uds/501chaps.htm<br>--- | --- | --- | 11/1997 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| Strategic Analysis | | | | | |
| | ▪ An analysis of force sufficiency and effectiveness conducted by the DOD Components to support the development and evaluation of the defense strategy. Such analyses address both forces and enablers (e.g., intertheater and intratheater lift capability, required language skill, and regional expertise capabilities). | DODI 8260.01<br>Definitions | <br>E1.4. | <br>6 | 01/2007 |
| Supporting Task | | | | | |
| | ▪ Specific activities that contribute to accomplishment of a joint mission-essential task. Supporting tasks associated with a command or agency's mission-essential task list are accomplished by the joint staff or subordinate commands or agencies. | CJCSM 3500.04E<br>Glossary | <br>GL-6 | <br>63 | 08/2008 |
| System | | | | | |
| | ▪ A functionally, physically, and/or behaviorally related group of regularly interacting or interdependent elements; that group of elements forming a unified whole. | JP 1-02<br>Appendix | <br>A-1 | <br>523 | 03/2007 |
| System Design Document (SDD) | | | | | |
| | ▪ The System Design Document (SDD) is created during the CTM. The SDD Implement LVC-DE phase describes the Live/Virtual/Constructive Distributed Environment (LVC-DE) configuration for the JME and is a unique configuration of the LVC-DE baseline system that supports a particular capability test. The SDD will support the JME physical design specifications (Joint Mission Environment System Design Document Template). | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| System of Systems (SoS) | | | | | |
| | ▪ A set or arrangement of interdependent systems that are related or connected to provide a given capability. The loss of any part of the system could significantly degrade the performance or capabilities of the whole. The development of an SoS solution will involve trade space between the systems as well as within an individual system performance. | CJCSM 3170.01C<br>Glossary | <br>GL | <br>GL-21 | 05/2007 |
| | ▪ Systems that include hardware, software, data, personnel, procedures, and facilities (DOTMLPF & MOD Lines of Development). | Dandashi, SE DSIG-OMB UML Profile for DoDAF/MODAF; http://www.acq.osd.mil/osjtf/SOS/7_UML %20Profile%20for%20DODAF-MODAF_20050922.ppt<br>--- | <br>--- | <br>Slide 4 | 06/2005 |
| System Threat Assessment Report (STAR) | | | | | |
| | ▪ The basic authoritative threat assessment, tailored for and focused on, a particular (i.e., single) U.S. major defense system. It describes the threat to be countered in the projected threat environment. The threat information should reference DIA-validated documents. | DoD 5200.1-M<br>Definitions | <br>DL | <br>DL1.1.<br>30. | 03/1994 |
| System Under Test (SUT) | | | | | |
| | ▪ An implemented capability increment during an acquisition program, which is the focus of evaluation during a capability test. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| System/System of Systems Attribute | | | |
| | ▪ A quantitative or qualitative performance characteristic of a system or system of systems that make a significant contribution to the characteristics of the future joint force as defined in the Capstone Concept for Joint Operations or to other characteristics deemed necessary by the sponsor (e.g., suitability, survivability) across doctrine, organization, training, materiel, leadership & education, personnel, and facilities. System/System of Systems Attributes can include Key Performance Parameters (KPPs); Key System Attributes (KSAs); Critical Technical Parameters (CTPs); and system-level measures of performance (MOPs), measures of effectiveness (MOEs), or measures of suitability (MOS). | CTM v3.0<br>--- --- --- | 04/2009 |
| Systems and Services View (SV) | | | |
| | ▪ The SV captures system, service, and interconnection functionality providing for, or supporting, operational activities. DOD processes include warfighting, business, intelligence, and infrastructure functions. The SV system functions and services resources and components may be linked to the architecture artifacts in the OV. These system functions and service resources support the operational activities and facilitate the exchange of information among operational nodes.<br>– SV-1 -- Systems/Services Interface Description:<br>Identification of systems nodes, systems, system items, services, and service items and their interconnections, within and between nodes.<br>– SV-2 -- Systems/Services Communications Description:<br>Systems nodes, systems, system items, services, and service items and their related communications laydowns.<br>– SV-3 -- Systems-Systems/Services-Systems/Services-Services Matrix:<br>Relationships among systems and services in a given architecture; can be designed to show relationships of interest, e.g., system-type interfaces, planned vs. existing interfaces, etc.<br>– SV-4a -- Systems Functionality Description:<br>Functions performed by systems and the system data flows among system functions.<br>– SV-4b -- Services Functionality Description:<br>Functions performed by services and the service data flow among service functions.<br>– SV-5a -- Operational Activity to Systems Function Traceability Matrix:<br>Mapping of system functions back to operational activities.<br>– SV-5b -- Operational Activity to Systems Traceability Matrix:<br>Mapping of systems back to capabilities or operational activities.<br>– SV-5c -- Operational Activity to Services Traceability Matrix:<br>Mapping of services back to operational activities.<br>– SV-6 -- Systems/Services Data Exchange Matrix:<br>Provides details of system or service data elements being exchanged between systems or services and the attributes of that exchange.<br>– SV-7 -- Systems/Services Performance Parameters Matrix:<br>Performance characteristics of Systems and Services View elements for the appropriate time frame(s).<br>– SV-8 -- Systems/Services Evolution Description:<br>Planned incremental steps toward migrating a suite of systems or services to a more efficient suite, or toward evolving a current system to a future implementation. | DoDAF 1.5, Volume I<br>1 1.4.3 1-8 | 04/2007 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | | | Source Date |
|---|---|---|---|---|---|
| | – SV-9 -- Systems/Services Technology Forecast:<br>Emerging technologies and software/hardware products that are expected to be available in a given set of time frames and that will affect future development of the architecture.<br>– SV-10a -- Systems/Services Rules Model:<br>One of three products used to describe system and service functionality—identifies constraints that are imposed on systems/services functionality due to some aspect of systems design or implementation.<br>– SV-10b -- Systems/Services State Transition Description:<br>One of three products used to describe system and service functionality—identifies responses of a system/service to events.<br>– SV-10c -- Systems/Services Event-Trace Description:<br>One of three products used to describe system or service functionality—identifies system/service-specific refinements of critical sequences of events described in the Operational View.<br>– SV-11 -- Physical Schema:<br>Physical implementation of the Logical Data Model entities, e.g., message formats, file structures, physical schema. | | | | |
| Tactics, Techniques, and Procedures (TTP) | | | | | |
| | ▪ Tactics -- The employment and ordered arrangement of forces in relation to each other. | JP 1-02<br>Appendix | <br>A-1 | <br>530 | 03/2007 |
| | ▪ Techniques -- Non-prescriptive ways or methods used to perform missions, functions, or tasks. | JP 1-02<br>Appendix | <br>A-1 | <br>537 | 03/2007 |
| | ▪ Procedures -- Standard, detailed steps that prescribe how to perform specific tasks. | JP 1-02<br>Appendix | <br>A-1 | <br>428 | 03/2007 |
| Task | | | | | |
| | ▪ An action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability.<br>(See Essential Task) | CJCSM 3500.04E<br>Glossary | <br>GL-6 | <br>63 | 08/2008 |
| Task Measures of Performance (TMOP) | | | | | |
| | ▪ Task Measures of Performance (TMOPs) are used to quantify mission task accomplishment. TMOPs are defined using the Universal Joint Task List (UJTL) structure of task, condition, and standard and are based on joint force commander mission tasks deemed essential to mission accomplishment using specified conditions and standards. | CTM v3.0<br>--- | <br>--- | <br>--- | 04/2009 |
| Technical Assessment | | | | | |
| | ▪ Technical Assessment activities measure technical progress and the effectiveness of plans and requirements. Activities within Technical Assessment include the activities associated with Technical Performance Measurement and the conduct of technical reviews. A structured review process should demonstrate and confirm completion of required accomplishments and exit criteria as defined in program and system planning. Technical reviews are discussed in detail in section 4.3. Technical assessment activities discover deficiencies or anomalies that often result in the application of corrective action. | Defense Acquisition University Guidebook<br>4 | <br>4.2.3.3 | <br>--- | 12/2008 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Technical Standards View (TV) | | | |
| | ▪ The TV is the minimal set of rules governing the arrangement, interaction, and interdependence of system parts or elements. Its purpose is to ensure that a system satisfies a specified set of operational requirements. The TV provides the technical systems implementation guidelines upon which engineering specifications are based, common building blocks are established, and product lines are developed. It includes a collection of the technical standards, implementation conventions, standards options, rules, and criteria that can be organized into profile(s) that govern systems and system or service elements for a given architecture.<br>– TV-1 -- Technical Standards Profile:<br>Listing of standards that apply to Systems and Services View elements in a given architecture.<br>– TV-2 -- Technical Standards Forecast:<br>Description of emerging standards and potential impact on current Systems and Services View elements, within a set of time frames. | DoDAF 1.5, Volume I<br>1 1.4.4 1-9 | 04/2007 |
| Test and Evaluation Master Plan (TEMP) | | | |
| | ▪ All programs on the OSD T&E Oversight List are required to submit for OSD approval a master plan that describes the total T&E planning from component development through operational T&E into production and acceptance. The program manager, with T&E WIPT providing support, is responsible for producing the TEMP. It is an important document in that it contains the required type and amount of test and evaluation events, along with their resource requirements. The TEMP is considered a contract among the program manager, OSD, and the T&E activities. The program manager must follow the approved TEMP to budget for T&E resources and schedules, which is why it is imperative that all T&E stakeholders participate early in the T&E Strategy development and make timely updates when events or resource requirements change. Stakeholders should include representatives from USD(AT&L) (e.g., SE/AS) and DOT&E, as those offices ultimately will approve the TEMP. | Defense Acquisition University Guidebook<br>9 9.10 --- | 12/2008 |
| Test and Evaluation Strategy (TES) | | | |
| | ▪ The TES is an early T&E planning document that describes the T&E activities starting with Technology Development and continuing through System Development and Demonstration into Production and Deployment. Over time, the scope of this document will expand, the TES will evolve into the TEMP due at Milestone B. The TES describes, in as much detail as possible, the risk reduction efforts across the range of activities (e.g., M&S, DT&E, OT&E, etc.) that will ultimately produce a valid evaluation of operational effectiveness, suitability, and survivability before full-rate production and deployment. It is a living document and should be updated as determined by the T&E WIPT during the Technology Development Phase. | Defense Acquisition University Guidebook<br>9 9.6.1.1 --- | 12/2008 |
| Test Approach | | | |
| | ▪ A description of the overall scope of the live, virtual, and constructive test including estimated number and size of events, and estimated test schedule. Live operations are highlighted and locations identified, such as a DOD range or contractor facilities. Detailed planning for live operations is normally conducted at individual ranges or facilities using local procedures. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Concept | | | |
| | ▪ The Test Goal, Objectives, and Approach (TGOA) necessary for a System Under Test (SUT) program manager to initially characterize a test to the organization orchestrating the LVC Distributed Environment. | CTM v3.0<br>--- --- --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Test Data | | | |
| | ▪ Quantitative or qualitative information collected during one or more test events. Data to be collected during a test are identified in the integrated data requirements list contained in the test plan. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Event | | | |
| | ▪ A collective term used to describe an event executed in consonance with an approved scenario. A test event occurs in a scheduled test venue using a combination of LVC components, with Service provided test resources, in a realistic joint mission environment for the purpose of generating and collecting SoS data. A test event is supported by a detailed test plan and typically consists of numerous test trials. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Goal | | | |
| | ▪ The Test Goal provides a high-level understanding of the Joint Capability under test and its contribution to achieving the Joint Mission. These goals should include the following definitions:<br>– The Joint Mission and statement for test, to include any key materiel and non-materiel system of systems information relevant to the test.<br>– The relevant portion of the Family of Joint Future Concepts, CONOPS, or UCP-assigned mission to which Joint Capability contributes, and the desired end state (operational outcome).<br>– The Joint system capabilities (key performance characteristics) that are to be tested to provide scope for the overall test.<br>– The enabling capabilities that may be required to achieve the desired mission outcomes.<br>– The traceability to relevant Defense Planning Scenarios (DPS) and Multi-Service Force Deployment (MSFD) documentation. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Objective | | | |
| | ▪ Test Objectives focus the test goal on a specific capability subset defined by a Critical Capability Issue (CCI) to capability crosswalk mission desired effects, tasks, conditions (e.g., threat and environment), and system/SoS elements to set the stage for developing a capability test design. The test objective should reference a critical joint issue; a focused subset of the capability crosswalk; and a test scenario, derived from the Joint Operational Context for Test (JOC-T). | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Plan | | | |
| | ▪ The test plan will provide sufficient detail to identify data and resource requirements to support the assessment/evaluation. It will list CCIs, Task MOPs, and attribute measures as well as describe test limitations, safety and security issues, specific test events, scenarios, schedule, measures, data collection (who, what, when, where, why, and how), reduction, and analysis.<br>– It will show linkages between data to be collected, information to be obtained, and conclusions needed. It will also show differences between LVC-DE scenarios versus operational scenarios and the tested system/SoS versus the planned operational system/SoS and describe how these differences (limitations) will be addressed. It can include a Data Analysis Plan (DAP). | CTM v3.0<br>--- --- --- | 04/2009 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Test Run | | | |
| | ▪ A Test Run is one instantiation of a test on a system or system of systems under a Joint Operational Context for Test (JOC-T). Multiple iterations of a test run with the same set of independent variables held at the same values would make a test trial. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Scenario | | | |
| | ▪ An overarching narrative reflecting a realistic LVC joint mission environment containing necessary elements from the Joint Operational Context for Test as required by the Test Concept and Evaluation Strategy. The test scenario describes proposed test events in joint operational terms, generally without regard to how the joint test scenario will be implemented (except identify any live, virtual, or constructive requirements or constraints). The test scenario describes relevant operational organizations, resources, missions, and threats that will interact with the client system under test. The test scenario provides insight into what operational entities are required and how they interact as test events are executed. Any operational constraints imposed by requirements should be specified, such as organization behaviors and rules of engagement. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Trial | | | |
| | ▪ A Test Trial is one or more test runs where a set of independent variables (joint mission vignettes, system of system material and non-material configurations, threat and environmental conditions) is held constant. | CTM v3.0<br>--- --- --- | 04/2009 |
| Test Vignette | | | |
| | ▪ Test vignettes are subsets of the overall Test Scenario. Each vignette is focused on one or more test objectives from the Test Concept. Using an analogy, a vignette is a scene and the scenario is the movie or play. Each vignette will be comprised of sets of system of systems combinations and test conditions, i.e., controlled variables (or factors) under which the test systems and participants will be subjected for a test trial or set of test trials to measure system of systems performance and joint mission effectiveness (JMe). | CTM v3.0<br>--- --- --- | 04/2009 |
| Threat Condition | | | |
| | ▪ Those threat (e.g., threat actions, threat order of battle, threat command and control structure, threat systems, threat force laydown) condition variables of an operational environment or situation in which a unit, system, or individual is expected to operate. | CTM v3.0<br>--- --- --- | 04/2009 |
| Threshold Value | | | |
| | ▪ A minimum acceptable operational value below which the utility of the system becomes questionable. | CJCSM 3170.01C<br>Glossary GL GL-21 | 05/2007 |
| Tolerance | | | |
| | ▪ The maximum permissible error or the difference between the maximum and minimum allowable values in the properties of any component, device, model, simulation or system relative to a standard or referent. Tolerance may be expressed as a percent of nominal value, plus and minus so many units of a measurement, or parts per million. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| | ▪ The character, state or quality of not interfering with some thing or action. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| Universal Joint Task List (UJTL) | | | |
| | ▪ A menu of capabilities (mission-derived tasks with associated conditions and standards, i.e., the tools) that may be selected by a joint force commander to accomplish the assigned mission. Once identified as essential to mission accomplishment, the tasks are reflected within the command joint mission essential task list. | JP 1-02<br>Appendix A-1 568 | 03/2007 |
| Use Case [Class] | | | |
| | ▪ The specification of a sequence of actions, including variants, that a system (or other entity) can perform, interacting with actors of the system. | DoDAF 1.5, Volume II<br>Annex C C-5 | 04/2007 |
| Validation | | | |
| | ▪ For the purpose of this Directive, the process of determining the degree to which a model is an accurate representation of the real-world from the perspective of the intended uses of the model. | DOD 5000.59<br>Definitions E2.17. 7 | 08/2006 |
| Validity | | | |
| | ▪ The quality of being inferred, deduced, or calculated correctly enough to suit a specific application. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| | ▪ The quality of maintained data that is found on an adequate system of classification (e.g., data model) and is rigorous enough to compel acceptance for a specific use. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| | ▪ The logical truth of a derivation or statement, based on a given set of propositions. | Recommended Practices Guide<br>Fidelity RPG --- 8<br>Special Topic | 09/2000 |
| Verification | | | |
| | ▪ For the purpose of this Directive, the process of determining that a model implementation accurately represents the developer's conceptual description and specifications. | DOD 5000.59<br>Definitions E2.18. 7 | 08/2006 |
| Vignette | | | |
| | ▪ A concise narrative description that illustrates and summarizes pertinent circumstances and events from a scenario. | CJCSI 3010.02B<br>Glossary GL GL-4 | 01/2006 |
| Virtual Simulation | | | |
| | ▪ A simulation involving real people operating simulated systems. Virtual simulations inject human-in-the-loop (HITL) in a central role by exercising motor control skills (e.g., flying an airplane), decision skills (e.g., committing fire control resources to action), or communication skills (e.g., as members of a C4I team). | DOD 5000.59<br>Definitions 36b A-6 | 08/2006 |

| Term | Definition | Source (Chapter, Paragraph, Page #) | Source Date |
|---|---|---|---|
| Ways | | | |
| | ▪ Doctrine, tactics, techniques, and procedures, competencies, and concepts. | Terms of Reference for Conducting a Joint Capability Area Baseline Reassessment; http://www.dtic.mil/futurejointwarfare/strategic/jca_tor9apr07.doc<br>--- --- --- | 04/2007 |
| | ▪ Applied to Testing in a Joint Environment, Ways are methods and processes, including the Capability Test Methodology (CTM), Capability Evaluation Metamodel (CEM), and the Joint Mission Environment Foundation Model (JFM) required to define and operate an instantiated Joint Mission Environment (JME) System of Systems (SoS). | CTM v3.0<br>--- --- --- | 04/2009 |

# ANNEX M – CHARACTERISTICS AND ATTRIBUTES

The evaluation strategy calls for three levels of measures:

- Measuring system of systems (SoS) contributions to joint mission effectiveness
- Measuring SoS and system task performance
- Measuring SoS and system characteristics

This annex focuses on measuring SoS and system characteristics.

A system is defined as:

- A functionally, physically, and/or behaviorally related group of regularly interacting or interdependent elements; that group of elements forming a unified whole. (JP 1-02)
- The organization of hardware, software, material, facilities, personnel, data, and services needed to perform a designated function with specified results, such as the gathering of specified data, its processing, and delivery to users. (DAU T&E Management Guide)

The acquisition process is focused on systems, supplies, or services, therefore traditional developmental test and evaluation (DT&E) and operational test and evaluation (OT&E) are structured to evaluate system performance, effectiveness, and suitability through operational and technical characteristics. A weakness in this process is that systems perform in a joint environment as a system of systems (SoS). An SoS is defined as a set or arrangement of interdependent systems that are related or connected to provide a given capability. The loss of any part of the system could significantly degrade the performance or capabilities of the whole. The development of an SoS solution will involve trade space between the systems as well as within an individual system performance (CJCSM 3170.01C). The interdependence of the systems will also apply to their characteristics, thus resulting in SoS level characteristics that must be evaluated. This is the rationale for the evaluation strategy to include evaluating system level characteristics and SoS level characteristics.

A characteristic is defined as a desirable trait, quality, or property that distinguishes how the future joint force should conduct military operations (CJCSI 3010.02B). A system or SoS can have a characteristic that may need to be measured. Typically, a system characteristic will contribute to the overall SoS characteristic. For example, lethality of a system will contribute to lethality of the overall SoS. Characteristics are measured through attributes.

An attribute is defined as a quantitative or qualitative characteristic of a system of systems that is expressed in terms of joint doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF). Attributes may apply at the system or SoS levels. System/SoS attributes can include key performance parameters (KPP); key system attributes (KSA); critical technical parameters (CTP); and system-level measures of performance (MOP), measures of effectiveness (MOE), or measures of suitability (MOS).

Characteristics and attributes are not fully standardized across Service acquisition programs. To that end, this annex lists previously defined characteristics and attributes which appear in existing doctrine in order to aid the analyst in identifying SoS characteristics and attributes that may need to be evaluated. The list is not intended to be all inclusive, but to represent possible

evaluation criteria for an SoS. The table is sorted on the term and source columns, and includes entries from various joint and Service sources.

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Accessibility | Accessibility is a general term used to describe the degree to which a product (e.g., device, service, environment) is accessible by as many people as possible. Accessibility can be viewed as the "ability to access" the functionality, and possible benefit, of some system or entity. Accessibility is often used to focus on people with disabilities and their right of access to entities, often through use of assistive technology. | http://en.wikipedia.org/wiki/Accessibility | | | | Mar-09 |
| Accountability | Accountability is a concept in ethics with several meanings. It is often used synonymously with such concepts as responsibility, answerability, enforcement, blameworthiness, liability and other terms associated with the expectation of account-giving. As an aspect of governance, it has been central to discussions related to problems in both the public and private (corporation) worlds. | http://en.wikipedia.org/wiki/Accountability | | | | Mar-09 |
| Accuracy | Measure example: Horizontal and vertical geolocation accuracy in feet. | CJCSM 3500.04C | Annex A Appendix C | Enclosure B | B-C-A-27 | Jul-02 |
| Accuracy | In the fields of science, engineering, industry and statistics, accuracy is the degree of closeness of a measured or calculated quantity to its actual (true) value. Accuracy is closely related to precision, also called reproducibility or repeatability, the degree to which further measurements or calculations show the same or similar results.[1]<br>1 John Robert Taylor (1999). An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books. pp. 128-129. ISBN 093570275X. http://books.google.com/books?id=giFQcZub80oC&pg=PA128. | http://en.wikipedia.org/wiki/Accuracy | | | | Mar-09 |
| Accuracy | The degree to which a parameter or variable or set of parameters or variables within a model or simulation conform exactly to reality or to some chosen standard or referent. | Recommended Practices Guide | Fidelity RPG | Special Topic | 8 | Sep-00 |
| Adaptability | How well M&P adapts to different enterprises, changing environments, compressed deliverable timelines etc. | JTEM Effectiveness and Suitability<br>Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Adaptability - m | Measure example: Hours to construct an adaptive plan against one target. "Yes/No" for: Present a comprehensive description of key aspects and procedures of nuclear adaptive planning at the level of understanding of the decision maker. | CJCSM 3500.04C | Annex A Appendix C | Enclosure B | B-C-A-43, 47 | Jul-02 |
| Adaptable/ Tailorable | An adaptable/tailorable joint force is versatile in handling threat missions with equal success; scalable in applying appropriate mass and weight of effort; agile in shifting between different types of missions without loss of momentum; responsive to changing conditions and environments; and whose leaders are intellectually empowered by a background of experience and education. Adaptability ensures that the joint force can rapidly shift from one operation to another across the range of military operations, and adjust operations based on changing conditions. An adaptive mindset and flexible force capabilities are essential for success in countering the full spectrum of anticipated threats and challenges and enhance the joint force ability to respond with unmatched speed of decision and action. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |
| Affordability | The degree to which a complex system costs are feasible over the life cycle. For example, selection and use of components that are COTS (commercial off the shelf) in lieu of more expensive custom components | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Affordability | The extent to which something is affordable, as measured by its cost relative to the amount that the purchaser is able to pay. | http://en.wiktionary.org/wiki/affordability | | | | Mar-09 |
| Agile | An agile joint force has the ability to move quickly and seamlessly to defuse (or help defuse) a crisis situation or effectively operate inside the decision loop of even the most capable adversary. Agility is about timeliness--thinking, planning, communicating, and acting in a manner that allows effective and efficient adaptation to an unfolding situation. Agility permits JFCs to exploit fleeting opportunities, protect friendly vulnerabilities, and adapt rapidly to changes in the operational environment--a characteristic essential to a force that is expected to succeed across the range of military operations. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Attribute | A quantitative or qualitative characteristic of an element or its actions. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-4 | Jun-07 |
| Availability | The ease with which the product can be accessed and used. A design consideration focused on the readiness of a system to perform its intended function. Related and depends on maintainability, serviceability, and reliability of a complex system. For example, a poorly designed complex system may be ready for use only small fraction of the times during which it is called into action. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Availability | The degree to which a system, subsystem, or equipment is operable and in a committable state at the start of a mission, when the mission is called for at an unknown, i.e., a random, time. Simply put, availability is the proportion of time a system is in a functioning condition. | http://en.wikipedia.org/wiki/Availability | | | | Mar-09 |
| Availability | A measure of the degree to which an item is in an operable and committable state at the start of a mission when the mission is called for at an unknown (random) time. In OT&E, Ao is the usual measure. (See Operational Availability.) | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 1 | Mar-09 |
| Availability | A measure of the degree to which an item is operable and can be committed at the start of a mission when the mission is called for at an unknown (random) point in time.[Defense Acquisition University Glossary.] | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | a | 2 | Oct-05 |
| Availability | When conducting OT&E, Availability is normally expressed as Operational Availability (Ao) which is a measure of the probability that a system will be operating or capable of operation when required. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 7 | c | 5 | Oct-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Availability - m | Measure example: Months since formal review of worldwide nuclear asset availability. | CJCSM 3500.04C | Annex A Appendix C | Enclosure B | B-C-A-45 | Jul-02 |
| Capability | The capacity to be used, treated, or developed for a specific purpose. Capability descriptions must contain the following elements: Key characteristics (attributes) with appropriate parameters and metrics. (CJCSI 3170.01) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-5 | Jun-07 |
| Capability | Measure example: Percent of passed Nuclear Weapon Tech Inspections of Service Component Command capability to handle, store, and maintain nuclear weapons with a Satisfactory or better. | CJCSM 3500.04C | Annex A Appendix C | Enclosure B | B-C-A-46 | Jul-02 |
| Compatibility | The capability of two or more items or components of equipment or materiel to exist or function in the same system or environment without mutual interference. [Defense Acquisition University Glossary]. Compatibility may apply to a specific investigation of a system's electrical, electromagnetic, physical, and man-machine interface characteristics. Because of such applications, compatibility may also be addressed as part of the operational effectiveness evaluation in OTA test plans and reports. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | b | 2 | Oct-05 |
| Compatibility | The capability of two or more items or components of equipment or materiel to exist or function in the same or function in the same system or environment without mutual interference. (CJCSI 6212.01A) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-7 | Jun-07 |
| Compatibility | One of the elements of operational suitability. The capability of a system or subsystem to operate in its intended environment without adverse effects to or from other systems. Compatibility includes physical, functional, electrical and electronic, and environmental issues. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 2 | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Completeness | The sufficiency of JTEM M&P input/output products and processes to address customer needs | JTEM Effectiveness and Suitability Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |
| Completeness - m | Measure example: Percent completeness of commander's guidance (i.e., coverage of functional areas). | CJCSM 3500.04C | Annex C Appendix C | Enclosure B | B-C-C-109 | Jul-02 |
| Composability | Composability is a system design principle that deals with the inter-relationships of components. A highly composable system provides recombinant components that can be selected and assembled in various combinations to satisfy specific user requirements. The essential attributes that make a component composable are that it be:<br>self-contained (modular)<br>it can be deployed independently - note that it may cooperate with other components, but dependent components are replaceable<br>stateless<br>it treats each request as an independent transaction, unrelated to any previous request | http://en.wikipedia.org/wiki/Composability | | | | Mar-09 |
| Consistency | Alignment of product and processes structures within the M&P and between relevant external DoD/service M&P | JTEM Effectiveness and Suitability Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |
| Correctness | In theoretical computer science, correctness of an algorithm is asserted when it is said that the algorithm is correct with respect to a specification. Functional correctness refers to the input-output behavior of the algorithm (i.e., for each input it produces the correct output). | http://en.wikipedia.org/wiki/Correctness | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Credibility | Credibility refers to the objective and subjective components of the believability of a source or message.<br>Traditionally, credibility has two key components: trustworthiness and expertise, which both have objective and subjective components. Trustworthiness is a based more on subjective factors, but can include objective measurements such as established reliability. Expertise can be similarly subjectively perceived, but also includes relatively objective characteristics of the source or message (e.g., credentials, certification or information quality). Secondary components of credibility include source dynamism (charisma) and physical attractiveness. | http://en.wikipedia.org/wiki/Credibility | | | | Mar-09 |
| Credibility (characteristics of a good model) | A model's credibility results from the logical consistency of its mathematical relationships, and a history of successful (i.e., correct) predictions. A history of successful predictions lends credibility to a model, but full validation—proof that the model's prediction is in accord with reality—is very difficult to attain since observational evidence on those predictions is generally very scarce. Systems that address new problems often require that new models be developed<br>for their trade studies. In that case, full validation is out of the question, and the system engineer must be content with models that have logical consistency and<br>some limited form of outside, independent corroboration. | NASA<br>Systems Engineering Handbook<br>SP-610S | 5 | 5.1.2 | 86 | Jun-95 |
| Data corruption | Data corruption refers to errors in computer data that occur during transmission or retrieval, introducing unintended changes to the original data. Computer storage and transmission systems use a number of measures to provide data integrity, the lack of errors. | http://en.wikipedia.org/wiki/Data_corruption | | | | Mar-09 |
| Decision Time: Command and Control (C2) Center | The only parameter currently in use for the C2 agent is the latency associated with the amount of decision time it takes to assign the mission to an available aircraft. The time is measured in seconds and effectively introduces a delay in between target identification and CAS mission authorization. | System of Systems Test Planning in a Complex Joint Environment<br>NPGS Thesis<br>Christopher M. Wegner | 3 | B.6.a | 23 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Dependability | Dependability is a value showing the reliability of a person to others because of his/her integrity, truthfulness, and trustfulness, traits that can encourage someone to depend on him/her.<br>The wider use of this noun is in Systems engineering.<br>Dependability as applied to a computer system is defined by the IFIP 10.4 Working Group on Dependable Computing and Fault Tolerance as:<br>"[..] the trustworthiness of a computing system which allows reliance to be justifiably placed on the service it delivers [..]" [1]<br>an alternative and broader definition is provided by IEC IEV 191-02-03:<br>"dependability (is) the collective term used to describe the availability performance and its influencing factors : reliability performance, maintainability performance and maintenance support performance"[2]<br>This concept can be further extended to encompass mechanisms to increase and maintain the Dependability of a system [3].<br>Dependability can be thought of as being composed of three elements:<br>Attributes - A way to assess the Dependability of a system<br>Threats - An understanding of the things that can affect the Dependability of a system<br>Means - Ways to increase the Dependability of a system<br>1 IFIP WG10.4 on Dependable Computing and Fault Tolerance<br>2 [1] (search for dependability)<br>3 A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic Concepts and Taxonomy of Dependable and Secure Computing," IEEE Transactions on Dependable and Secure Computing, vol. 1, pp. 11-33, 2004. | http://en.wikipedia.org/wiki/Dependability | | | | Mar-09 |
| Dependability | A measure of the degree to which an item is operable and capable of performing its required function at any (random) time during a specified mission profile, given item availability at the start of the mission. (DSMC Glossary) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-9 | Jun-07 |
| Deployability | The ability of a unit, weapon system, or element thereof, to relocate to a desired area of operations or to a staging area without unacceptable delays. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-9 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| deployability | deployability (plural deployabilities)<br>The extent to which something is deployable | http://en.wiktionary.org/wiki/deployability | | | | Mar-09 |
| Diagnostics | Diagnostics may be expressed as (1) a measure of false alarms (number, percent, probability, rate, etc.) (2) the percent of correct detection given that a fault has occurred (Pcd), (3) the percent of correct fault isolation (and/or fault location) given a correct detection (Pcfi and/or Pcfl), and (4) Mean Time To Fault Locate (MTTFL). | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 7 | d | 5 | Oct-05 |
| Diagnostics | The ability of integrated diagnostics (automated, semi-automated, and manual techniques taken as a whole) to fault-detect and fault-isolate in a timely manner. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 6 | c | 4 | Oct-05 |
| Disposability | The ease with which a system can be destroyed or salvaged at the end of the system life cycle. A design consideration focused on the ease with which a system can be destroyed or salvaged at the end of the system life cycle. For example, use of recyclable materials and easy disassembly to reduce costs of disposal. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Documentation | Documents used to determine suitability e.g., operator and maintenance instructions, repair parts lists, support manuals, and manuals related to computer programs and system software. (Defense Acquisition University Glossary). | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | m | 3 | Oct-05 |
| Durability | durability (plural durabilities)<br>Permanence by virtue of the power to resist stress or force. | http://en.wiktionary.org/wiki/durability | | | | Mar-09 |
| Enduring/ Persistent | This has both a mental and physical aspect. The mental aspect can be expressed as will, while the physical aspect can be expressed as the staying power of the joint force--in both cases, sustaining ours while breaking the adversaries. This characteristic is especially important given the interaction between the anticipated environment, joint force Operations, and unanticipated events in any complex and adaptive system. It demands that the joint force possess the depth and capacity to sustain operations over time, regardless of the situation or adversary. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Evolvability | Evolvability is a concept within the Darwinian understanding of biological evolution. Darwin's theory of evolution by natural selection requires that plants, animals, and other organisms be able to produce offspring that are sometimes better adapted to the circumstances of life than the parents are. It is these offspring that survive and reproduce, and the adaptive traits thus increase in number if they are passed down to the offspring. If the only changes to be found in offspring were deleterious, adaptive evolution could not occur. So, the ability to produce enough advantageous variation to allow adaptive evolution to occur is what is called "evolvability". | http://en.wikipedia.org/wiki/Evolvability | | | | Mar-09 |
| Expeditionary | An expeditionary joint force is organized, postured and capable of rapid and simultaneous deployment, employment, and sustainment. Implicit in this is a joint force that converges mission-tailored capabilities at the desired point of action from dispersed locations around the globe, regardless of anti-access or area-denial environments. As elusive and adaptive adversaries seek refuge in remote and inaccessible areas, the norm will be short-notice operations, austere operational environments, incomplete information and the requirement to fight on arrival throughout the battlespace and to dominate potential adversaries for the duration of a campaign. The future joint force will be immediately employable even in austere conditions and largely independent of existing infrastructure. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |
| Expeditionary (continued) | As a situation evolves, these elements will be readily capable of transitioning to sustained operations, blending into new capability packages to execute follow-on or different operations, or dispersing until otherwise required. The term "expeditionary" also describes the joint force mindset that inculcates an expeditionary perspective into all aspects of force planning, training, and education. The future joint force will increasingly require a mechanism to enable global sourcing of military forces and capabilities; in order to leverage the most responsive, best positioned forces at the time of need. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Extensibility | In software engineering, extensibility (sometimes confused with forward compatibility) is a system design principle where the implementation takes into consideration future growth. It is a systemic measure of the ability to extend a system and the level of effort required to implement the extension. Extensions can be through the addition of new functionality or through modification of existing functionality. The central theme is to provide for change while minimizing impact to existing system functions. | http://en.wikipedia.org/wiki/Extensibility | | | | Mar-09 |
| Failure rate | the number of times that an item fails in a specified period of time. It is a calculated value that provides a measure of reliability for a product, expressed as a rate (failures per unit time). For example, if a component has a failure rate of 2 failures per thousand operating hours, then it is anticipated that 'similar' components will follow the same failure pattern.<br>Failure rate is the inverse of the Mean Time Between Failures (MTBF) value for constant rate failure systems. Constant rate failure assumes that the system failure rate will be consistent for the same operating conditions. For example: If a component has an MTBF value of 500,000 hours, and the failure rate is desired in failures per million hours, the failure rate would be:<br>Failure Rate = (1,000,000 hours) / (500,000 hours) = 2 failures per million hours | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Fast | Key to effectively controlling tempo is the ability to be faster than the adversary or situational events. The speed at which forces maneuver and engage, or decisions are made, or relief is provided, will largely determine operational successes or failures. Successfully overcoming future challenges may require speed of action across all domains. Acting fast is in itself a force multiplier and often a requisite for the effective application of military capabilities. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |
| Fast - m | Measure example: The speed a target is capable of maneuvering. (In KMPH) | CJCSM 3500.04C | Enclosure C | C 2.6.10 Target Speed | C-57 | Jul-02 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Feasibility | Can be built within technical, cost & schedule means. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Fidelity | Fidelity is a notion that at its most abstract level implies a truthful connection to a source or sources. Its original meaning dealt with loyalty and attentiveness to one's duty to a lord or a king, in a broader sense than the related concept of fealty. Both derive from the Latin word fidēlis (A III adjective), meaning "faithful or loyal"In the fields of scientific modeling and simulation, fidelity refers to the degree to which a model or simulation reproduces the state and behavior of a real world object, feature or condition. Fidelity is therefore a measure of the realism of a model or simulation [1]. Simulation fidelity has also been described in the past as 'degree of similarity'[2].1 "SISO-REF-002-1999: Fidelity Implementation Study Group Report". http://www.sisostds.org/index.php?tg=fileman&idx=get&id=5&gr=Y&path=SISO+Products%2FSISO+Reference+Products&file=99S-SIW-167.doc. 2 Hays, R.; Singer, M. (1989). Simulation fidelity in training system design: Bridging the gap between reality and training. Springer-Verlag. | http://en.wikipedia.org/wiki/Fidelity | | | | Mar-09 |
| Flexibility (Adaptability) | The ability of a component or system to adjust to changing requirements, environments, or contexts and still satisfactorily perform mission (without redesign). A design consideration focused on the ability of a component or system to adjust to changing requirements, environments, or contexts and still satisfactorily perform mission (without redesign). For example, design of a complex system capable of making rapid adjustments to climate changes to maintain operation. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Flexibility (engineering) | Flexibility is used as an attribute of various types of systems. In the field of engineering systems design, it refers to designs that can adapt when external changes occur. Flexibility has been defined differently in many fields of engineering, architecture, biology, economics, etc. In the context of engineering design one can define flexibility as the ability of a system to respond to potential internal or external changes affecting its value delivery, in a timely and cost-effective manner. Thus, flexibility for an engineering system is the ease with which the system can respond to uncertainty in a manner to sustain or increase its value delivery. It should be noted that uncertainty is a key element in the definition of flexibility. Uncertainty can create both risks and opportunities in a system, and it is with the existence of uncertainty that flexibility becomes valuable. | http://en.wikipedia.org/wiki/Flexibility_(engineering) | | | | Mar-09 |
| Human Factors | The systematic application of relevant information about human abilities, characteristics, behavior, motivation, and performance. It includes principles and applications in the areas of human engineering, anthropometrics, personnel selection, training, life support, job performance aids, and human performance evaluation. [Defense Acquisition University Glossary.] Within the context of this definition, human factors also may be addressed as part of the operational effectiveness evaluation in OTA test plans and reports. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | i | 3 | Oct-05 |
| Impact Radius: Network Enabled Weapon (NEW) | The impact radius defines a circular area of lethality around the center of the bomb's impact location. Upon impact, the model determines whether the target's true location is within the area and, in combination with the probability of kill (Pk), assigns a target kill or miss accordingly. The impact radius is measured in meters. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.b | 20 | Jun-07 |
| In-Flight Target Update (IFTU) Interval: Joint Terminal Attack Controller (JTAC) | The JTAC does not continuously provide target location updates to other agents on the network. Instead, updates are transmitted at a frequency defined by the assigned IFTU—measured in seconds. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.f | 19 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Interchangeability | Interchangeability can refer to:<br>Interchangeability (telecommunication): A condition in which exist two or more items with characteristics making them equivalent in performance and durability, making them fully exchangeable.<br>Interchangeability (computer science): The ability that an object can be replaced by another object without affecting code using the object.<br>Interchangeable parts: In mechanical engineering, the ability to select components for assembly at random and fit them together within proper tolerances.<br>Interchangeable random variables in mathematics | http://en.wikipedia.org/wiki/Interchangeability | | | | Mar-09 |
| Interoperability | The ability of systems, units, or forces to provide and receive services from other systems, units, or forces, and to use the services so interchanged to enable them to operate effectively together. The conditions achieved among communications-electronics systems or communications-electronics items when information or services can be exchanged directly between them and/or their users. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-17 | Jun-07 |
| Interoperability | Interoperability is a property referring to the ability of diverse systems and organizations to work together (inter-operate). The term is often used in a technical systems engineering sense, or alternatively in a broad sense, taking into account social, political, and organizational factors that impact system to system performance. | http://en.wikipedia.org/wiki/Interoperability | | | | Mar-09 |
| Interoperability | The capability of systems, units, or forces to provide services to or accept services from other systems, units, or forces, and to use the services exchanged to operate effectively (DoDINST 5000.2). Effective exchange of information is emphasized. For example, a radar is interoperable with a gun system if the radar causes the gun to point at the target; the Carrier Aircraft Inertial Navigation System must be interoperable with the Ship's Inertial Navigation System for initial alignment; a fuze must be interoperable with the warhead in order for the firing signal to get through. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 6 | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Interoperability | 1. The ability of systems, units, or forces to provide services to and accept services from other systems, units, or forces and to use the services so exchanged to enable them to operate effectively together. 2. The conditions achieved among communications-electronics systems, or items of communications-electronics equipment, when information or services can be exchanged directly and satisfactorily between them or their users. The degree of interoperability should be defined when referring to specific cases. [Defense Acquisition University Glossary and Joint Pub 1-02.] Interoperability is often addressed as part of the operational effectiveness evaluation in OTA test plans and reports. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | d | 2 | Oct-05 |
| Interoperability | The ability of systems, units, or forces to provide services to and accept services from other systems, units, or forces and to make use the services, units, or forces and to use the services so exchanged to enable them to operate effectively together. (CJCSI 3170.1D) | Naval Capability Evolution Process Guidebook Volume 1 Version 1.1 | Appendix A | Definitions | A-2 | May-05 |
| Interoperability - m | Measure example: Hours / days to review interoperability of required databases and disseminationsystems. | CJCSM 3500.04C | Annex A Appendix C Enclosure B | SN 2.1 Plan and Direct Strategic Intelligence Activities | B-C-A-24 | Jul-02 |
| Interoperable | Interoperability is a necessary prerequisite to integrated and interdependent joint operations. The future joint force will be able to share and exchange knowledge and services between units and commands at all levels. The interoperable joint force can act in an integrated and ultimately an interdependent way among joint force components and capabilities, facilitating more effective interoperability with interagency and multinational partners. Interoperability implies systems, capabilities and organizations working in harmony across all joint force elements; however, it involves more than systems and equipment. Interoperability includes a cultural change at all levels that extends through DOTMLPF. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Interoperable - m | Measure example: Percent of each theater's apportioned vehicles interoperable for fuel. | CJCSM 3500.04C | Annex A Appendix C Enclosure B | SN 7.5 Ensure Interoper-ability | B-C-A-132 | Jul-02 |
| Joint Interoperability | Joint Interoperability is an E-test designed to examine the use of systems which must exchange information or services with non-Navy systems and platforms; that is, Army or Air Force and in some cases, Marines or Coast Guard. For instance, in designing an SP test for a submarine antenna, the capability of the antenna to assist the platform in communicating with Army helicopters, USAF aircraft and satellites, and a Marine CP would have to be examined. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 6 | Mar-09 |
| Knowledge Empowered | The future joint force will emphasize better decisions made faster throughout all levels of command. The fundamentals of this knowledge empowerment are experienced and empowered decision makers benefiting from an enhanced understanding of the environment, potential adversaries and cultures, as well as enhanced collaborative decision-making processes. Although we will never eliminate the fog of war, an increased level of understanding should empower leaders throughout the joint force. This will enable them to anticipate and act as opportunities are presented, apply innovative solutions, mitigate risk, and increase the pace, coherence, and effectiveness of operations even in complex environments. A knowledge empowered force, capable of effective information sharing across all agencies and partners, will be able to make better decisions quicker, increasing joint force effectiveness. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Knowledge Empowered - m | Measure example: Percent of staff possess current knowledge of strategic intent and plans. | CJCSM 3500.04C | Annex B | ST 5.1.4 Monitor Worldwide and Theater Strategic Situation | B-C-B-80 | Jul-02 |
| knowledge: Designing Decision Support Systems for Revolutionary Command and ControlDomains | The final level of worker competencies, the category of knowledge-based behaviors (KBB), represents those behaviors that require complex cognitive processing beyond that of applying an established procedure for a known situation. Situations that require KBB are generally open-ended with no set of previously prescribed rules available to solve a problem, and it is in the KBB phase of cognitive processing that workers confront unanticipated and dynamic situations. Information must be evaluated, often from various sources, and a decision must be made in response to an event. | Designing Decision Support Systems for Revolutionary Command and Control Domains A Dissertation to the faculty of the School of Engineering and Applied Science, University of Virginia, Mary L. Cummings | 2 | 2.1.5 | 41 | Jan-04 |
| Latency: Weapon Control Network (WCN) | Latency is a measure of how long it takes for a message to be transmitted across the network and is measured in seconds. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.5.b | 23 | Jun-07 |
| Launch Aircraft: Launch Distance, Update Requirement, Number of Attempts, and Reattempt Interval | This section describes four different factors for the aircraft agent. The launch distance represents the stand-off distance where the aircraft will initially attempt to release the NEW. If the target location information has not been successfully relayed to the aircraft within the time period set by the update requirement factor, the aircraft will not launch the weapon. This is designed to limit the launch of a weapon when the target is no longer within sensor range of the JTAC. Based on the number of attempts setting, the aircraft may make subsequent attempts to launch the NEW, at a frequency governed by the reattempt interval. While awaiting multiple attempts to launch the weapon, the aircraft continues to fly toward the last known target location. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.4.b | 22 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Launch Aircraft: Speed | The aircraft speed setting is varied over a range representative of a typical strike aircraft and is constant throughout a particular model run. Aircraft speed is measured in meters/second. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.4.a | 22 | Jun-07 |
| Learnability | In software testing learnability, according to ISO 9126, is the capability of a software product to enable the user to learn how to use it. Learnability may be considered as an aspect of usability, and is of major concern in the design of complex software applications.<br>Learnability is defined in the Standard glossary of terms used in software testing published by the International Software Testing Qualifications Board. | http://en.wikipedia.org/wiki/Learnability | | | | Mar-09 |
| Lethal | This is the ability to destroy an adversary and/or his systems in all conditions and environments when required. It includes the use of kinetic and/or non-kinetic means, while leveraging technological advances in greater precision and more devastating target effects at both longer-ranges and in close combat. | CCJO | Appendix C | 4.E.5 | 22 | Aug-05 |
| Lethal - m | Measures example: Percent of attacks using nonlethal means on selected targets, achieve<br>desired nonlethal effect. | CJCSM 3500.04C | Annex B Appendix C Enclosure B | ST 3.2.2 | B-C-B-51 | Jul-02 |
| Lethality | The ability of a munitions system (or laser, high power microwave) to cause damage that results in the loss or degradation of the ability of a target system to complete its designated mission(s). | AFOTEC OT&E Guide, 5th Edition | Enclosure B | Glossary | B-19 | Jun-07 |
| Lethality | The ability of a munition (or laser, high power microwave, and so forth) to cause damage that will cause the loss or degradation in the ability of a target system to complete its designated mission(s). | Army Regulation 73-1Test and EvaluationTest and Evaluation Policy | Glossary | Section II | 55 | Apr-06 |
| Lethality - m | Measures example: e.g., rate of kills given a hit. | CJCSM 3500.04C | Appendix B Enclosure B | 4. e | B-B-4 | Jul-02 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Logistic Supportability | The degree to which the planned logistics (including test equipment, spares and repair parts, technical data, support facilities, and training) and manpower meet system availability and wartime usage requirements. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 7 | Mar-09 |
| Logistics Supportability | The degree of ease to which system design characteristics and planned logistics resources (including the logistics support (LS) elements) allow for the meeting of system availability and wartime usage requirements. [Defense Acquisition University Glossary]. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | k | 3 | Oct-05 |
| M&P Effectiveness | Within the context of the JTEM JT&E, effectiveness is how well JTEM-developed Capability Test Methodology (CTM) overall outputs satisfy end customer requirements, how well the outputs of every CTM process meet the input requirements of internal customers, and how well the inputs from CTM suppliers meet the requirements of the methodology. JTEM effectiveness specifically addresses :<br>(1) the extent to which the goals of the method and processes are attained for designing and executing system-of-systems tests in the JME;<br>(2) the extent to which the goals of the methods and processes are achieved for assessing performance pertaining to capabilities supporting joint missions. JTEM JT&E effectiveness metrics assess the following criteria areas:· Usability (product template and instruction/guidebook usefulness, helpfulness, and simplicity).· Consistency (alignment of product and processes structures within the M&P, and between relevant external DOD/service M&P).· Workflow (leanness of process sequencing, product input/output mappings).· Completeness (the sufficiency of JTEM M&P input/output products and processes to address customer needs).· Repeatability (the degree to which different groups of JTEM M&P users demonstrate similar actions and produce similar output products).Timeliness (the latency of performing JTEM processes or the amount of processes performed in a work period). | CTM v3.0 | | | | Apr-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| M&P Effectiveness - m | Measure example: Percent degradation of mission effectiveness (lack of equipment interoperability). | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 6.2.6 | B-C-C-139 | Jul-02 |
| M&P Suitability | Within the context of JTEM JT&E, suitability is the degree to which JTEM M&P can be efficiently implemented and sustained in a Live, Virtual, Constructive (LVC) distributed range enterprise tasked with designing and executing system-of-systems tests in a JME. JTEM M&P suitability evaluation includes resource utilization, minimization, interoperability, and reuse across non-materiel and materiel criteria areas including:·<br>- Doctrine translates to M&P and policy related to JTEM (relates to issues concerning compatibility with current T&E M&P/policy, interoperability with external DOD domain M&P/policy including acquisition, JCIDS, training, and experimentation).·<br>- Organization (relates to migration/extensions from current T&E organizations and organization-policy change requirement issues).·<br>- Training (relates to M&P and external DOD business practice training issues).·<br>- Materiel (relates to M&P supporting materiel (hardware, software) and LVC distributed range materiel issues).·<br>- Leadership and education (relates to M&P leadership/governance change requirement issues, M&P personnel educational foundation issues).·<br>- Personnel (relates to M&P personnel availability issues and personnel-organization change issues).<br>- Facilities (relates to M&P supporting facility and LVC distributed range facility issues). | CTM v3.0 | | | | Apr-09 |
| M&S Interoperability | The ability of a model or simulation to provide services to and accept services from other models and simulations, and to use the services so exchanged to enable them to operate effectively together. (reference DoD M&S Glossary) | Defense Systems Management College Simulation Based Acquisition: A New Approach LTC Michael V. R. Johnson, Sr., USA LTC Mark F. McKeon, USMC LTC Terence R. Szanto, USAF | Appendix A | Glossary | A-10 | Dec-98 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Maintainability | The ability of an item to be retained in or restored to specified conditions when maintenance is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance and repair. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-20 | Jun-07 |
| Maintainability | The ability of an item to be retained in, or restored to, a specified skill level, using prescribed procedures and resources, at each prescribed level of maintenance and repair. (reference DSMC Glossary of Defense Acquisition Acronyms and Terms, 8th Edition) | Defense Systems Management College Simulation Based Acquisition: A New Approach LTC Michael V. R. Johnson, Sr., USA LTC Mark F. McKeon, USMC LTC Terence R. Szanto, USAF | Appendix A | Glossary | A-10 | Dec-98 |
| Maintainability | "The" ease with which maintenance (routine or failure) can be performed on a complex system. A design consideration focused on ease with which maintenance (routine or failure) can be performed on a complex system. It is a design characteristic whereas "maintenance" is the actual activity of "fixing" or "repairing" a non-functioning system. For example, component standardization (off the shelf replacements), ease of access to critical components, and marking would all be considered in the design phases of a complex system to enhance maintainability. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Maintainability | A characteristic of design and installation, expressed as the probability that an item will be retained in or restored to a specified condition within a given period of time, when the maintenance is performed in accordance with prescribed procedures and resources.<br>The ease with which maintenance of a functional unit can be performed in accordance with prescribed requirements | http://en.wikipedia.org/wiki/Maintainability | | | | Mar-09 |
| Maintainability | The capability of an item to be retained in or restored to specified conditions when maintenance is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance and repair. MTFL, MCMTOMF, and MR are frequently calculated in maintainability evaluations. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 7 | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Maintainability | Maintainability consists of three major areas: time to repair OMFs, total corrective maintenance time, and maintenance burden or maintenance ratio. Maintainability may be expressed as (1) Mean Corrective Maintenance Time for Operational Mission Failure Repairs (MCMTOMF), (2) Mean Corrective Maintenance Time for all incidents (MCMT), (3) Maximum (e.g., 90 Percentile Time) Corrective Maintenance Time for Operational Mission Failures (MaxCMTOMF), (4) Maximum (e.g., 90 Percentile) Corrective Maintenance Time for all incidents (MaxCMT), and (5) various maintenance ratios (MR), e.g., Maintenance Man-Hours Per Operating Hour, Mile, Round, etc. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 7 | b | 4 | Oct-05 |
| Maintainability | The ability of an item to be retained in, or restored to, specified condition when maintenance is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance and repair. [Defense Acquisition University Glossary.] | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | g | 2 | Oct-05 |
| Maintainability - m | Measure example: The qualities of reliability, maintainability, and sustainability built into military systems. DESCRIPTORS: High (few breakdowns and those fixed without major effort) Acceptable (operates at or near established reliability standards maintainable in theater) Low (often breaks down or must be repaired by specialists in rear areas) | CJCSM 3500.04C | Enclosure C | C 2.2.5.3 | C-42 | Jul-02 |
| Maintainability and Maintenance | Maintenance actions include both corrective maintenance (i.e., repairs as a result of failures) and preventive maintenance. The time required for and inherent ease and economy with which a maintenance action can be performed is a direct function of how well maintainability was considered in design.<br>The length of time required for a given maintenance action is also affected by the skill of the maintenance personnel, the maintenance policy and concept, and effectiveness of maintenance manuals and procedures. | DOD Guide for Achieving Reliability, Availability, and Maintainability | 3 | 3.2.4.1 | 70 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Maintenance Ratio | The most common expression for Maintenance Ratio (MR), is Maintenance Man-hours per Operating Hour, which is an indication of the maintenance burden associated with the system. The cumulative number of maintenance man-hours during a given period divided by the cumulative number of operating hours. If appropriate, other terms such as miles or rounds may be substituted for hours. Scheduled as well as corrective maintenance, in keeping with the users maintenance requirements, are included without regard to their effect on mission or availability of the system. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | g | 5 | Oct-05 |
| Manpower Supportability | The identification and acquisition of military and civilian personnel with the skills and grades required to operate and support a material system over its lifetime at peacetime and wartime rates. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | j | 3 | Oct-05 |
| Maximum Corrective Maintenance Time (MaxCMT) | That time below which a specified percentage of all corrective maintenance tasks must be completed. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | f | 5 | Oct-05 |
| Maximum Corrective Maintenance Time for Operational Mission Failures | (MaxCMTOMF): That time below which a specified percentage of corrective maintenance tasks must be completed to restore the system to operation after an Operational Mission Failure. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | e | 5 | Oct-05 |
| Mean Corrective Maintenance Time (MCMT) | The total number of clock-hours of corrective, on-system, active repair time due to all corrective maintenance divided by the total number of incidents requiring corrective maintenance. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | d | 5 | Oct-05 |
| Mean Corrective Maintenance Time for Operational Mission Failures (MCMTOMF) | The total number of clock-hours of corrective, on-system, active repair time, which was used to restore failed systems to mission-capability status after an operational mission failure (OMF) occurs, divided by the total number of OMFs. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | c | 5 | Oct-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Mean Time Between Failures (MTBF) | This measures the reliability for items that are repairable. It can be described as the number of hours that pass before a component, assembly, or system fails. It is a commonly used in discussing the reliability or maintainability for a complex system.<br>For example: If a component has a failure at a rate of 2 failures per million hours, the MTBF would be:<br>MTBF = (1,000,000 hours) / (2 failures) = 500,000 hours | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Mean Time Between Operational Mission Failures (MTBOMF) | The total operating time (e.g., driving time, flying time, or system-on time) divided by the total number of OMFs. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | a | 5 | Oct-05 |
| Mean Time Between Unscheduled Maintenance (MTBUM) | The total operating time divided by the total number of incidents requiring unscheduled maintenance. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | b | 5 | Oct-05 |
| Mean Time to Failure (MTTF) | measures the reliability of parts that are considered to be non-repairable. Therefore, MTTF is the mean Time it takes to reach the first failure for a non-repairable system or component. Since MTTF is a mean value, it is intended to represent failure statistically over a large number of similar systems/components. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Modularity | Modular programming is a software design technique that increases the extent to which software is composed from separate parts, called modules. Conceptually, modules represent a separation of concerns, and improve maintainability by enforcing logical boundaries between components. Modules are typically incorporated into the program through interfaces. A module interface expresses the elements that are provided and required by the module. The elements defined in the interface are visible to other modules. The implementation contains the working code that corresponds to the elements declared in the interface. | http://en.wikipedia.org/wiki/Modularity_(programming) | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Multi-Target Degradation Factor (MTDF): Joint Terminal Attack Controller (JTAC) | In scenarios containing more than one target, the JTAC's ability to detect multiple targets may be degraded to represent the difficulty associated with simultaneous engagements. The MTDF is a factor between 0 and 1 applied multiplicatively to the Pd. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.d | 19 | Jun-07 |
| Natural Environmental Effects and Impacts | (1) Environment. Includes the air, water, land, plants, animals, and other living organisms, man-made structures, historical and cultural resources, and the interrelationships that exist among them and with people. [The Defense Acquisition University Glossary.] (2) Environmental Effects. The effects of the natural environment on the system. For example, corrosion is a natural environmental effect caused by weather, ocean conditions, etc. (3) Environmental Impacts. The system's impact on the natural environment as a result of its operational use, maintenance, transportation, and storage. For example, impacts include pollution (noise, air, and water), threat to endangered species, threat to public health, etc. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | 1 | 3 | Oct-05 |
| Networked | All joint force elements will be connected and synchronized in time and purpose to facilitate integrated and interdependent operations across the global battlespace. A networked joint force can extend the benefits of decentralization--initiative, adaptability, and increased tempo--without sacrificing the coordination or unity of effort emblematic of centralization. The joint force will capitalize on being networked by making user-defined information and expertise available anywhere within the network, and will exploit network connectivity among dispersed joint force elements to improve information sharing, collaboration, coordinated maneuver, and integrated situational awareness. Networks should extend to interagency and multinational partners, where possible, to support and enhance unified action. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |
| Operability | Operability is the ability to keep a system in a functioning and operating condition. Operability is considered one of the ilities and is closely related to supportability and maintainability. | http://en.wikipedia.org/wiki/Operability | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Operational Availability (Ao) | The degree (expressed as a decimal between 0 and 1, or the percentage equivalent) to which one can expect a piece of equipment or weapon system to work properly when it is required. Operational Availability is calculated by dividing uptime by the sum of uptime and downtime. It is the quantitative link between readiness objectives and supportability. [Defense Acquisition University Glossary.] It can also be calculated by the number of systems that are ready, divided by the number possessed (e.g., the number of times the system was available, divided by the number of times the system was required) for on-demand systems. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 8 | h | 6 | Oct-05 |
| Operational Availability (Ao) - m | Measure example: Hours after CJCS Warning Order to determine availability of suitable munitions within theater. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 4.1 | B-C-C-71 | Jul-02 |
| Operational Capability | An OC is a system attribute or grouping of attributes that users and subject matter experts have identified as being crucial to the achievement of critical mission elements and/or operational objectives and are, therefore, of significant value to the warfighter. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-23 | Jun-07 |
| Operational effectiveness | The overall degree of mission accomplishment of a system when used by representative personnel in the expected (or planned) environment. Some examples of environment are: natural, electronic, threat, and so forth for operational employment of the system considering organization, doctrine, tactics, survivability, vulnerability, and threat (including countermeasures; initial nuclear weapons effects; nuclear, biological, and chemical contamination threats). (Army Regulation 73–1) | Army DT/OT&E Policy and Guidance (Briefing) | | | Slide 3 | Apr-06 |
| Operational Effectiveness | Measure of the overall ability to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, supportability, survivability, vulnerability, and threat. | CJCSI 3170.01F | Glossary | GL | GL-17 | May-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Operational Reliability | The probability that an operationally ready system will react as required to accomplish its intended mission or function as planned, excluding the effects of enemy action, may be specified as an estimated or an achieved reliability. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-24 | Jun-07 |
| Operational Reliability - m | Measure example: Percent of C4 support systems meet command reliability standards. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 5.1 | B-C-C-97 | Jul-02 |
| Operational Sufficiency | The breadth of the operations addressed while employing new or modified capabilities within the context of representative employment and support concepts. The evaluation is considered operationally sufficient if it provides both the decision maker and warfighter with results drawn from test events executed across sufficient operational conditions to identify the capabilities and limitations associated with employment as well as determine effectiveness, suitability. Test/core teams address operational sufficiency during Initial Test Planning and test planning by applying operational perspective and experience while designing test events to ensure capabilities and concepts are adequately addressed within the planned evaluation. Test/core teams address operational sufficiency during test execution and reporting by using operational perspective and experience to apply judgment and ensure events executed, and results observed, are evaluated within the context of envisioned operations and support characterizing the capabilities provided and operational effects achieved. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-25 | Jun-07 |
| Operational Sufficiency - m | Measure example: Yes/No; Forecast branches reviewed for legal sufficiency. | CJCSM 3500.04C 1 July 2002 | Annex C Appendix C Enclosure B | OP 5.2.1 | B-C-C-106 | Jul-02 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Operational Suitability | The degree to which a system can be placed satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistics supportability, natural environmental effects and impacts, documentation, and training requirements. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-25 | Jun-07 |
| Operational Suitability | The degree to which a system can be satisfactorily placed in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistic supportability, and training requirements. (Army Regulation 73–1) | Army DT/OT&E Policy and Guidance (Briefing) | | | Slide 3 | Apr-06 |
| Operational Suitability | The degree to which a system can be satisfactorily placed in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistic supportability, and training requirements. | Army Regulation 73-1 Test and Evaluation Test and Evaluation Policy | Glossary | Section II | 56 | Apr-06 |
| Operational Suitability | The degree to which a system can be placed satisfactorily in field use with consideration given to reliability, maintainability, availability, logistic supportability, compatibility, interoperability, training, human factors, safety, documentation, transportability, wartime usage rates, manning requirements, and natural and environmental effects and impacts. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 11 | Mar-09 |
| Operational Suitability - m | Measure example: Percent of Service DT&E/OT&E test facility, simulation, and modeling capabilities subject to systems and processes to minimize duplication. | CJCSM 3500.04C | Annex A Appendix C Enclosure B | SN 7.2.4 | B-C-A-126 | Jul-02 |
| Portability | The ability of the product to be transferred from one environment to another. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Portability | Portability is the general characteristic of being readily transportable from one location to another, and may specifically refer to:<br>- Portability (social security), the portability of social security benefits<br>- Software portability, the portability of a piece of software to multiple platforms<br>- Telephone number portability keeping one telephone number while switching ones account to another telephony provider.<br>- Porting, the ability of a program to be ported from one system to another in computer science | http://en.wikipedia.org/wiki/Portability | | | | Mar-09 |
| Precise | The ability to act directly upon key elements and processes demands precisely executed joint actions. Precision extends beyond surgical strikes to the exact application of all joint force capabilities to achieve greater success at less risk. Knowledge gained in all dimensions will enhance the capability of the JFC to understand a situation, determine the effects desired, select a course of action and the forces to execute it, accurately assess the effects of that action and reengage as necessary. Regardless of its application in combat or noncombat operations, the capability to engage precisely allows commanders to shape situations or battlespace in order to generate the desired effects while minimizing unintended effects and contributing to the most effective use of resources. The overall effect of precision is far-reaching with considerable payoff in terms of combat effectiveness. | CCJO | 4. Solution Recom-mended Practices Guide | 4.E.5 | 22 | Aug-05 |
| Precise - m | Measure example: Minutes to derive precise/mensurated DMPI/DPI coordinates for preplanned and NRT munitions employment. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 2.2.5 | B-C-C-40 | Jul-02 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Precision | Precision has the following meanings: Concepts<br>Accuracy and precision, measurement deviation from true value and its scatter<br>Precision (arithmetic), the number of digits from which a value is expressed<br>Precision (information retrieval), the percentage of documents returned that are relevant<br>Precision (computer science), a measure of the detail in which a quantity is expressed | http://en.wikipedia.org/wiki/Precision | | | | Mar-09 |
| Precision | The quality or state of being clearly depicted, definite, measured or calculated. A quality associated with the spread of data obtained in repetitions of an experiment as measured by variance; the lower the variance, the higher the precision A measure of how meticulously or rigorously computational processes are described or performed by a model or simulation. | Recommended Practices Guide | Fidelity RPG | Special Topic | 8 | Sep-00 |
| Predictability | Predictability (also called banality) is the degree to which a correct prediction or forecast of a system's state can be made either qualitatively or quantitatively. Although the second law of thermodynamics can tell us about the equilibrium state that a system will evolve to, and steady states in dissipative systems can sometimes be predicted, there exists no general rule to predict the time evolution of systems far from equilibrium, e.g. chaotic systems, if they do not approach some kind of equilibrium. Their predictability usually deteriorates with time. To quantify predictability, the rate of divergence of system trajectories in phase space can be measured (Kolmogorov-Sinai entropy, Lyapunov exponents). | http://en.wikipedia.org/wiki/Predictability | | | | Mar-09 |
| Probability of Detection (Pd): Network Enabled Weapon (NEW) | Once in autonomous seeker mode, the NEW relies on its own Pd todetermine if it senses the target during a particular time step. Unlike the JTAC, however,the Pd is always a random draw, according to the parameter setting. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.e | 21 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Probability of Detection (Pd): Joint Terminal Attack Controller (JTAC) | Probability of Detection (Pd): Targets within the JTAC's sensor range are detected according to the assigned Pd. During each model time step, a random draw determines whether or not the target is detected. The logic within the model can be adjusted such that the Pd becomes one (certainty), after initial detection occurs, to simulate focused attention on the target. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.c | 18 | Jun-07 |
| Probability of Kill (Pk): Network Enabled Weapon (NEW) | Combined with the impact radius, the NEW's Pk value determines whether or not a target is killed. The Pk is limited to very high (i.e., favorable) values to replicate the likelihood of correct weapon detonation upon impact. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.f | 21 | Jun-07 |
| Producability | The ease with which a complex system can be produced. A design consideration focused on the enhancing the ease with which a complex system can be produced. For example, placement of components in a configuration that allows ease of build and replacement of failed components during system build. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Producibility | The relative ease of manufacturing an item or system. This relative ease is governed by the characteristics and features of a design that enables economical fabrication, assembly, inspection and testing using available manufacturing techniques. (reference DSMC Glossary of Defense Acquisition Acronyms and Terms, 8th Edition) | Defense Systems Management College Simulation Based Acquisition: A New Approach LTC Michael V. R. Johnson, Sr., USA LTC Mark F. McKeon, USMC LTC Terence R. Szanto, USAF | Appendix A | Glossary | A-13 | Dec-98 |
| Readiness | State of preparedness of forces or weapon system or systems to meet a mission or to warfight. Based on adequate and trained personnel, material condition, supplies/reserves of support system and ammunition, numbers of units available, etc. [Defense Acquisition University Glossary]. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 6 | a | 4 | Oct-05 |
| Readiness - m | Measure example: Percent Of forces supplied by theater commander, rated C1 for readiness and no forces supplied by theater commander, below C2. | CJCSM 3500.04C | Annex B Appendix C Enclosure B | ST 4.2 | B-C-B-57 | Jul-02 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Recoverability | Following combat damage, the ability to take emergency action to prevent loss of the system, to reduce personnel casualties, or to regain weapon system combat mission capabilities. Recoverability is considered a subset of survivability. (Defense Acquisition Desk book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-29 | Jun-07 |
| Recoverability - m | Measure example: Minutes to develop, maintain, and update target databases to includeestimates of target recoverability, resupply, and regeneration. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 2.2.5 | B-C-C-40 | Jul-02 |
| relevance: characteristics of a good model | Relevance is determined by how well a model addresses the substantive cost-effectiveness issues in the trade study. | NASA Systems Engineering Handbook SP-610S | 5 | 5.1.2 | 86 | Jun-95 |
| Reliability | The ability of a system and its parts to perform its mission without failure, degradation, or demand on the support system. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-29 | Jun-07 |
| Reliability | The ability of a system and its parts to perform its mission without failure, degradation, or demand on the support system. (reference DSMC Glossary of Defense Acquisition Acronyms and Terms, 8th Edition) | Defense Systems Management College Simulation Based Acquisition: A New Approach LTC Michael V. R. Johnson, Sr., USA LTC Mark F. McKeon, USMC LTC Terence R. Szanto, USAF | Appendix A | Glossary | A-13 | Dec-98 |
| Reliability | Mission and non-mission failures that require repair. The lower limit on the number of failures is determined by the inherent level of reliability designed and built into the system. However, poor manufacturing, inadequate maintenance, operations in conditions beyond those specified for the design, and "acts of God" can increase the number. In addition to determining a lower bound on failures, the reliability characteristics of an item should be considered in determining the number and types of preventive maintenance actions that are either required or are economically desirable. | DOD Guide for Achieving Reliability, Availability, and Maintainability | 3 | 3.2.4.1 | 70 | Aug-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Reliability | A design consideration focused on the ability of a component or system to perform its intended function. Quantitatively, reliability is the probability of a complex system satisfactorily performing its purpose during a given time period under specified operating conditions. Probability is generally considered to be quantitatively derived through historical or test data. Satisfactory performance involves the system producing specified "outputs", considered to be indicative of success, over a time period. The probability of a complex system satisfactorily performing its purpose during a given time period under specified operating conditions. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Reliability | In general, reliability (systemic def.) is the ability of a person or system to perform and maintain its functions in routine circumstances, as well as hostile or unexpected circumstances.<br><br>The IEEE defines it as ". . . the ability of a system or component to perform its required functions under stated conditions for a specified period of time." | http://en.wikipedia.org/wiki/Reliability | | | | Mar-09 |
| Reliability | The duration or probability of failure-free performance under stated conditions. In OT&E, reliability is usually reported in one of two ways:• Mission Reliability (R). For equipment operated only during a relatively short duration mission (as opposed to equipment operated more or less continuously), the probability of completing the mission without an operational mission failure.• MTBOMF. Mean time between operational mission failures. For more or less continuously operated equipment or systems. MTBOMF measures reliability as it relates to the overall mission of the equipment or system being tested and is the total operating time divided by the number of operational mission failures. MTBOMF is the figure used in the calculation of overall mission Reliability (R). MTBOMF is sometimes modified to Mean Flight Hours Between Operational Mission Failures (MFHBOMF). | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 13 | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Reliability | Reliability consists of two major areas: mission reliability and logistics related reliability. (1) Mission Reliability. The probability that a system will perform its required mission critical functions for the duration of a specified mission under conditions stated in the mission profile. [Defense Acquisition University Glossary.] Mission reliability can also be stated as the probability a system can complete its required operational mission without an operational mission failure (OMF). An OMF is a failure that prevents the system from performing one or more mission essential functions. For some systems, mission reliability may be better expressed as a function of Mean Time (miles, rounds, etc.) Between Operational Mission Failure (MTBOMF). (See paragraph 8 for definition.) | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 7 | a | 4 | Oct-05 |
| Reliability | Reliability consists of two major areas: mission reliability and logistics related reliability. (2) Logistics Reliability. The measure of the ability of an item to operate without placing a demand on the LS structure for repair or adjustment. Logistics reliability recognizes the effects of occurrences that place a demand on the LS structure without regard to the effect on mission or function. (Defense Acquisition University Glossary.) Measures of logistics reliability include the probability that no corrective (or unscheduled) maintenance, unscheduled removals, or unscheduled demands for spare parts will occur following the completion of a specific mission profile. Logistics reliability may be expressed as a function of Mean Time Between Unscheduled Maintenance (MTBUM) (see paragraph 8 for definition) and/or Mean Time Between Demand for Spares. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 7 | a | 4 | Oct-05 |
| Reliability | The ability of a system and its parts to perform its mission without failure, degradation, or demand on the support system (Defense Acquisition University Glossary). | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | e | 2 | Oct-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Reliability - m | Measure example: Percent of C4 support systems meet command reliability standards. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 5.1 | B-C-C-97 | Jul-02 |
| Reliability: Weapon Control Network (WCN) | The reliability of the network is a parameter that governs the random probability of a successful message transmission. For the purposes of this research, messages transmitted are either fully received or not at all—there are no instances of partial (or garbled) communications. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.5.a | 23 | Jun-07 |
| Repeatability | Repeatability is the variation in measurements taken by a single person or instrument on the same item and under the same conditions. A measurement may be said to be repeatable when this variation is smaller than some agreed limit. According to the Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, repeatability conditions include:the same measurement procedure the same observer the same measuring instrument, used under the same conditions the same location repetition over a short period of time. Repeatability methods were developed by Bland and Altman (1986). The repeatability coefficient is a precision measure which represents the value below which the absolute difference between two repeated test results may be expected to lie with a probability of 95%. The standard deviation under repeatability conditions is part of precision and accuracy. | http://en.wikipedia.org/wiki/Repeatability | | | | Mar-09 |
| Repeatability | The degree to which different groups of JTEM M&P users demonstrate similar actions and produce similar output products) | JTEM Effectiveness and Suitability Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |
| Reproducibility | backups and restores | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Reproducibility | Reproducibility is one of the main principles of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently. | http://en.wikipedia.org/wiki/Reproducibility | | | | Mar-09 |
| Resilient | To operate successfully, the future joint force must be able to protect and sustain its capabilities from the effects of adversaries or adverse conditions. It must also be able to withstand pressure or absorb punishment without permanently losing its focus, structure, momentum, or integrity. Resilience provides joint forces with the ability to sustain performance at high levels, despite losses, setbacks, or similar developments. The future joint force must be resilient to meet the demands of being successful across the ROMO [Range of Military Operations] in an uncertain future security environment. | CCJO | 4. Solution | 4.E.5 | 22 | Aug-05 |
| Resources | Resources include the number of maintenance personnel available as well as the number and availability of spare and repair parts, support equipment, repair manuals, tools, etc. | DOD Guide for Achieving Reliability, Availability, and Maintainability | 3 | 3.2.4.1 | 70 | Aug-05 |
| Resources - m | Measure example: Percent of required C4 systems resources identified. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 5.1.2 | B-C-C-99 | Jul-02 |
| Responsiveness | The responsiveness of an interactive system describes how quickly it responds to user input (i.e. the rate of communication with the system). It is one of the criteria under the principle of robustness (from a usability principle). The other three are observability, recoverability, and task conformance.<br>A good example of a responsiveness of an interactive system is the use of an Internet browser. When requesting a page from a website (user input), it usually takes time for the web server to send all the page contents to the client browser. The speed of transmission of all the page data to the browser would be the responsiveness of the system. | http://en.wikipedia.org/wiki/Responsiveness | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| responsiveness: characteristics of a good model | Responsiveness of a model is a measure of its power to distinguish among the different alternatives being considered in a trade study. A responsive lunar base cost model, for example, should be able to distinguish the costs associated with different system architectures or designs, operations concepts, or logistics strategies. | NASA Systems Engineering Handbook SP-610S | 5 | 5.1.2 | 87 | Jun-95 |
| Reusability | In computer science and software engineering, reusability is the likelihood a segment of source code can be used again to add new functionalities with slight or no modification. Reusable modules and classes reduce implementation time, increase the likelihood that prior testing and use has eliminated bugs and localizes code modifications when a change in implementation is required. Reusability implies some explicit management of build, packaging, distribution, installation, configuration, deployment, maintenance and upgrade issues. If these issues are not considered, software may appear to be reusable from design point of view, but will not be reused in practice. | http://en.wikipedia.org/wiki/Reusability | | | | Mar-09 |
| Robustness | Robustness is the quality of being able to withstand stresses, pressures, or changes in procedure or circumstance. A system, organism or design may be said to be "robust" if it is capable of coping well with variations (sometimes unpredictable variations) in its operating environment with minimal damage, alteration or loss of functionality. | http://en.wikipedia.org/wiki/Robustness | | | | Mar-09 |
| rules: Designing Decision Support Systems for Revolutionary Command and Control Domains | Whereas SBBs are generally reactive, the next level of worker competencies, the rule-based behaviors (RBBs) category generally focuses on actions that require more than a simple or well-rehearsed reaction. A RBB occurs when an actor makes decisions or takes action as a result of a prescribed set of rules. Based on experience, what is a RBB may be a SBB for another person. | Designing Decision Support Systems for Revolutionary Command and Control Domains A Dissertation to the faculty of the School of Engineering and Applied Science, University of Virginia, Mary L. Cummings | 2 | 2.1.5 | 41 | Jan-04 |
| Safety | The ability of a component or system intended to eliminate or reduce the risk of injury, incapacitation, or death to humans interacting within complex systems. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Safety | Safety is the state of being "safe" (from French sauf), the condition of being protected against physical, social, spiritual, financial, political, emotional, occupational, psychological, educational or other types or consequences of failure, damage, error, accidents, harm or any other event which could be considered non-desirable. This can take the form of being protected from the event or from exposure to something that causes health or economical losses. It can include protection of people or of possessions. | http://en.wikipedia.org/wiki/Safety | | | | Mar-09 |
| Safety | Freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. [Defense Acquisition University Glossary] | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | h | 3 | Oct-05 |
| Safety - m | Measure example: Percent of OPLANs address time-phased operating and safety levels (of supply). | CJCSM 3500.04C | Annex B Appendix C Enclosure B | ST 4.3.2.2 | B-C-B-69 | Jul-02 |
| Saftey | A design consideration focused on the ability of a component or system intended to eliminate or reduce the risk of injury, incapacitation, or death to humans interacting within complex systems. For example, inclusion of a failsafe design to prevent electrical shock to system maintainers. | How many “ilities” are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Scalability | In telecommunications and software engineering, scalability is a desirable property of a system, a network, or a process, which indicates its ability to either handle growing amounts of work in a graceful manner, or to be readily enlarged.[1] For example, it can refer to the capability of a system to increase total throughput under an increased load when resources (typically hardware) are added. An analogous meaning is implied when the word is used in a commercial context, where scalability of a company implies that the underlying business model offers the potential for economic growth within the company. André B. Bondi, 'Characteristics of scalability and their impact on performance', Proceedings of the 2nd international workshop on Software and performance, Ottawa, Ontario, Canada, 2000, ISBN 1-58113-195-X, pages 195 - 203 | http://en.wikipedia.org/wiki/Scalability#cite_note-0 | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Seamlessness | The quality of having no visible seams, often applied idiomatically to any construct, even the most vague (i.e. "seamless networking"). Seamlessness is a system quality attribute, often applied to converging or merged technologies, that refers to the degree to which the technologies present a consistent structure and paradigm in interfaces and operations, so that the transition from one technology to another is not disruptive or confusing either in usage or integration. | http://en.wikipedia.org/wiki/Seamlessness | | | | Mar-09 |
| securability | securability (countable and uncountable; plural securabilities) (uncountable) The characteristic of being securable (countable) The extent to which something is securable; especially the ability of a system to provide different levels of secure access | http://en.wiktionary.org/wiki/securability | | | | Mar-09 |
| Sensor Range: Network Enabled Weapon (NEW) | The sensor range identifies the point at which the NEW acquires autonomous recognition of the target. The performance within the model is similar to the "cookie cutter" approach described for the JTAC agent. Once the NEW is within this three-dimensional range to the target, it no longer receives IFTUs through the network; instead relying on its own sensor characteristics to guide it to the target. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.c | 20 | Jun-07 |
| Sensor Range: Joint Terminal Attack Controller (JTAC) | The JTAC sensor range is a radial distance defining a circular representation of the agent's ability to identify and track targets. The sensor characteristics are commonly referred to as "cookie cutter," indicating a defined probability of detection for targets within range and a probability of zero for targets outside the range. Sensor range is measured in meters. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.b | 18 | Jun-07 |
| Serviceability (computer) | In software engineering and hardware engineering, serviceability (also known as supportability,) is one of the -ilities or aspects (from IBM's RASU (Reliability, Availability, Serviceability, and Usability)). It refers to the ability of technical support personnel to install, configure, and monitor computer products, identify exceptions or faults, debug or isolate faults to root cause analysis, and provide hardware or software maintenance in pursuit of solving a problem and restoring the product into service. Incorporating serviceability facilitating features typically results in more efficient product maintenance and reduces operational costs and maintains business continuity. | http://en.wikipedia.org/wiki/Serviceability_(computer) | | | | Mar-09 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Simplicity | Simplicity is the property, condition, or quality of being simple or un-combined. It often denotes beauty, purity or clarity. Simple things are usually easier to explain and understand than complicated ones. Simplicity can mean freedom from hardship, effort or confusion. It may also refer to a simple living lifestyle. Can also be the basis, or bare foundation to terms of understanding, proving no depth or intricate reasoning. | http://en.wikipedia.org/wiki/Simplicity | | | | Mar-09 |
| skills: Designing Decision Support Systems for Revolutionary Command and Control Domains | The first level of worker competencies is known as skill-based behavior (SBB). On the lowest level of the cognitive continuum, a SBB results when a worker takes some action that is reflexive and without conscious planning. | Designing Decision Support Systems for Revolutionary Command and Control Domains A Dissertation to the faculty of the School of Engineering and Applied Science, University of Virginia, Mary L. Cummings | 2 | 2.1.5 | 41 | Jan-04 |
| Speed: Network Enabled Weapon (NEW) | The true speed of flight is, again, a complicated physical process to model. The simplifying approach for this study is to vary the speed as a parameter across a typical range between maximum glide performance and minimum time to target. Every effort was made to consult with Subject Matter Experts (SMEs) in order to accurately represent weapon performance. Speed is measured in meters/second. The author and model developer consulted in person with SMEs from Joint Command and Control for NEW and SDB II program offices, Eglin AFB, on 26 April 2007. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.a | 20 | Jun-07 |
| Speed: Joint Terminal Attack Controller (JTAC) | The JTAC is typically a ground-based controller deployed within the immediate vicinity of potential or known targets. He possesses sensor capabilities, communication devices, and equipment necessary to locate, identify, and track targets. If necessary, the JTAC can be assigned waypoints from his initial location to traverse during execution of the scenario. The speed setting controls how fast the JTAC moves between the assigned waypoints.<br>Speed is measured in meters/second. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.a | 18 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Supportability | The degree to which system design characteristics and planned logistics resources, including manpower, meet system peacetime readiness and wartime utilization requirements. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-30 | Jun-07 |
| Supportability | The degree of ease to which system design characteristics and planned logistics for resources, including the logistic support elements, allows for the meeting of system availability and wartime utilization requirements. (reference DSMC Glossary of Defense Acquisition Acronyms and Terms, 8th Edition) | Defense Systems Management College Simulation Based Acquisition: A New Approach LTC Michael V. R. Johnson, Sr., USA LTC Mark F. McKeon, USMC LTC Terence R. Szanto, USAF | Appendix A | Glossary | A-14 | Dec-98 |
| Supportability | The ability of a component to interface with the infrastructure necessary to provide support for service and support. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Supportability | A design consideration focused on the ability of a component to interface with the infrastructure necessary to provide support for service and support. This includes such considerations as manpower, logistics, personal training, test and support equipment, storage, etc. For example, construction of a major complex system to "move dirt" that cannot be fueled locally at the site of use. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Supportability - m | Measure example: Days to determine supportability of allied support request. | CJCSM 3500.04C | Annex B Appendix C Enclosure B | ST 8.1.2 | B-C-B-134 | Jul-02 |
| Supporting Measure of Suitability | Measure of system performance reflecting the degree to which a system can be placed satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistics supportability, natural environmental effects and impacts, documentation, and training requirements. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-31 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Survivability | The capability of a system and its crew to avoid or withstand man-made hostile environments without suffering an abortive impairment of its ability to accomplish its designated mission. Survivability is comprised of susceptibility, vulnerability, and recoverability. (Defense Acquisition Desk Book) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-31 | Jun-07 |
| Survivability | The capability of a system and crew to avoid or withstand manmade | Army Regulation 73-1 Test and Evaluation Test and Evaluation Policy | Glossary | Section II | 58 | Apr-06 |
| Survivability | The ability of a system to withstand external threats to continued performance of designed mission. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Survivability | A design consideration focused on the ability of a system to withstand external threats to continued performance of designed mission. For example, placing additional external shielding and software design to prevent external intrusion in a network system. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Survivability | Survivability is the ability to remain alive or continue to exist. The term has more specific meaning in certain contexts. | http://en.wikipedia.org/wiki/Survivability | | | | Mar-09 |
| Survivability | The capability of a system to avoid or withstand manmade, hostile environments without suffering an abortive impairment of its capability to accomplish its designated mission. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 14 | Mar-09 |
| Survivability - m | Measure example: Hours to disperse forces to ensure survivability. | CJCSM 3500.04C | Annex A Appendix C Enclosure B | SN 3.2.2 | B-C-A-44 | Jul-02 |
| Susceptibility | The degree to which a weapon system is open to effective attack due to one or more inherent weaknesses. (Susceptibility is a function of operational tactics, countermeasures, and probability of the enemy fielding a threat.) Susceptibility is considered a subset of survivability. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-31 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Susceptibility | The degree to which a weapon system is open to effective attack due to one or more inherent weaknesses. Susceptibility is a function of operational tactics, countermeasures, probability of enemy fielding a threat, and so forth. Susceptibility is considered a subset of survivability. | Army Regulation 73-1 Test and Evaluation Test and Evaluation Policy | Glossary | Section II | 58 | Apr-06 |
| Susceptibility | The degree to which a device, equipment, or weapons system is open to effective attack due to one or more inherent weaknesses. (Susceptibility is a function of operational tactics, countermeasures, probability of the enemy fielding a threat, etc.) Susceptibility is considered a subset of survivability. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 14 | Mar-09 |
| Sustainability | Sustainability, in a broad sense, is the ability to maintain a certain process or state. It is now most frequently used in connection with biological and human systems. | http://en.wikipedia.org/wiki/Sustainability | | | | Mar-09 |
| Sustainability | The ability to maintain the necessary level and duration of operational activity to achieve military objectives. Sustainability is a function of providing for and maintaining those levels of ready forces, materiel and consumables necessary to support military effort. (Defense Acquisition University Glossary | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 6 | b | 4 | Oct-05 |
| Target Location Error (TLE): Joint Terminal Attack Controller (JTAC) | TLE defines the accuracy associated with the JTAC's perception of the true location of the target. It is represented in the model by a Bivariate Normal distribution and is defined by two parameters, the mean and standard deviation. Jay L. Devore, Probability and Statistics for Engineering and the Sciences, 6th Edition, 2004, Thomson Brooks/Cole Publishers, 2004, p. 543. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.1.e | 19 | Jun-07 |
| Target Location Error: Network Enabled Weapon (NEW) | The TLE is identical to the process described for the JTAC. The Gaussian mean and standard deviation parameters, however, will take on a different range of values in order to more closely represent NEW sensor capabilities. It is represented in the model by a Bivariate Normal distribution and is defined by two parameters, the mean and standard deviation. Jay L. Devore, Probability and Statistics for Engineering and the Sciences, 6th Edition, 2004, Thomson Brooks/Cole Publishers, 2004, p. 543. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.2.d | 20 | Jun-07 |
| Target: Pop-Up Time | Each target in the scenario may be designated a particular time step in which it first appears. The pop-up time presents the opportunity to investigate system effectiveness against a HVT that may result in a redirecting a NEW already in flight. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.3.b | 21 | Jun-07 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Target: Speed | For each scenario, the target(s) are assigned predetermined waypoints to traverse. The target may be set to cycle through the waypoints continuously or stop when it reaches the last one. The target speed is used to differentiate between categories of targets. For example, vehicular targets typically travel at much greater speeds than dismounted (i.e., foot patrol) targets. Speed is measured in meters/second. | System of Systems Test Planning in a Complex Joint Environment NPGS Thesis Christopher M. Wegner | 3 | B.3.a | 21 | Jun-07 |
| Technical Credibility | Addresses the depth of the technical information produced by the test. Technical credibility can be determined only after test completion. A technically credible test provides the decision maker and the warfighter with an indication of decision risk. Decision risk should be addressed by characterizing the weapon system capabilities with the likelihood of an operational event happening and the consequences of the event's occurrence. Sample size, confidence bounds, and repeatability are important considerations. | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-33 | Jun-07 |
| Testability | The ease and efficiency with which self-testing and external testing can be accomplished | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Testability | A design consideration focused on the ease and efficiency with which self-testing and external testing can be accomplished. For example, designing a complex system that measures wear of critical components and alerts system operators of maintenance need prior to system failure. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Testability | Testability, a property applying to an empirical hypothesis, involves two components: (1) the logical property that is variously described as contingency, defeasibility, or falsifiability, which means that counterexamples to the hypothesis are logically possible, and (2) the practical feasibility of observing a reproducible series of such counterexamples if they do exist. In short, a hypothesis is testable if there is some real hope of deciding whether it is true or false of real experience. Upon this property of its constituent hypotheses rests the ability to decide whether a theory can be confirmed or falsified by the data of actual experience. It means it can be tested, but it also can be proven wrong later.<br>In engineering this refers to the capability of an equipment or system to be tested | http://en.wikipedia.org/wiki/Testability | | | | Mar-09 |
| Timeliness | The latency of performing JTEM processes or the amount of processes performed in a work period – not a Rock Drill focus) | JTEM Effectiveness and Suitability<br>Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |
| Training and Training Support | The processes, procedures, techniques, training devices, and equipment used to train civilian and active duty and reserve military personnel to operate and support a materiel system. This includes individual and crew training; new equipment training; initial, formal, and On-The-Job (OJT) training; and Logistics Support (LS) planning for training equipment and training device acquisitions and installations. A traditional element of LS. (Defense Acquisition University Glossary.) | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | n | 3 | Oct-05 |
| Transparency: characteristics of a good model | Transparency...occurs when the model's mathematical relationships, algorithms, parameters, supporting data, and inner workings are open to the user. The benefit of this visibility is in the traceability of the model's results. Not everyone may agree with the results, but at least they know how they were derived. Transparency also aids in the acceptance process. It is easier for a model to be accepted when its documentation is complete and open for comment. Proprietary models often suffer from a lack of acceptance because of a lack of transparency. | NASA<br>Systems Engineering Handbook<br>SP-610S | 5 | 5.1.2 | 87 | Jun-95 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Transportability | The capability of materiel to be moved by towing, self propulsion, or carrier via any means such as railways, highways, waterways, pipelines, oceans, and airways. (Joint Pub 1-02) | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-35 | Jun-07 |
| Transportability | Percent of TPFDD sealifted and air units closed by LAD. Percent of units close by LAD. Percent of sustainment movement requirements met by available sealift (during execution). Hours to determine transportation feasibility. Hours to identify available common-user lift assets worldwide to support critical, short-notice requirements in support of national or theater military strategies. Hours to identify requirements for lift assets (in crisis situation). Months from approval of JSCP to approved OPLAN/TPFDD. Percent of airlift requirements miss LAD because of unforeseen lack of enroute support. | CJCSM 3500.04C | Annex A, Appendix C | Enclosure B | B-C-A-8,9 | Jul-02 |
| Transportability | The ease with which a complex system, requiring movement, can be relocated, disassembled, packed, etc. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Transportability | A design consideration focused on the ease with which a complex system, requiring movement, can be relocated, disassembled, packed, etc. For example, a major assemble might be designed such that it is modularized – to provide for ease of movement of the system. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Transportability | The capability of materiel/personnel to be moved by towing, self propulsion or carrier via any means, such as railways, highways, waterways, pipelines, oceans, and airways. Full consideration of available and projected transportation assets, mobility plans and schedules and the impact of system equipment and support items on the strategic mobility of operating military forces are required to achieve this capability. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | c | 2 | Oct-05 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Understanding | Understanding (also called intellection) is a psychological process related to an abstract or physical object, such as a person, situation, or message whereby one is able to think about it and use concepts to deal adequately with that object.<br><br>An understanding is the limit of a conceptualization. To understand something is to have conceptualized it to a given measure. | http://en.wikipedia.org/wiki/Understandability | | | | Mar-09 |
| Usability | The design of a component or system for human use. A design consideration focused on the design of a component or system for human use (human factors engineering). Human considerations include the following characteristics:<br>Anthropometric factors – dealing with the dimensions and characteristics of the human body (example, height, weight, arm reach, etc.). For example, chair height adjustment to compensate for variation in operator heights.<br>Human Sensory factors – dealing with the senses (sight, hearing, touch, smell, taste) as they relate to the design of a complex system. For example, color coding of equipment for ease of operation (assumes that the people interfacing with the system must not be color blind).<br>Physiological factors – considers the range of stresses on the human body during the performance of tasks related to operation of a complex system. For example, excessive heat generation from a machine can influence system performance by reducing human capacity to work.<br>Psychological factors – considers the range of considerations that pertain to the human mind in performance of work in a complex system (emotions, behaviors, attitudinal responses, etc.). For example, providing additional operator "control" to prevent inattentiveness in operation of a complex system. | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Usability | Usability is a term used to denote the ease with which people can employ a particular tool or other human-made object in order to achieve a particular goal. Usability can also refer to the methods of measuring usability and the study of the principles behind an object's perceived efficiency or elegance.<br>In human-computer interaction and computer science, usability usually refers to the elegance and clarity with which the interaction with a computer program or a web site is designed. The term is also used often in the context of products like consumer electronics, or in the areas of communication, and knowledge transfer objects (such as a cookbook, a document or online help). It can also refer to the efficient design of mechanical objects such as a door handle or a hammer. | http://en.wikipedia.org/wiki/Usability | | | | Mar-09 |
| Usage Rates | (1) Wartime Usage Rates. The quantitative statement of the projected manner in which the system is to be used in its intended wartime environment. (2) Peacetime Usage Rates. The quantitative statement of the projected manner in which the system is to be used in its intended peacetime environment. | MOA on Operational Suitability Terminology & Definitions to be used in Operational Test & Evaluation (OT&E) | 5 | f | 2 | Oct-05 |
| Usage Rates - m | Measure example: Percent of end items completely depleted under expected usage rates before industry reopens production. | CJCSM 3500.04C | Annex A Appendix C Enclosure B | SN 4.2.3 | B-C-A-73 | Jul-02 |
| Useability of process documentation | Product template and instruction/guidebook usefulness, helpfulness, and simplicity | JTEM Effectiveness and Suitability Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |
| User friendliness: characteristics of a good model | Upfront user friendliness is related to the ease with which the system engineer can learn to use the model and prepare the inputs to it. Backend user friendliness is related to the effort needed to interpret the model's results and to prepare trade study reports for the tentative selection using the selection rule. | NASA Systems Engineering Handbook SP-610S | 5 | 5.1.2 | 87 | Jun-95 |

| Term | Definition | Source | Chapter | Section or Paragraph | Page # | Source Date |
|---|---|---|---|---|---|---|
| Viability | If started, can it be finished? | How many "ilities" are there? Dr. Charles Keating, Department of Engineering Management & Systems Engineering, Old Dominion University | | | | 2007 |
| Vulnerability | The characteristics of a system that cause it to suffer a definite degradation (loss or reduction of capability to perform the designated mission) as a result of having been subjected to a certain level of effects in an unnatural (man-made) hostile environment. Vulnerability is considered a subset of survivability. (Joint Pub 1-02). | AFOTEC OT&E Guide, 5th Edition | Attach-ment B | Glossary | B-36 | Jun-07 |
| Vulnerability | The characteristic of a system that causes it to suffer a definite degradation (loss or reduction of capability to perform its designated mission) as a result of having been subjected to a certain (defined) level of effects in an unnatural (manmade) hostile environment. Vulnerability is considered a subset of survivability. | Army Regulation 73-1 Test and Evaluation Test and Evaluation Policy | Glossary | Section II | 61 | Apr-06 |
| Vulnerability | The characteristics of a system that causes it to suffer a degradation (loss or reduction of capability to perform the designated mission) as a result of having been subjected to a certain (defined) level of effects in an unnatural (man made) hostile environment. Vulnerability is considered a subset of survivability. | http://www.cotf.navy.mil/files/otdfiles/glossary.pdf | Glossary | | 16 | Mar-09 |
| Vulnerability-m | Measure example: Percent of OPSEC Measures selected tied to Vulnerability Analysis. | CJCSM 3500.04C | Annex C Appendix C Enclosure B | OP 6.2.14 | B-C-C-145 | Jul-02 |
| Workflow | Leanness of process sequencing, product input/output mappings | JTEM Effectiveness and Suitability Rock Drill 3 Survey | M&P Effective-ness Criteria | | Slide 3 | Sep-07 |

## ANNEX N - REFERENCES


Army Evaluation Center (AEC), *Evaluator and Analyst Handbook*, Version 2.0, Alexandria, VA, Methodology and Analysis Division, Technical Support Directorate, Army Evaluation Center, 2008

Banks, J., J. S. Carson, B. L. Nelson, D. M. Nicol., *Discrete-Event Simulation*, 4th ed. Prentice-Hall, Upper Saddle River, NJ., 2005

Barbour, G. L., "Failure Modes and Effects Analysis by Matrix Methods," *Proceedings of the 1977 Annual Reliability and Maintainability Symposium*, 1977

Bayliss, J., *Strategy in the Contemporary World: An Introduction to Strategic Studies*, Oxford Oxfordshire, Oxford University press, 2002

Ben-Gal, I., and J. Bukchin, "The Ergonomic Design of Workstations Using Virtual Manufacturing and Response Surface Methodology," IIE Transactions 34: 375-391, 2002

Bettonvil, B., J. P. C. Kleijnen., "Searching for Important Factors in Simulation Models with Many Factors: Sequential Bifurcation," *European Journal of Operational Research* 96 180-194, 1997

Bonder, S., Army Operations Research, "Historical Perspectives and Lessons Learned," *Operations Research* 50 (1), 2002

Box, G. E. P., and K. B. Wilson, "On the Experimental Attainment of Optimum Conditions," J. Roy. Statist. Soc. Ser. B Metho.13: 1-45, Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester, England, 1951

Box, G. E. P., and N. R. Draper, *Empirical Model-Building and Response Surfaces*, New York: Wiley, 1987

Box, G. E. P., W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, John Wiley & Sons, Inc., New York, 1978

Box, G. E. P., W. G. Hunter, and J. S. Hunter, *Statistics For Experimenters: An Introduction to Design, Data Analysis and Model Building*, 2nd ed. New York: Wiley, 2005.

Bratley, P., B.L. Fox, and L.E. Schrage., *A Guide to Simulation*, 2nd. ed. Springer-Verlag, New York, 1987

Calvano, C. N., and John, P., "Systems Engineering in an Age of Complexity," *Systems Engineering* 7 (1), 2004

Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey, *Graphical Methods for Data Analysis*, Wadsworth, 1983

Chang, K.-H., and H. Wan., "Stochastic Trust Region Gradient-Free Method (STRONG): A New Response-Surface-Based Algorithm in Simulation Optimization," *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 346-354. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2007

Cioppa, T. M., *"Efficient Nearly Orthogonal and Spacefilling Experimental Designs for High-Dimensional Complex Models,"* Doctoral dissertation, Operations Research Department, Naval Postgraduate School, Monterey, California, 2002

Cioppa, T. M., and T. W. Lucas, "Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes," *Technometrics* 49(1): 45-55, 2007.

CJCSI 3170.01G, *Joint Capabilities Integration and Development System*, 1 March, 2009

CJCSM 3170.01C, *Operation of the Joint Capabilities Integration and Development System*, 1 May 2007

CJCSM 3500.04E U*niversal Joint Task List (UJTL) Manual*, 25 August 2008.

CONDOR 7.1.1. Available via http://www.cs.wisc.edu/condor/ [Last Accessed June 17, 2008].

Conover, W. J. & Iman, R. L., *"Rank Transformations As a Bridge between Parametric and Nonparametric Statistics," American Statistician*, *35*, 124-129. [1] http://is.ba.ttu.edu/conover/Dr.Conover.htm [2] http://citeseer.ist.psu.edu/context/1743341/0, 1981

Davis, Paul K., and Bigelow, James H., *Motivated Metamodels, Synthesis of Cause-Effect Reasoning and Statistical Metamodeling*, RAND, 2003

DoD Instruction (DoDI) 5000.2, E.6.1.d, December 8, 2008

Dorfman, R., "The Detection of Defective Numbers of Large Populations," *Annals of Mathematical Statistics* 14: 436-440., 1943

Draper, N. R., and H. Smith., *Applied Regression Analysis*, New York: Wiley, 1980.

Dussault, H. B., "Automated FMEA Status and Future," Reliability and Maintainability Symposium, 1984, Proceedings, Annual, 1984

Ferber, J., *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*, Boston, MA, Addison-Wesley, 1999

Ferguson, George A., Takane, Yoshio, *Statistical Analysis in Psychology and Education*, Sixth Edition. Montréal, Quebec: McGraw-Hill, Ryerson Limited, 2005

Fiebrandt, M. J., and J. Wilson, "System of Systems Measurement Framework in Evaluation of Joint Mission Effectiveness," AIAA, US Air Force Test and Evaluation (T&E) Days, 2008

Friedman, J. H., "Multivariate Adaptive Regression Splines (with discussion)," *Annals of Statistics* 19(1): 1-67, 1991.

Fu, M., "Optimization for Simulation: Theory vs. Practice," *INFORMS Journal on Computing* 14 (3): 192-215, 2002

Gunn, K. M., F. Dansashi, and D. J. Edwards, "A Service Oriented Architecture (SAO) Approach to Department of Defense Architecture Framework (DoDAF) Architecting," *Proceedings of the 13th International Command and Control Research Technology Symposium*, Bellevue, WA, 2008

Hand, D., H. Mannila, and P. Smyth., *Principles of Data Mining*, Cambridge, Massachusetts: MIT Press, 2001

Helsel, D. R., and R. M. Hirsch, Statistical Methods in Water Resources: Techniques of Water-Resourses Investigations," Book 4, chapter A3, U.S. Geological Survey, 522 pages.[3] http://pubs.water.usgs.gov/twri4a3, 2002

Hernandez, A. S., M. Carlyle, and T. W. Lucas, "Using Integer Programming to Generate Flexible Nearly Orthogonal Latin Hypercubes," Working paper, Operations Research Department, Naval Postgraduate School, Monterey, California, 2008.

Herrin, S., Iverson, D., and L. Spukovska, *Matrix Failure Modes and Effects Analysis as a Knowledge Base for a Real Time Automated Diagnosis Expert System*, Report # 20010043669, Moffett Field, CA, Ames Research Center. Available via http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010043669_2001064803.pdf [accessed 18 July 2008]

Holland, J. H., *Hidden Order: How Adaptation Builds Complexity*, Reading, MA, Addison-Wesley, 1995

Horne, G.E., *Maneuver Warfare Distillations: Essence Not Verisimilitude*, 1999 Winter Simulation Conference, 1999

Horne, Gary E., and Ted E. Meyer, "Data Farming: Discovering Surprise," *Proceedings of the 2004 Winter Simulation Conference*, 2004

Hughes, John A., Jon O'Brien, Tom Rodden, Mark Roundefield, and Steve Blythin, *"Designing with Ethnography: A Presentation Framework for Design, Designing Interactive Systems," Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods and techniques*, Amsterdam, The Netherlands, 1997

IDA, Capabilities Based T&E, http://www.ida.org/researchareas/resourceandsupportanalyses/training%20readiness%20and%20personnel%20issues.php [Last accessed 28 Jan 2009]

Joint Staff J7 Joint Experimentation, Transformation, and Concepts Division (JETCD) website, http://www.dtic.mil/futurejointwarfare/index.html [Last accessed 28 Jan 2009]

Kelton, W. D., R. P. Sadowski, and D. P. Sturrock, *Simulation with Arena*, 4th ed. New York: McGraw-Hill, 2007.

Khuri, A. I., "Multiresponse Surface Methodology," S. Ghosh, C. R. Rao, eds. *Handbook of Statistics*, Volume 13. Elsevier, Amsterdam, The Netherlands, 1996

Kittler, J. E., W. Menard, and K. A. Phillips, "Weight Concerns in Individuals with Body Dysmorphic Disorder," NIH Public Access, Author Manuscript, Eat Behav. 2007 January; 8(1): 115–120. , 2007

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa., "A User's Guide to the Brave New World of Simulation Experiments," *INFORMS Journal on Computing* 17 (3): 263-289 (with online companion), 2005

Law, A. M., *Simulation Modeling and Analysis*, 4th ed. New York: McGraw-Hill 2007

Lee, K., S. Cane, A. Salwa, and C. E. Martinez, "An Architecture Approach for Command, Coordination, and Communications Capabilities Assessment," *Proceedings of the 12th International Command and Control research and Technology Symposium*, Newport, RI, 2007

Leemis, L.M. and S.K. Park, *Discrete-Event Simulation: A First Course*. Pearson Prentice Hall, Upper Saddle River, NJ, 2006

Lindsey, Scott, *How to Write an Effect*, Naval Proceedings, April, 2006

Lucas, T. W., S. M. Sanchez, F. Martinez, L. R. Sickinger, and J. W. Roginski, "Defense and Homeland Security Applications of Multi-Agent Simulations," *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M. H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 138-149. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2007

Lucas, T. W., S. M. Sanchez, L. P. Brown, and W. Vinyard, "Better Designs for High-Dimensional Explorations of Distillations," *Maneuver Warfare Science 2002*, Marine Corps Combat Development Command, 2002

Macal, Charles M. and Michael J. North, "Tutorial on Agent-Based Modeling and Simulation Part 2: How to Model with Agents," *Proceedings of the 2006 Winter Simulation Conference*, IEEE 1-4244-0501-7/06., 2006

Michelson-Morley Experiment, 1887

Miller, A. J., *Subset Selection in Regression*, Chapman and Hall, Boca Raton, FL. CRC Press, 1990

Miller, G. A., Notes for SE4008 (Systems Engineering and Integration), Naval Postgraduate School, unpublished, 2008

Montgomery, D. C., *Design and Analysis of Experiments*, 5th ed. New York: Wiley, 2000

Myers, R. H., and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, 2nd ed. John Wiley & Sons, New York, 2002

NIST/Sematech, e-Handbook of statistical methods, Available via <http://www.itl.nist.gov/div898/handbook/> [accessed July 1, 2006], 2005

Osborne, D. M., R. L. Armacost, and J. Pet-Edwards, "State of the Art in Multiple Response Surface Methodology," 1997 IEEE International Conference on Systems, Man, and Cybernetics, 3833-3838, 1997

Park, K. S., and K. J. Kim, *Optimizing Multi-Response Surface Problems: How to Use Multi-Objective Optimization Techniques*, 2005

Rees, L. P., E. R. Clayton, and B. W. Taylor, "Solving Multiple Response Simulation Models Using Modified Response Surface Methodology Within a Lexicographic Goal Programming Framework," *IIE Transactions* 17(1): 47-57, 1985

Ring, S.J., D. Nicholson, J. Thilenius, and S. Harris, "An Activity-Based Methodology for Development and Analysis of Integrated DoD Architectures," 2004 Command and Control Research and Technology Symposium, The Power of Information Age Concepts and Technologies, C2 Tools and Metrics Track, Mitre, 2004

Ryan, T. P., *Modern Experimental Design*, Hoboken, New Jersey: Wiley, 2007

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and Analysis of Computer Experiments (includes Comments and Rejoinder)," *Statistical Science* 4: 409-435, 1989

"Sample Data Analysis Plan for Joint Test and Evaluation Events," Joint Test and Evaluation Methodology, Joint Test and Evaluation, March 2009

"Sample Data Management Plan for Joint Test and Evaluation Events," Joint Test and Evaluation Methodology, Joint Test and Evaluation, March 2009

Sanchez, S. M., "Robust Design: Seeking the Best of All Possible Worlds," *Proceedings of the 2000 Winter Simulation Conference*, eds. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 69-76. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2000

Sanchez, S. M., and T. W. Lucas, “Exploring the World of Agent-Based Simulations: Simple Models, Complex Analyses,” *Proceedings of the 2002 Winter Simulation Conference*, eds. E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. Charnes, 116-126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2002

Sanchez, S. M., and P. J. Sanchez, “Very Large Fractional Factorials and Central Composite Designs,” *ACM Transactions on Modeling and Computer Simulation*, 15(4): 362-377, 2005

Sanchez, S. M., H. Wan, and T. W. Lucas, “Two-Phase Screening Procedures for Simulation Experiments,” *ACM Transactions on Modeling and Computer Simulation*, forth-coming, 2008

Sanchez, “Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments,” *Proceedings of the 2006 Winter Simulation Conference*, 2006

Santner, T. J., B. J. Williams, W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York.

SAS Institute Inc., *JMP Statistics and Graphics Guide*, Release 7. Cary, NC, the SAS Institute, Incorporated, 2007

Schamburg, J., and D. Brown, “A Generalized Multiple Response Surface Methodology for Complex Computer Simulation Applications,” *Proceedings of the 2004 Winter Simulation Conference*, eds. R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 958-966, Available via <http://www.informs-sim.org/wsc04papers> [accessed July 15, 2008], 2004

Schamburg, J., “A Modified Response Surface Methodology for Knowledge Discovery,” Ph.D. dissertation, University of Virginia, Charlottesville, Virginia, 2004

Seaman, J. W., S. C. Walls, S. E. Wide, and R. G. Jaeger, “Caveat Emptor: Rank Transform Methods and Interactions,” *Trends Ecol. Evol.*, 9, 261-263, 1994

SEED Center for Data Farming, “Simulation experiments & efficient designs” [online], Available via <http://harvest.nps.edu> [accessed May 15, 2008]

Shadish, W., T. Cook, and D. Campbell, *Experimental and Quasi-Experimental Design for Generalized Causal Inference*, New York: Houghton Mifflin Co., 2001

Sheehan, Jack, “Missions and Means Framework, Measures and Evaluation Techniques,” briefing, 2006

Taguchi, G., *System of Experimental Design*, vols. 1 and 2, White Plains, New York: UNIPUB/Krauss International, 1987

Tan, P. N., M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Boston, Massachusetts: Addison-Wesley, 2005

Taylor, J. G., *Modeling and Simulation of Land Combat*, L. G. Callahan, ed., Atlanta, GA, Georgia Institute of Technology, 1983

Testing in a Joint Environment Roadmap Steering Group (TSSG), 26 June 2008.

Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977

U.S. Joint Forces Command (USJFCOM), *Commanders Handbook for an Effects-Based Approach to Joint Operations*, 2006

US Air Force Test and Evaluation Center (AFOTEC), *AFOTEC OT&E Guide*, 5th ed., Kirtland AFB, NM, AFOTEC, 2007

US Operational Test and Evaluation Force (OPTEVFOR), *Analyst's Handbook*, Norfolk, VA, OPTEVFOR, 2007

www.en.wikipedia.org/wiki/Anova, Wikipedia definition for Analysis of variance, 13 Jan 2009

www.en.wikipedia.org/wiki/Box_plots, Wikipedia definition of Box plot, 13 Jan 2009

www.en.wikipedia.org/wiki/Cross-correlation, Wikipedia definition of Cross-correlation, 13 Jan 2009

www.en.wikipedia.org/wiki/Regression_analysis, Wikipedia definition of Regression analysis, 13 Jan 2009

Zeigler B. P., K. Praehofer, T. G. Kim, *Theory of Modeling and Simulation*, 2nd ed. Academic Press, San Diego, CA, 2000

*This page intentionally left blank.*

# ANNEX O – CHECKLISTS

Chapter 2 Analysis Handbook Checklists

| Recommended Input Info/Sources | Products |
|---|---|
| ***1.1 Develop Capability/SoS Description*** | |
| ▪ Joint Capability Areas (JCA)<br>▪ Initial Capabilities Document (ICD)<br>▪ Capability Development Document (CDD)<br>▪ Capability Production Document (CPD)<br>▪ Doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) Change Request (DCR)<br>▪ Other Test Plans (Developmental Test [DT], Operational Test [OT], Live Fire [LF] Test)<br>▪ Approved Analysis of Alternatives (AoA) Plan<br>▪ Analytical Baseline (DPS, MSFD, FYAB) | ❑ a. Capability/SoS Description<br>❑ b. Capability/SoS OV-1 |

**Figure 2-5. CTM 1.1 Develop Capability/SoS Description**

| Recommended Input Info/Sources | Products |
|---|---|
| ***1.3 Develop JOC-T*** | |
| ▪ Capability/SoS Description<br>▪ Joint Capability Areas<br>▪ Universal Joint Task List (UJTL)<br>▪ Joint Operations Concept (JOpsC) Family (Joint Operating Concept [JOC], Joint Integrating Concept [JIC], Joint Functional Concept [JFC])<br>▪ Initial Capabilities Document (ICD)<br>▪ Capability Development Document (CDD)<br>▪ Capability Production Document (CPD) | ❑ a. Joint Operational Context for Test (JOC-T) |
| ***1.3.1 Analyze Mission*** | |
| ***1.3.2 Analyze Blue*** | |
| ***1.3.3 Analyze Environment*** | |
| ***1.3.4 Analyze Threat*** | |
| ***1.3.5 Compose JOC-T*** | |
| ***1.3.6 Validate JOC-T*** | |

**Figure 2-8. CTM 1.3 Develop JOC-T**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***1.2 Develop Evaluation Strategy*** | | |
| | ☐ | Evaluation Strategy |
| | | |
| ***1.2.1 Identify and Collect Evaluation Inputs*** | | |
| ▪ Initial Capabilities Document/Capability Development Document<br>▪ Analysis of Alternatives<br>▪ Joint Capabilities Evaluation<br>▪ Joint Operational Context for Test (JOC-T)<br>▪ Data Analysis Plan (DAP) template | ☐ | a. Functional Area Analysis Inputs |
| | ☐ | b. Functional Needs Analysis Inputs |
| | ☐ | c. Functional Solutions Analysis Inputs |
| | ☐ | d. Capability Description Inputs |
| | ☐ | e. Relevant Capability Analysis Designs and Results |
| | ☐ | f. Relevant Test & Evaluation Results |
| | | |
| ***1.2.2 Develop CCIs and COIs*** | | |
| ▪ Functional Area Analysis Inputs<br>▪ Functional Needs Analysis Inputs<br>▪ Functional Solutions Analysis Inputs<br>▪ Relevant Capability Analysis Designs and Results<br>▪ Relevant T&E Results<br>▪ Joint Operational Context for Test (JOC-T)<br>▪ Joint Capability Area (JCA) - Universal Joint Task List (UJTL)/Service task mapping | ☐ | a. Critical Capability Issues (CCI) |
| | ☐ | b. Critical Operational Issue (COI) |
| | | |
| ***1.2.3 Establish T&E Strategy Framework*** | | |
| ▪ Joint Operational Context for Test (JOC-T)<br>▪ Critical Capability Issues<br>▪ Critical Operational Issues<br>▪ Capability Crosswalk | ☐ | Evaluation Strategy Elements to include the following: |
| | ☐ | a. Independent factors: Threat conditions |
| | ☐ | b. Independent factors: Environmental conditions |
| | ☐ | c. Independent factors: System/SoS configurations across DOTMLPF |
| | ☐ | d. Dependent measures: Mission measures of effectiveness (MMOEs) |
| | ☐ | e. Dependent measures: Task measures of performance (TMOPs) |
| | ☐ | f. Dependent measures: System/SoS attributes |
| | | |
| ***1.2.4 Develop Risks and Mitigations*** | | |
| ▪ Evaluation Strategy | ☐ | a. Identified Risks and Mitigations |

| Recommended Input Info/Sources | Products | |
|---|---|---|
| | | |
| ***1.2.5 Develop Infrastructure Strategy*** | | |
| ▪ JOC-T | ❑ | a. Infrastructure Strategy |
| ▪ Evaluation Strategy | ❑ | |

**Figure 2-14. CTM 1.2 Develop Evaluation Strategy**

| Recommended Input Info/Sources | Products |
|---|---|
| ***1.4 Develop/Refine Capability Crosswalk*** | |
| ▪ Evaluation Strategy (across life cycle including Development Test [DT] and Operational Test [OT]) | ❑ a. Capability Crosswalk Matrix |

**Figure 2-18. CTM 1.4 Develop/Refine Capability Crosswalk**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***2.1 Develop Test Concept*** | | |
| | ❑ | Test Concept |
| | | |
| ***2.1.1 Establish Overall Test Goal*** | | |
| ▪ Test and Evaluation Strategy<br>▪ Test and Evaluation Master Plan<br>▪ Other Test Plans<br>▪ Initial Capabilities Documents<br>▪ Capability Development Documents<br>▪ Capability Production Documents<br>▪ Joint Capability Areas<br>▪ DOTMLPF Change Requests | ❑ | a. Test Goal |
| | | |
| ***2.1.2 Establish Test Objectives*** | | |
| ▪ Test Goal<br>▪ Test and Evaluation Strategy | ❑ | a. Test Objectives |
| | | |
| ***2.1.3 Develop Test Approach*** | | |
| ▪ Test Objectives | ❑ | a. Test Approach |

**Figure 3-9. CTM 2.1 Develop Test Concept**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***2.2 Refine Evaluation Strategy*** | | |
| | ❑ | Refined Evaluation Strategy |
| | | |
| ***2.2.1 Develop Initial Data Analysis Requirements*** | | |
| ▪ Test and Evaluation Strategy<br>▪ Test and Evaluation Master Plan<br>▪ Other Test Plans<br>▪ Initial Capabilities Documents, Capability Development Documents, and Capability Production Documents<br>▪ Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel and Facilities (DOTMLPF) Change Request<br>▪ System of System/System Description<br>▪ Critical Capability Issues (CCI)/Critical Operational Issues (COI)<br>▪ Joint Operational Context for Test (JOC-T) | ❑ | a. Capability evaluation subset |
| | ❑ | (1) Test CCI/COI |
| | ❑ | (2) Test factors |
| | ❑ | (3) Test measures |
| | ❑ | (4) Test data elements |
| | | |
| ***2.2.2 Identify Additional Modeling Requirements*** | | |
| ▪ Known Modeling Requirements | ❑ | a. Additional Modeling Requirements |
| | | |
| ***2.2.3 Develop Test Scenario*** | | |
| ▪ Capability Evaluation Subset<br>o Factors<br>o Measures<br>o Data Elements<br>▪ Critical Capability Issues (CCI)/Critical Operational Issues (COI)<br>▪ Joint Operational Context for Test (JOC-T) | ❑ | a. Test Scenario |

**Figure 3-10. CTM 2.2 Refine Evaluation Strategy**

| Recommended Input Info/Sources | Products |
|---|---|
| ***3.1 Develop Test Design*** | |
| | ☐ Test Design |
| | |
| ***3.1.1 Develop Analysis Approach*** | |
| ▪ Analytical model capabilities<br>▪ Test measures and data elements<br>▪ Efficient test trial design of experiment | ❑ a. Analytical approach |
| | |
| ***3.1.2 Design Test Trials*** | |
| ▪ Vignettes | ❑ a. Test trial matrix |
| | |
| ***3.1.3 Develop Data Collection Requirements*** | |
| ▪ Test Trials | ❑ a. Integrated Data Requirements List |
| | |
| ***3.1.4 Identify Additional Plan Test Modeling Requirements*** | |
| ▪ Vignettes<br>▪ Test trials<br>▪ Integrated Data Requirements List | ❑ a. Analytic Model Capabilities |
| | |
| ***3.1.5 Develop Vignettes*** | |
| ▪ Test scenarios<br>▪ Test trials | ❑ a. Vignettes |
| | |
| ***3.1.6 Verify and Validate Scenario and Vignettes*** | |
| ▪ Test scenario | ❑ a. Verified and validated test scenario |
| ▪ Vignettes | ❑ b. Verified and validated test vignettes |

**Figure 3-11. CTM 3.1 Develop Test Design**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***4.1 Design LVC-DE Configuration*** | | |
| | ❑ | LVC-DE Configuration |
| | | |
| ***4.1.1 Perform JME LVC Logical Design*** | | |
| ▪ Joint Operational Context for Test<br>▪ DoDAF Products (AV-2, TV-1, OV-1, 2, 4, 5, 6c, 7, SV-1, 4a, 4b, 6, 10c | ❑ | JME LVC-DE Logical Design |
| | | |
| | | |
| ***4.1.3 Perform JME LVC Physical Design*** | | |
| ▪ JME LVC-DE Logical Design<br>▪ DoDAF Products (AV-2, TV-1, OV-1, 2, 4, 5, 6c, 7, SV-1, 4a, 4b, 6, 10c<br>▪ Data Analysis Plan<br>▪ Vignettes<br>▪ V&V Plan (Initial) | ❑ | JME LVC-DE Physical Design |
| | ❑ | IDRL – Physical Design mapping |
| | | |

**Figure 4-3. CTM 4.1 Design LVC-DE Configuration**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***5.1 Develop Event Management Plan*** | | |
| | ❑ | Event Management Plan |
| | | |
| ***5.1.2 Develop Data Management Plan*** | | |
| ▪ Test Plan<br>▪ Data Analysis Plan<br>▪ Test Trial Matrix<br>▪ Vignettes<br>▪ DoDAF Products (AV-2, TV-1, OV-1, 2, 4, 5, 6c, 7, SV-1, 4a, 4b, 6, 10c, 11 | ❑ | Data Management Plan |

**Figure 5-4. CTM 5.1 Develop Event Management Plan**

| Recommended Input Info/Sources | Products | |
|---|---|---|
| ***5.2 Run Event*** | | |
| | | |
| ***5.2.4 Capture and Archive Data*** | | |
| ▪ Data Management Plan<br>▪ Data Analysis Plan<br>▪ Test Plan | ❑ | Archived Data |
| ***5.2.6 Assess Event*** | | |
| ▪ Data Management Plan<br>▪ Data Analysis Plan<br>▪ Test Plan<br>▪ Test Data | ❑ | Quicklook Evaluation |

**Figure 5-5. CTM 5.2 Run Event**

| Recommended Input Info/Sources | | Products |
|---|---|---|
| ***6.1 Process Test Data*** | | |
| ▪ Data Management Plan<br>▪ Data Analysis Plan<br>▪ Test Plan | ❑ | a. Processed Test Data |
| | ❑ | b. Reconstructed Mission Level Data |
| | ❑ | c. Reconstructed Task Level Data |
| | ❑ | d. Reconstructed System Level Data |

**Figure 6-5. CTM 6.1 Process Test Data**

| Recommended Input Info/Sources | | Products |
|---|---|---|
| ***6.2 Analyze Data*** | | |
| | ❑ | Capabilities Analysis Information |
| | | |
| ***6.2.1 Prepare Analysis Tools*** | | |
| ▪ Data Analysis Plan | ❑ | a. Analysis Tools |
| | ❑ | b. Analysis Techniques |
| | | |
| ***6.2.2 Conduct Analysis*** | | |
| ▪ Processed Test Data | ❑ | a. Joint Mission Effectiveness Evaluation |
| | ❑ | b. Task Performance Evaluation |
| | ❑ | c. SUT/SoS Performance Evaluation |

**Figure 6-6. CTM 6.2 Analyze Data**

| Recommended Input Info/Sources | | Products |
|---|---|---|
| ***6.3 Evaluate SoS and JMe*** | | |
| ▪ Data Analysis Plan | ❑ | Joint Capabilities Evaluation (JCE) |
| | | |
| ***6.3.1 Integrate Pre-Event Analysis Results*** | | |
| ▪ Pre-Event Analysis Results<br>▪ Processed Test Data | ❑ | a. Integrated Analysis Results |
| | | |
| ***6.3.2 Synthesize SoS – Tasks – JMe Evaluations*** | | |
| ▪ Integrated Analysis Results | ❑ | a. Synthesized JMe Evaluation |
| | ❑ | b. Synthesized Task Performance Evaluation |
| | ❑ | c. Synthesized SUT/SoS Performance Evaluation |
| | | |
| ***6.3.3 Identify Significant Findings*** | | |
| ▪ Synthesized JMe Evaluation<br>▪ Synthesized Task Performance Evaluation<br>▪ Synthesized SUT/SoS Performance Evaluation | ❑ | a. CCI Significant Findings |
| | ❑ | b. COI Significant Findings |
| | | |
| ***6.3.4 Make Recommendations*** | | |
| ▪ CCI Significant Findings<br>▪ COI Significant Findings | ❑ | a. CCI Recommendations |
| | ❑ | b. COI Recommendations |

**Figure 6-8. CTM 6.3 Evaluate SoS and JMe**