CHOOSE AN ITEM.

NAME OF THE SYSTEM UNDER TEST

INTEGRATED EVALUATION FRAMEWORK (IEF)

COMOPTEVFOR 3980 (XXXX-OT-XX)
Ser XXX/XXX

The cover page will identify the system (increment) and include a picture of the system, date, serial number of the document, and distribution statements. It will also contain the COMOPTEVFOR logo and appropriate FOUO/Classification markings.

DATE This is the signature date.

DISTRIBUTION STATEMENT B. Distribution authorized to U.S. Government agencies only; test and evaluation document. Other requests for this document shall be referred to CNO (N84) or COMOPTEVFOR via DTIC using DTIC form 55.

(If there is no FOUO information in the CDR's ltr and only your survey sheets have FOUO (when filled in), remove this statement.)

This document contains information exempt from mandatory disclosure under the FOIA. Exemption 5USC 552(b)(5) applies.

COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE
NORFOLK, VIRGINIA

CHOOSE AN ITEM.

BLANK PAGE

Start defining acronyms here for first use. Do not need to be redefined in the enclosure.

COMOPTEVFOR INTEGRATED EVALUATION FRAMEWORK (IEF)

The 1st paragraph will describe the high level purpose of the document including the system name and objectives. The second paragraph will describe the formal approval and updating of the IEF, and any major deviations from normal practices.

This is the IEF for the System Under Test (SYSTEM ACRONYM), Chief of Naval Operations (CNO) Project No. TEIN. The framework is intended to:

(Modify bullets below as appropriate.)

- Document mission and capabilities analyses and test design conducted during the Mission-Based Test Design (MBTD) process.
- Detail the missions, tasks, and subtasks to be supported by the system, the conditions under which these elements must be performed, the data required to support performance evaluation, test methods, and test events to be accomplished, and test resource requirements.
- Describe the overarching Operational Test (OT) strategy and document an up-front view of testing, coordinated between Commander, Operational Test and Evaluation Force (COMOPTEVFOR), and Director, Operational Test and Evaluation (DOT&E) (for oversight programs).
- Serve as the foundational document to support OT data gathering during Integrated Testing (IT).
- Identify the minimum data requirements (IT, OT, M&S) to evaluate the System Under Test (SUT) effectiveness and suitability across the operational environment.
- Provide the foundation for the OT input to the Test and Evaluation Master Plan (TEMP) including early identification of resources.
- Provide a basis for the integration of OT objectives with Developmental Test (DT), Contractor Test (CT), and Live Fire Test and Evaluation (LFT&E) objectives.
- Provide a basis for Operational Assessments (OA), Initial Operational Test and Evaluation (IOT&E), and Follow-on Operational Test and Evaluation (FOT&E).

Enclosure (1) is provided for planning purposes. Testing supported by this document will be accomplished per reference (a), the Operational Test Director's (OTD) Manual. Updates to

the IEF will be prepared and released anytime substantive program or requirements changes occur. At a minimum, updates will be issued upon the decision to update the TEMP. The IEF will be reviewed to determine if an update is required at the completion of the Critical Design Review (CDR) and following the release of an updated Operational Requirements Document (ORD), Capabilities Design Document (CDD), or Capabilities Production Document (CPD). (as appropriate)

J. R. PENFIELD

Distribution: PLANNED DISTRIBUTION (FOR EXAMPLE):
OSD (DOT&E)
OUSD (AT&L)(DASD DT&E)
ASN RDA (DASN(RDT&E/CHSENG))

ADDITIONAL DISTRIBUTION OPTIONS – NOTE that these should not be blindly copied and pasted, but need to be scrubbed for applicability to the SUT. Consult with 01A and the Division A codes for appropriate recipients:

CNO (appropriate N codes) (N84, N842, etc.)
COMNAV SYSCOM ( ) ***(Cognizant Program Manager, e.g., PMS or PMA)***
COMSPAWARSYSCOM (PEO C4I) (for C4I systems)
COMNAVNETWARCOM (NNWC C4)
COMNAVNETWARCOM (NNWC CS) (for Combat systems)
COMNAVAIRPAC (N3, N7, N8, N41, N42, N43, N45)
COMNAVAIRLANT (N3, N7, N8, N41, N42, N43, N45)

Supporting Test Squadron – as appropriate

AIRTEVRON ONE
AIRTEVRON NINE
AIRTEVRON TWO TWO

Close up blank lines in distribution list when done. For more information on formatting the distribution list, see the DON Correspondence Manual available on the Y: drive.

XXX SYSTEM

INTEGRATED EVALUATION FRAMEWORK (IEF)

DISTRIBUTION STATEMENT B. Distribution authorized to U.S. Government agencies only; test and evaluation document dated _____. Other requests for this document shall be referred to CNO (N84) or COMOPTEVFOR via DTIC using DTIC form 55.

This document contains information exempt from mandatory disclosure under the FOIA. Exemption 5USC 552(b)(5) applies

(If there is no FOUO information in the enclosure and only your survey sheets have FOUO (when filled in), remove this statement.)

Enclosure (1)

BLANK PAGE

To generate the TOC, Styles must be used in the section/appendix headings: Heading 1 (TOC 1), Heading 2 (TOC 2), Heading 7 (TOC 1), and Heading 8 (TOC 2).

CONTENTS

TABLES

CHOOSE AN ITEM.

Use style Heading 1 for all SECTION headings. Type the title (and classification, if required). No numbering, spacing, or tabbing required. Refer to the Quick Style Gallery.

SECTION 1 - INTRODUCTION

1.1 PURPOSE

Use style Heading 2 for all section level 2 headings.

Describe the overarching purpose of the document. If this is a revision to previous IEFs, describe the revisions made.

EXAMPLE

This IEF for System Under Test (SYSTEM ACRONYM), CNO Project No. TEIN documents the results of the MBTD process. MBTD was executed for this program per reference (a). The IEF supports revisions to the TEMP and planning for IOT&E and FOT&E (as required). The OA test plan (if applicable) will be developed using this framework, taking into account the results of IT and any changes in planned scope of test. Prior to IOT&E, a revision to the IEF is anticipated to support a revision to the TEMP.

1.2 SYSTEM DESCRIPTION

1.2.1 SUT

Use style Heading 3 for all section level 3 headings.

The SYSTEM ACRONYM SUT…

Provide a description of the expected final configuration of the SUT and the environment in which it is intended to operate. If the SUT replaces an existing system, be clear how the new system is meant to improve over legacy (task execution, reliability, etc.). If this IEF revision tests enhancements over a previously tested version/increment, place special emphasis on what modifications have been made or what upgrades have been incorporated, and how performance should improve. This section will be used to help the reader understand the scope of test. The reader must be able to understand where the SUT stops (SUT boundary) and the SoS begins. Explain those outputs from the SUT that support the SoS. The reader must also understand what the system does, to properly review task execution, capabilities, and the resulting test strategy.

Most frameworks are written for the entire system (current and future), but some are written for programs with limited remaining scope and testing. If that is the case, this is the paragraph to explain how your framework is limited to certain pieces of a system and/or capabilities.

1.2.2 System of Systems (SoS)

The SYSTEM ACRONYM SoS…

Provide a basic description of the SoS. This must encompass the accomplishment of all missions detailed by the MBTD. Determining SoS boundaries is not always intuitive. The SoS description should capture the systems required to execute the missions which the SUT is supporting. The reader must understand those SoS inputs to the SUT required for SUT mission accomplishment. If SOS enhancements (or the interactions between the SOS and SUT enhancements) are significant to test results (i.e. regression analysis), explain that here.

Blank pages are only added if the section ends on an odd numbered page. Each section will begin with an odd numbered page.

BLANK PAGE

SECTION 2 - TEST DESIGN

2.1 EFFECTIVENESS CRITICAL OPERATIONAL ISSUES (COI)

Be certain that your COIs are consistent with the standard Navy missions and the associated default mission threads (located in the IEF database). If you have any nonstandard COIs, explain why. It is possible that not all your SUT COIs will be analyzed as a part of this IEF (i.e., the IEF update supports an FOT&E where only some COIs apply). If so, state that in the paragraph below.
EXAMPLES

The following effectiveness COIs reflect the analysis of SUT capabilities and the missions it supports. The SUT is net enabled and IA is included as an additional COI to support analysis and reporting on the program's IA capabilities.

(Example of an additional COI)

2.1.1 E-1, Air Warfare (AW)

Will the SYSTEM support the AW mission?

2.1.2 E-2, Strike Warfare (STW)

Will the SYSTEM…?

2.1.3 E-3, Information Assurance (IA)

Will SYSTEM IA protect, detect, react, and restore capabilities support completion of its missions?

2.2 SUITABILITY COIS

Use the four standard COIs; Reliability, Maintainability, Availability, and Logistic Supportability. If you add other suitability COIs or remove any of the standard four, explain why.
EXAMPLES

The following suitability COIs reflect the standard COIs for COMOPTEVFOR, as set by reference (a).

2.2.1 S-1, Reliability

Will SYSTEM reliability support mission accomplishment?

2.2.2 S-2, Maintainability

Will the SYSTEM be maintainable by Fleet personnel?

2.2.3 S-3, Availability

Will SYSTEM availability support mission accomplishment?

2.2.4 S-4, Logistic Supportability

Will the SYSTEM be logistically supportable?

2.3 (U) STATISTICAL/EXPERIMENTAL DESIGN

The following is a "template" whose purpose is to guide test teams in their writing section 2.3 of the IEF. It is important to treat this as a GUIDE and not as a BOILERPLATE to be repeated verbatim. Details of operational tests vary from one SUT to another. Similarly, details of the statistical design vary. Therefore, the details in section 2.3 should conform to the specific system's mission-oriented assessment.

(U) The following sections describe the statistical test procedures for assessing mission effectiveness/suitability of SSC in all mission areas.

2.3.1 (U) E-1, AMW

2.3.1.1 (U) Critical Tasks and Measures

This paragraph introduces the table and points the reader towards appendices B and C for additional details. It also introduces the concept of response variables.

(U) The critical tasks comprising the AMW mission are enumerated in Table 2-1 along with associated measures and response variables (RVs) that will be used to assess AMW mission success. Table B-4 illustrates how the tasks, subtasks, and the measures of the AMW mission relate.

The table below is required for every COI. Identify the critical tasks for each COI and the critical measures associated with those tasks. Mark RVs and KPPs as appropriate.

For FOT&E updates to an IEF, it is important to distinguish between critical tasks/measures that directly apply to the enhancement or new capability of the SUT, and those that were retained from IOT&E to support regression testing. Do not simply reuse table 2-1 from IOT&E. In the FOT&E example, any measures retained as critical for the sole purpose of regression testing must be marked as note 1.

Table 2-1. (U) AMW Critical Tasks/Measures
UNCLASSIFIED//FOR OFFICIAL USE ONLY

Task Title	Critical Measure	
1.1.4 – Onload Vehicles and Cargo	M1[1]	Payload Capacity
	M18[1]	Manpower
	M49	Well Deck Craft Load Time
1.5.1 – Conduct Overwater Transit	M17	Seaworthiness
	M54 (RV)	Sortie Time
1.5.3 – Conduct Land	M19	Inland Accessibility – Operations above the HWM

Table 2-1. (U) AMW Critical Tasks/Measures UNCLASSIFIED//FOR OFFICIAL USE ONLY		
Task Title	**Critical Measure**	
transit over the High Water Mark (HWM)	M20 (RV)	Probability of Successful HWM Transition (P_{HWMTS})
1.5.5 – Off-load Vehicles and Cargo	M1[1]	Payload Capacity
	M18[1]	Manpower
Note 1 - These critical measures specifically focus on regression testing.		

2.3.1.2 (U) Response Variable – Sortie Time (ST), M54

Key elements of the "Response Variable" paragraph are (1) test objective for this RV; (2) description and definition of the RV including its unit of measure (seconds, minutes, meters, miles, etc.) and how it is to be calculated (with equations and input variables if necessary); (3) distributional characteristics of the RV; (4) threshold. Note: "Sortie Time (ST)" used in this this example is a continuous RV, as compared to, for example, a binomial variable.

RVs are critical measures on which statistical analyses are carried out to characterize how the factors affect them.

The most common objective is "characterization," which refers to analyzing and graphing the values of the RV in different conditions defined by the combinations of controlled factors in the design.

(U//FOUO) The objective is to characterize ST across the operational conditions described by the factors (section 2.3.1.2.1.1 below), including the main effects and interactions of these factors. ST is defined as the required time for an SSC to complete a sortie. It is measured from the point when the SSC moves past the Line of Departure (LOD) until SSC transits a 25-nm route to the oceanfront HWM previously prepared craft landing zone/landing site, offloads cargo, completes the 25-nm return transit to the LOD, and finally requests a green well. Measurement of ST excludes external time delays created by environment including unacceptable surf conditions, navigational restrictions, white traffic, unplanned beachmaster actions, and mechanical failures of the payloads and gripes. ST is a continuous variable assumed to be normally and independently distributed. The specified threshold is a mean ST of 120 minutes. The expected range of ST values is 90 to 130 minutes. Because no historical test data exist, a standard deviation (sigma) of 10 minutes was roughly estimated by dividing the expected range of ST by 4: 40 minutes ÷ 4 = 10 minutes. (This rough rule of thumb is based on Tchebysheff's theorem.)

2.3.1.2.1 (U) Conditions

2.3.1.2.1.1 (U) Controlled Conditions (Factors)

In documenting the factors, do the following: (1) use the name and number of the condition from Table B-1; (2) describe the levels; (3) provide an explanation of why the factor is included. Focus on how the factor is expected to affect the quantitative results of the RV. There may be times when apparently important variables are *not* included as factors in the experimental design. In these cases, a one- or two-sentence explanation of why they are not included should be presented. Such statements will most often appear with recordable conditions that were no set as controlled conditions.

(U) Two factors are expected to substantially affect ST:

- (U) **Cargo Type (C 4.2) – 4 levels (Load 1, Load 2, Load 3, Load 4**) described below:
  - (U) **Load 1 -** Extreme Single Load Configuration 1 – One Medium Tactical Vehicle Replacement (MTVR) w/Water Trailer, one M9 Armored Combat Earthmover (ACE), one Armored Truck, Utility Vehicle w/Cargo Trailer, and one Tractor, Rubber-tired, Articulate Steering, Multipurpose (TRAM).
  - (U) **Load 2 -** Extreme Single Load Configuration 2 – One M1A1 Tank w/Track Width Mine Plow (TWMP).
  - (U) **Load 3 -** Extreme Single Load Configuration 3 – Two MTVRs w/two M777 155-mm Lightweight (L/W) Howitzers.
  - (U) **Load 4 -** Extreme Single Load Configuration 4 – Seven Armored Truck, Utility Vehicles.

  Different loads of cargo may impact sortie time due to different weights, types of gripes, and maneuvering constraints within the confines of the craft. **Load 2** is anticipated to affect ST most significantly.

- (U) **Light (C 1.3.2.1) – 2 levels (Day, Night)**
  - (U) **Day**
  - (U) **Night**

  It is anticipated that sorties conducted during night may increase ST because of reduced visibility and increased difficulty of maneuvering without light.

2.3.1.2.1.2 (U) Constant Conditions

For constant conditions: (1) use the condition name/number from Table B-1; (2) describe the level to be used for test; (3) explain why it is held constant at the chosen level.

(U) Two conditions will be held constant at specific levels because they are specifically called for in the requirements documents:

- (U) **Transit Distance (C 4.12) – Medium (10 – 25NM)**
  This is the most likely distance given that most amphibious assaults will occur with the ships over the horizon from the landing zone/landing site, approximately 20nm.
- (U) **Staging Area (C 2.5.4.2.3) – Ashore**
  The staging area will be a previously prepared craft landing zone/landing site on a readily available beach. This zone is most operationally representative.

2.3.1.2.1.3 (U) Recordable Conditions

For recordable conditions: (1) use the condition name/number from Table B-1; (2) explain why it is important enough to record, but not important enough to control.

(U) The conditions listed below are likely to impact ST, but are treated as recordable because of the intrinsic difficulty to control them during test, and/or due to SME determination that the impact should be minimal:

- (U) **Terrain Slope (C 1.1.1.3)**
  If the average steepness or grade of the landing area is steep, ST may be affected.
- (U) **Obstacles to Movement (C 1.1.3.4)**
  The presence of obstacles to movement may cause ST to be longer.
- (U) **Ocean Currents (C 1.2.1.2)**
  The strength of the current may impact the maneuvering of the craft possibly affecting the ST.
- (U) **Sea State (C 1.2.1.3)**
  The roughness of the seas may impact the maneuvering of the craft possibly affecting the ST.
- (U) **Significant Wave Height (C 4.9)**
  As wave height increases, maneuvering the craft becomes more difficult and is likely to affect ST.
- (U) **Waterspace (C 4.10)**
  The availability of space to maneuver may impact how the craft must maneuver and where it must transit affecting ST.

2.3.1.2.2 (U) Test Design

The Test Design section describes (1) the layout of the statistical design (e.g., full factorial, fractional factorial, optimal design, split plot factorial, single factor design, single sample design, and so on) and why the design was chosen; (2) The statistical analysis (e.g., ANOVA, logistic regression analysis, etc.); (3) Other special information required to complete the description of the design (for example, the fact that certain combinations of factors are disallowed). Because test design and the next section on sample size and power are intimately connected, some reference may be made to that section.

(U) A four-by-two full factorial design with one disallowed combination was chosen. The Load 1 Cargo Type will not be run at Night, as that is prohibited by procedure. Load 1 runs during the day are doubled to maintain a balanced design across the Cargo Type factor. Three replications of the design result in a total of 24 runs. The rationale for the number of replications is presented in Section 2.3.1.2.3 below. Analysis of Variance (ANOVA), confidence intervals, and graphical displays will be used to analyze the data, characterize ST across the operational envelope, and determine whether threshold is met. Table C-6 describes all runs including excursion runs.

2.3.1.2.3 (U) Sample Size and Statistical Power Analysis

The “Sample Size and Statistical Power Analysis” section presents the following for reporting power at the factor level: (1) Type I error rate (alpha--α) expressed as confidence (1 – α). Although α is usually set to 0.20 at COTF, there are occasions where 0.20 is considered too risky in which cases α is set to 0.10 or 0.05. (2) If a continuous response variable is used, standard deviation (sigma--σ) based on prior test data or an estimate. At times, there are no prior data on which to estimate standard deviation. In these cases, there are two options. The first is to have SMEs estimate the range of values of the response variable, after which sigma is estimated roughly by dividing the range by 4. If the range cannot be estimated, the signal-to-noise ratio (SNR) is used. SNR is explained in the Best Practices in Statistical Analyses documents posted in the Y drive 01C Best Practices folder.

Generally speaking, SNR can be thought of as a standardized effect size expressed as a multiplier applied to the (unknown) standard deviation. The size of SNR corresponds to the sensitivity of testing a factor or interaction effect. Although there is no hard and fast rule for setting the SNR, a SNR value of 1.0 - 1.5 for a factor is often considered reasonable. However, there may be occasions where larger values are used due to practical sample size limitations or expectations that the effect is not important unless large differences across levels of factors/interactions are shown.

(U//FOUO) Confidence (1-α) was set to 80%. Cargo Type is the most important factor, with four levels. Table 2-2 shows the relationship between changing sample size and effect size in examining this factor. SMEs decided that 15 minutes (i.e., 1.5*sigma) is an operationally meaningful effect size for comparing the main effect of Cargo Type on ST. Based on the choice of effect size, three replications of the test design were needed to provide 24 runs, yielding the indicated power. Power for testing other main effects and interactions at this chosen sample size is presented in Table 2-3 along with the related SNR.

Table 2-2 focuses on the "most important" factor. This factor drives the overall sample size for the test. "Most important" refers to one of two factor characteristics. (1) In comparison to the other factors, it has the largest number of levels and thus requires the largest sample for test. (2) It may be judged to have the most operationally impactful effect on SUT performance. In Table 2-2 *Cargo Type* is identified as the most important (or driving) factor. The table shows the tradeoff between power, sample size, and effect size and provides the basis for weighing risks and choosing a reasonable sample size. There may be designs for which table 2-2 is not appropriate to explain the chosen effect size. In this case, use an appropriate method to explain why the test sample size is chosen.

Table 2-2. (U) Power Analysis for Cargo Type (Type I Error rate set at 0.20)

Sample Size	Effect Size (presented as SNR)			
	0.5	1.0	1.5	2.0
8	30%	54%	67%	76%
16	44%	65%	75%	84%
24	55%	74%	**82%**	90%
32	65%	81%	88%	95%

Table 2-3 is an efficient way of summarizing power analysis at the factor level for all factors and logically follows Table 2-2. The table shows the tradeoff between power and effect size for the remaining factors or interactions not presented in table 2-2.

Table 2-3. (U) Power Analysis for Other ST Main Effects/Interactions (Type I Error rate set at 0.20)

Factor Effect	Effect Size (presented as SNR)			
	0.5	1.0	1.5	2.0
Light	40%	66%	78%	86%
Cargo Type by Light	27%	56%	69%	79%

2.3.1.3 (U) Response Variable – Probability of Successful HWM Transition (P_{HWMTS}), M20

This is an example in which the RV is binomial rather than continuous. With a binomial RV, standard deviation and SNR do not ordinarily apply, and effect size is expressed in terms of intervals on the binomial scale (e.g., 0.30 – 0.40, 0.40 – 0.65, etc). Effect size, sample size, desired confidence and power are reported as in the case with a continuous RV.

(U) The objective is to characterize P_{HWMTS} across the levels of the controlled factor (section 2.3.1.3.1.1 below). P_{HWMTS} is the probability that transition across the HWM can be achieved. It is measured as the number of successes divided by the number of attempts. P_{HWMTS} is a discrete variable with a binomial distribution. The specified threshold is 0.80. The expected result for the most difficult loads is 0.80, while the easiest load should allow performance at 0.95. Because no historical test data exist, a range of possible results are considered.

2.3.1.3.1 (U) Conditions

2.3.1.3.1.1 (U) Controlled Conditions (Factors)

Notice that only one factor is controlled. Although this is a simple experimental design, the format for describing the design is quite similar to a full factorial or more complex designs.

Also notice that the reader was referred to a previous section of the document to view the levels of the Cargo Type factor. This approach streamlines and reduces repetition and complexity of the document. Use of this option may not always be possible because of instances where the levels of a factor may vary from one RV to another.

(U) One factor is hypothesized to substantially affect P_{HWMTS}:

- (U) **Cargo Type (C 4.2) – 4 levels (Load 1, Load 2, Load 3, Load 4)** described above in section 2.3.1.2.1.1.

2.3.1.3.1.1 (U) Constant Factor Levels (conditions)

(U) One condition will be held constant:

- (U) **C 4.5 Well Deck Type (set to LSD 41/49 class)**
  This level of the factor was purposely chosen because it requires the most distance to be traveled and, as such, will likely have the greatest impact on the response variable. FOT&E will demonstrate onload of certain cargo types in other well decks but are not a part of the statistical design.

2.3.1.3.1.2 (U) Recordable Factor Levels

(U) The conditions listed below are likely to impact P_{HWMTS}, but are treated as recordable because of the intrinsic difficulty to control them during test, and/or due to SME determination that the impact should be minimal:

- (U) **Terrain Slope (C 1.1.1.3)**
  If the average steepness or grade of the landing area is steep, HWM transition may be affected.
- (U) **Sea State (C 1.2.1.3)**
  The roughness of the seas may impact the ride of the ship, causing difficulty at HWM transition.

2.3.1.3.2 (U) Test Design

(U) Three replications of the four-level, single-factor design will be run resulting in a total of 12 test runs. The rationale for the number of replications is presented in Section 2.3.1.3.3 below. Logistic Regression, confidence intervals, and graphical displays will be used to analyze the data, characterize P_{HWMTS} across the operational envelope, and compare results to threshold. Table C-8 describes all runs.

2.3.1.3.3 (U) Statistical Power and Sample Size

(U//FOUO) Confidence (1- α) was set to 80%. Table 2-4 shows the relationship between changing sample size and effect size in examining the Cargo Type factor. SMEs decided that 0.1 is an operationally meaningful effect size for comparing the main effect of Cargo Type on P_{HWMTS}. Based on the choice of effect size, three replications of the test design were needed to provide 12 runs, yielding the indicated power.

Table 2-4. (U) Power Analysis for Cargo Type (Type I Error rate set at 0.20)

Sample Size	Effect Size (binomial proportion value)			
	0.05 = 0.85-0.80	0.07 = 0.88-0.80	0.1 = 0.90-0.80	0.15 = 0.95-0.80
4	30%	54%	67%	76%
8	44%	65%	75%	84%
12	55%	74%	**82%**	90%
16	65%	81%	88%	95%

In the next section, there is an optional description of Critical Measures associated with the current targeted mission. If the Critical Measures are well understood, or well described in Appendices C and D, the test team may choose not to describe them here.

2.3.1.4 (U) Critical Measure Discussion

When there are no controlled factors, the test objective may be to characterize the critical measure in an overall sense. In this case, the following wording might be used: "The objective is to characterize overall performance on critical measures {name the CM} by computing its mean and confidence interval, and comparing to threshold" {if a threshold exists}.

(U) Critical measures associated with the AMW mission requiring explanation are described below. Appendices C and D provide detail on data collection for all measures.

2.3.1.4.1 (U//FOUO) M1: Maximum Payload Capacity (MPC) (74 short tons)

(U//FOUO) The MPC will be assessed by observing the transport of military vehicles and equipment culminating with the M1A1 Main Battle Tank configured with the TWMP weighing 74 short tons. Planned demonstrations of MPC will conform to the factorial design to be executed in conjunction with vignette IT 1-5 (Transit) while conducting the Design Reference Mission (DRM). The transit vignette varies the fuel type, transit surface, and transit distance to provide a robust set of conditions for evaluating the SSC's PC in all operationally realistic conditions. MPC is not expected to be significantly affected by any of these conditions.

2.3.1.4.2 (U//FOUO) M49: Well Deck Craft Load Time (WDCLT) (≤55min)

(U//FOUO) WDCLT is the time required to onload prestaged loads. It is measured from the point in time when the craft comes off cushion in the aft-most spot of an LSD-41 Class well deck until the time the craft master requests a green well and reports that SSC is ready to depart. WDCLT is a continuous variable assumed to be log-normally and independently distributed. The specified threshold is a mean of 55 minutes (4.0 in LN).

- Confidence Interval: 80% 1-sided [3.9-4.0 (49-54 min)]
- Recordable conditions: None
- Sample size: 4.

2.3.2 (U) E-2, MOB

2.3.2.1 (U) Critical Tasks and Measures

(U) Table 2-5 delineates SSC's critical tasks comprising MOB. The associated critical measures appearing in the table will be used to resolve the MOB mission. Table B-4 illustrates how the tasks, subtasks, and all measures of the MOB COI relate.

Table 2-5. (U) MOB Critical Tasks/Measures
UNCLASSIFIED//FOR OFFICIAL USE ONLY

Task Title	Critical Measure	
2.3.2 – Detect and Track Contacts	M45	Surface Contact Detection and Tracking Capacity (Maximum Number)
	M46	Surface Contact Detection and Tracking Capacity (Minimum Range)

Table 2-5. (U) MOB Critical Tasks/Measures UNCLASSIFIED//FOR OFFICIAL USE ONLY		
Task Title	**Critical Measure**	
	M47	Surface Contact Detection and Tracking Capacity (Maximum range)
	M48	Radar Coverage Capability
2.4.1 – Maintain Situational Awareness	M26	Situational Awareness Capability
	M63	Command, Control, Communications, Computers, and Navigation (C4N) Equipment Interoperability
	M72	Human Factors (Commonality)
2.10.1 – Receive Ship's Services	M2	L-Class Ship Well Deck Compatibility and Interoperability
	M8	Receive Ship's Services
	M18	Manpower

There are no response variables for the MOB COI given that all critical measures listed above are non-stochastic or qualitative in nature. Appendices C and D adequately describe the conduct of tests, primarily demonstrations of task execution under varying conditions.

2.3.3 (U) S-1, Reliability

2.3.3.1 (U) Critical Tasks and Measures

(U) Table 2-6 delineates SSC's critical measures for Reliability. There are no critical tasks for this COI. Table B-4 illustrates how the tasks, subtasks, and all measures of the MOB COI relate.

Table 2-6. (U) Reliability Critical Tasks/Measures	
Critical Measure	
M82	Mean Time Between Operational Mission Failure – Hardware ($MTBOMF_{HW}$)
M83	Mean Time Between Operational Mission Fault – Software ($MTBOMF_{SW}$)

BLANK PAGE

SECTION 3 - TEST EXECUTION

3.1 OPERATIONAL EVALUATION APPROACH

Describe the approach to conduct the independent evaluation of the system. Identify the periods during integrated testing that may be useful for operational assessments and evaluations. Outline the approach to conduct the dedicated IOT&E and collect data for COIs resolution. Each relevant phase of test (OA/IT/IOT&E) should be described.
EXAMPLE

OT&E of the SYSTEM ACRONYM will be conducted in three phases. At the completion of IT-B1, an OA, (OT-B1), will be conducted to investigate risk areas identified during IT-B1 and identify additional risk areas associated with effectiveness, suitability, and survivability. Data from IT-B1 and OT-B1 will be used to provide decision makers with an assessment of risk associated with the successful completion of IOT&E of the SYSTEM ACRONYM in support of the Milestone C decision and subsequent Low Rate Initial Production (LRIP) decision. The second phase of OT&E, Initial OT&E (IOT&E) (OT-C1), will be conducted independently by COMOPTEVFOR at the completion of IT-C1 to evaluate the effectiveness and suitability of the SeaDragon™ UUV as well as the readiness of the system for Fleet introduction based on relevant test data available from all phases of testing. The final phase of dedicated OT&E, FOT&E (OT-C2), will be conducted as necessary to support deficiency correction or to evaluate system capabilities not tested during IOT&E.

3.1.1

To the greatest extent possible, all three OT&E phases will be conducted in an operationally representative environment with Fleet crews and equipment. Because surveyed and instrumented underwater ranges are essential to accurate data collection, operating environments will be carefully selected from among available ranges to create a realistic threat environment. Threat-representative opposing forces, environmental conditions, and target types, locations, and orientation will be incorporated into test events to maximize operational realism of these range-based test events.

3.1.2

Because the planned conditional variations may not be encountered at the time of test, the OT team will review actual test conditions associated with each subtask for each completed vignette to identify any resulting test limitations. Data will

be analyzed throughout IT. The OT team may determine that adjustments in vignette design, procedures, and/or conditional variations are necessary based on this ongoing analysis. This analysis may also lead the OT team to recommend the ITT pursue regression or follow-on testing, especially with regard to design/configuration changes.

3.2 OT VIGNETTE STRATEGY

Identify the vignettes that will be used to collect OT data. Title and a brief description are sufficient. Point towards appendix C for details on each vignette. If appropriate (i.e., multiple phases of test are planned), provide a table of planned test vignettes to be executed during the relevant test period. EXAMPLE

3.2.1 Vignettes

The vignettes used to exercise the SUT are summarized below. Full descriptions of each vignette, including the data required for each, are included in appendix C.

- IT-1-1, HDCM. To evaluate ISIS in an environment with multiple smaller contacts, such as trawlers and pleasure craft, an HDCM vignette will be conducted. This vignette will focus on how well ISIS supports the ship in managing contacts and maximizing Closest Point of Approach (CPA) to contacts of concern.
- OT-6-2, Suitability. Suitability COIs will be assessed using data collected during the suitability vignette which will be run in parallel with the vignettes. [list as appropriate] -DEMO conducted at XXX in XXX will also be used to augment the suitability data set.

3.2.2 Schedule of Events

Briefly describe the plan to execute the vignettes during each test phase.

3.3 MODELING AND SIMULATION (M&S)

Describe the key models and simulations and their intended use, including key threat simulators and/or simulation(s). Include the OT objectives to be addressed using M&S. Identify data needed and the planned accreditation effort. Identify how the OT scenarios will be supplemented with M&S. Identify who will perform the M&S verification, validation, and accreditation. Make sure there is understanding of the capabilities and limitations of the model. Adhere to COMOPTEVINST 5000.1B.

3.3.1 Example Model 1 (E-1, S-7)

Detail the specific model, including a very short summary of how that model will undergo VV&A. Explain the necessity to use modeling and the contribution to test, providing for validity of data to OT. Identify any critical measures or 1st level tasks for the respective COIs that the models will support evaluating. State the plan for COMOPTEVFOR accreditation.

EXAMPLE

Conduct of the ISR, STW, and SUW missions by the Waycool UAS will be tested in a robust EW environment. The Woodstock Offensive Signal Generator (WOSG) will be used to produce a multitude of EW signals ranging from standard navigation radars to threat illumination radars to provide additional data needed to evaluate measures MX through MY. The signal list was selected from anticipated parameters of the operating environment and from ONI threat radar data. The WOSG was verified and validated prior to IT-B1 and accredited for this test by COMOPTEVFOR. Insufficient signal density at the test range required that simulation of EW be used to properly evaluate the Detect and Defend capabilities of the UAS. In addition, the use of threat simulations is necessary to determine AV survivability throughout mission execution. Detection of and reaction to threat signals is an essential system function.

3.4 LIMITATIONS TO TEST

The subsequent paragraphs should identify by category (severe/major/minor) each limitation and the COI(s) they affect. Examples include threat realism, resource availability, limited operational (military, climatic, CBNR, etc.) environments, limited support environment, maturity of tested systems or subsystems, safety, etc., that may impact the resolution of affected COIs. Each limitation shall be binned in the appropriate paragraph (2.4.1 through 2.4.3) and include:

(1) Descriptions of limitation
(2) Description of measures taken to mitigate the limitation
(3) The COIs affected in parenthesis after the title of each limitation.

3.4.1 Severe Limitations

The following limitation(s) precludes COI resolution and will adversely impact the ability to form conclusions regarding effectiveness and suitability.

3.4.1.1 Example Limitation (E-1)

Provide description of limitation, measures taken to mitigate limitation. (List COIs affected in parenthesis following the title of each limitation.)

3.4.2 Major Limitations

The following limitation(s) may affect COI resolution but should not impact the ability to form conclusions regarding effectiveness and suitability.

3.4.2.1 AV Operations Training (E-1)

Provide description of limitation, measures taken to mitigate limitation. (List COIs affected in parenthesis following the title of each limitation.)
EXAMPLE

Testing of the ISR mission will be limited by the absence of a formal training course in AV operations. Fleet operators running the system will attend a preliminary course given by the manufacturer and be allowed several weeks of preparatory flight operations prior to testing. By completing these items, several Fleet operators should be available to represent proficient control of the system under mission conditions. Surveys completed by the operators will assess the validity of the training and provide feedback into the construction of standard Navy training. Personnel slated for giving the training course to future operators will also go through the preliminary

manufacturer course. They will complete surveys targeted at comparing this training to other Navy courses. Future test events and surveys will allow tracking of the evolution of the training program from this version to the final product. This limitation may affect COI resolution.

3.4.3 Minor Limitations

The following limitation(s) has minimal impact on COI resolution and will not impact the ability to form conclusions regarding effectiveness and suitability.

SECTION 4 - CONSOLIDATED RESOURCES

This table is intended to support resource requirements for TEMPs and subsequent test plans. Resource requirements first need to be identified for each vignette. Given the resources required for each vignette and the execution plan/schedule for those vignettes (discussed in paragraph 3.2), OTDs should be able to identify resource requirements for each phase of test.

Quantify the testing sufficiently (e.g., number of test hours/operating days planned, test articles, test events, test firings, manpower, range requirements, etc.) to allow a valid cost estimate to be created for TEMP input. On test articles, detail the number of days required (in port/at sea). For manpower, be specific on how many personnel required, how many days they are needed, and what level of training/expertise they must have. These tables are intended to support the identification of resource requirements needed for TEMP input. Test targets and expendables should include the type, number, and availability requirements for all targets, weapons, flares, chaff, sonobouys, etc. required for testing. Operational force test support includes specific aircraft, ship, submarine, unit, or exercise support requirements (COMOPTEVFOR personnel) including flight hours, at-sea time, or system operating time.

EXAMPLE

4.1 TEST EVENT RESOURCES

Table 4-1 lists the resources required for each test phase. This matrix is based on the current understanding of IT and OT progression.

Table 4-1. Test Event Resource Matrix

Phase Identifier	Test Articles	Test Sites and Instrumentation	Test Support Equipment	Test Targets and Expendables	Operational Test Force Support	Simulations, Models, and Test Beds	Manpower and Personnel Training	Special Requirements
IT-B2	Mission Planning Station Version 3.2 MPS software	Mission planning lab (5 hr)	Stopwatch	None	OTD/Analyst	None	Fully trained operator	None
OT-B	Mission Planning Station Version 3.2 MPS software	Mission planning lab (5 hours)	Stopwatch	None	OTD/Analyst	None	Fully trained operator	None
IT-C	Complete UUV System	Systems integration lab (10 hours)	UUV lift kit	None	OTD/Analyst	None	Fully trained operator	XYZ failure mode database
OT-C	Complete UUV System	NUTEC Range (4 days)	None	Mk 55 bottom mine (inert) (3) Mk 3 moored mine (inert) (4)	OTD/Analyst SSN (w/installed system) (4 days)	VMS Mine Simulation System	Fully trained operator	None

Use style Heading 7 for all APPENDIX level headings.

APPENDIX A - STATISTICAL DEFINITIONS

Use style Heading 8 for all appendix level 2 headings.

A.1 MBTD DEFINITIONS

The following list of terms is useful in understanding how attributes and measures will be used in evaluation. See COMOPTEVFORINST 3980.2 for

Use style Heading 9 for all appendix level 3 headings.

A.1.1 Attribute Allocation (Measures of Effectiveness (MOE), Measures of Suitability (MOS), SoS)

- MOEs and MOSs come from system documents (CDD, CPD, ORD, FRD). MOEs contribute to the assessment of system effectiveness, while MOSs contribute to the assessment of system suitability.
- SoS attributes do not apply to the SUT, but are applicable to the overarching SoS. Although data may be collected and these measures may be reported, they do not impact the resource requirements for a minimum and adequate test.

A.1.2 Measure Types (Specified, Derived, Other)

- Specified measures are extracted directly from the reference JCIDS Capability Document.
- Derived measures are extracted from other authoritative source documents (Navy Tactics, Techniques, and Procedures (NTTP), system specifications, Concept of Operations (CONOPS), etc.).
- Other measures are those measures that apply to the SUT but do not have a documented source. These include metrics based on subject matter expertise or created by the OTA to assess a particular task.

A.2 DOE GLOSSARY

The following list of statistical terms with definitions may be useful.

A.2.1

Null Hypothesis (H_0) – the proposition concerning system performance put forward at the beginning of a test and assumed to be true. The test data can be used to reject the null hypothesis or fail to reject it. If test data indicate that H_0 should be rejected, then the alternative hypothesis is considered supported. If test data do not indicate that H_0 should be rejected, then there are two possible explanations: (a) the statistical test was not sufficiently sensitive to reject H_0, or (b) the H_0 is, for practical purposes, true.

A.2.2

Alternative Hypothesis (H_1) – the proposition concerning system performance that is expected. If the null hypothesis is rejected, then this hypothesis is supported.

A.2.3

Type I error – Type I error occurs when H_0 is wrongly rejected.

A.2.4

Type I error rate – the probability of Type 1 error; also referred to as alpha (α).

A.2.5

Statistical confidence – the probability of not making a Type I error ($1-\alpha$).

A.2.6

Type II error – in a hypothesis test, a Type II error occurs when the null hypothesis is not rejected when it is in fact false; that is, $H_{0\ is}$ wrongly not rejected.

A.2.7

Type II error rate – the probability of Type II error; also referred to as beta (β).

A.2.8

Statistical power – the probability of avoiding a Type II error ($1-\beta$).

A.2.9

Effect size or delta – the expected, planned-for difference a test is designed to detect. In test planning, the size of the effect size has direct impact on the sample size in the test. The sensitivity of the test is related to the effect size that it can detect (detecting smaller effect sizes means a more sensitive test). A test is considered more sensitive when the effect size is small. Therefore, effect size is sometimes used to describe the sensitivity of the test. This parameter is used in test planning to estimate the appropriate sample size and ensure adequate power to detect the stated difference.

A.2.10

Standard deviation (σ) – a statistic that assesses the run-to-run variability of the response or “critical” variable used in the test.

A.2.11

Power analysis - the process of estimating a minimum sample size to meet the preplanned α, β, σ, and δ; or the process of estimating 1-β given the preplanned α, σ, δ, and sample size.

A.2.12

Controllable conditions/factors – a controllable variable that is thought to influence the response. The specific values of a factor are called levels.

A.2.13

Mean – the arithmetic average of a set of numbers.

A.2.14

Median – the numerical value in a data set below which 50% of the values falls.

A.2.15

Condition or factor – a variable that is thought to influence the response variable.

A.2.16

Test design – complete specification of the organized test runs with respect to controlled conditions. Where necessary, disallowed combinations and replications are included in the specification.

A.2.17

Test of one proportion – a test that statistically compares the observed binomial proportion to a threshold.

A.2.18

Test of two proportions – a test that statistically compares two binomial proportions.

A.2.19

One-sample t-test – a test that compares the sample mean to a threshold.

A.2.20

Two-sample t-test – a test that compares two sample means.

A.2.21

Main effect – difference in the response variable attributed to or caused by a single factor.

A.2.22

Interactions – an interaction between two factors implies that the differences between levels on the first factor change as a result of the levels of other factor.

A.2.23

Logistic regression – a statistical technique for analyzing dichotomous response variables. Through logistic regression, the effects of factors and their interactions on the dichotomous response variable can be assessed.

A.2.24

ANOVA (Analysis of Variance) – a set of statistical methods for examining the main effects and interactions of multiple controlled factors.

A.2.25

Normal distribution – a continuous distribution that shows the greatest frequency of occurrence around a central mean (bell curve). This is the most common approximation used for continuous variables such as detection range, time to complete a series of tasks, etc.

A.2.26

Binomial distribution – the theoretical distribution that applies to critical variables with two discrete outcomes (e.g., hit and miss, or detect and no detect outcomes). The binomial distribution is used to analyze binomial proportions, defined as the number of “successes” (however defined) divided by the total number of events.

A.2.27

$\chi 2$ (“Chi-squared”) distribution – a theoretical distribution occasionally used with response variables taking a skewed shape. The distribution is also used to construct confidence intervals for standard deviations from populations that follow a normal distribution.

A.2.28

Poisson distribution – a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event. This distribution can be used to estimate the occurrence of false alarm or false contact report rates.

A.2.29

Rejection-based null hypothesis test – a test in which the assumption from the outset is that the system under test does not meet threshold. Rejection of this form of the null statement leads to the conclusion that the system under test meets threshold. This type of test is typically used for measures where the criterion is mission critical.

A.2.30

Acceptance-based null hypothesis test – a test in which the assumption from the outset is that the system under test meets threshold. Rejection of an acceptance-based null hypothesis leads to the conclusion that the system under test does not meet threshold. This type of test is typically used for measures where the criterion is not mission critical.

APPENDIX B - MISSION AND CAPABILITIES ANALYSIS

The following tables are provided:

- Table B-1. Conditions Directory – a listing of conditions that are controlled or recorded to support post-test analysis.
- Table B-2. Attribute Matrix – a listing of all attributes and measures used to assess effectiveness and suitability of the SUT.
- Table B-3. Orphaned Attributes (if applicable) – attributes identified in requirement documentation that OT will not report on.
- Table B-4. Traceability Matrix – a linkage of the operator tasks (for each COI) to the measures and conditions associated with those tasks that will be used to assess the performance of the SUT.

Critical tasks and measures are presented in red. Items that are highlighted in gray are retained for traceability but do not apply to the SUT or are out of scope of this IEF. Definitions for acronyms used in the enclosed tables can be found in appendix X, Acronyms and Abbreviations.

AppB Workbook SUT mm-dd-yyyy.xlsx

Table B-1. Conditions Directory
Table B-2. Attribute Matrix
Table B-3. Orphaned Attributes
Table B-4. Traceability Matrix

The appendix B tables should be exported from the IEF database with 01B support.

For electronic routing, the exported Excel workbook can be inserted here, vice converting the excel output from the IEF database into MS Word format.

For routing or review of a hardcopy, the individual tables can be printed directly from Excel and inserted after this page. If the OTD sought 01B assistance, headers/footers and page numbers are automatically formulated and included in this workbook (as is this case for this example). .

Guidance for each of the tables is included in the example Excel workbook.

BLANK PAGE

APPENDIX C - TEST DESIGN

C.1 VIGNETTE-TO-SUBTASK-TO-CONDITIONS MATRIX

The embedded Excel file below contains the Vignette-to-Subtask-to-Conditions Matrix for each vignette and displays the operator tasks associated with each vignette, the controlled conditions for that vignette, and the resulting run matrix (if applicable). Each vignette is shown on an individual tab in the workbook.

AppC-A Workbook
SUT mm-dd-yyyy.xls›

Table C-1. Vignette-to-Subtask-to-Conditions Matrix (IT 1-1-1)
Table C-2. Vignette-to-Subtask-to-Conditions Matrix (IT 1-1-2)
Table C-3. Vignette-to-Subtask-to-Conditions Matrix (IT 1-1-3)

This workbook shows the conditional variations explained in section 2. Verify that the run matrix shown in the tables matches the discussion in section 2. The IEF database will export the shell for these tables. Certain fields (i.e., the run matrix) vary significantly dependent on the DOE and have to be populated manually. Seek assistance from the 01B CTF. The DOE notes section should simply contain a summary of the DOE results discussed in section 2 (RV, Type of Test, Effect Size, Confidence, Sample Size, Power, etc).

C.2 VIGNETTE DATA REQUIREMENTS AND TEST METHOD MATRIX

The embedded Excel file below contains the Vignette-to-Data Requirements-to-Test Method for each vignette. It identifies the data that testers need to collect during each vignette and describes the test method used to execute the vignette. It also captures the tasks and measures associated with each vignette.

AppC-B Workbook
SUT mm-dd-yyyy.xls›

Table C-4. Vignette-to-Data Requirements-to-Test Method Matrix (IT 1-1-1)

Data requirements should direct test participants to observe and record specific items that are needed to confirm satisfactory subtask performance based on measures. Each measure must be confirmed by data. Provide a full understanding of that data as part of the vignette. Listing data requirements by measure gives visibility to the adequacy of test. Note: While the inclusion of data sheets is optional, defining data requirements is not. Data requirements for each measure should be documented under each vignette and can be organized as appropriate. Although the template shows several examples of how data requirements are grouped, this is left to the OTDs discretion. Test methods should detail what will occur during the event, and what testers must do to collect the required data. Ensure test method has a logical flow that can be easily understood. Recommended headings (Pre Test, Test Execution, Post Test) are shown in the example but may be modified at the OTDs discretion.

The IEF database will export the shell for this workbook. As for the previous tables, the narrative entered in the data requirements and test method will be entered and formatted manually. Seek assistance from the 01B CTFs.

BLANK PAGE

APPENDIX D - DATA REQUIREMENTS

Table D-1, provided below, shows the relationship of measures to the data requirements that must be collected to satisfactory resolve each measure.

AppD Workbook SUT
mm-dd-yyyy.xlsx

Table D-1. Measures-to-Data Requirements Matrix

BLANK PAGE

APPENDIX E - EVENT RECORDS AND SURVEY

Provide data cards, logs, and surveys to be used by the OTD during IT and OT. If data requirements in the vignettes rely on the existence of data sheets, then each of the listed sheets must be created. Label each data item/survey question with the measure that it is intended to answer. Note: if data sources are not yet understood, data sheets are not required for the IEF.

EXAMPLE

E.1 QUESTIONNAIRE

The questionnaire will be filled out by all test participants.

- Operator Qualification/Experience Questionnaire

E.2 DATA SHEETS

The data sheets below will be completed by the system operators and other test participants under the supervision of the OTD.

- D-1, Sonar Operator Data
- D-2, Fire-Control Operator Data

E.3 EVENT LOGS

The event logs below will be completed by the OTD or trusted agent with the assistance of system operators and other test participants

- L-3, TRACKEX Log
- L-4, M-DEMO/Repair Log

E.4 SURVEYS

The surveys listed below will be administered per the data requirements of each vignette. Fleet personnel are the primary targets of the questions, but other test participants and trusted agents are also eligible to complete them.

- S-5, Mission 1 Effectiveness Survey
- S-6, Mission 2 Effectiveness Survey
- S-7, M-DEMO Survey
- S-8, Training Suitability Survey

Often, those participating in testing are asked to fill out a general questionnaire to gather information about the roles and duties of the participants.

This questionnaire usually asks for the participants name, rate/rank and their watch standing role; schools attended and training; years using the system and in the service. Tailor the questionnaire to the relevant background information you will need to help make sense of the answers they will be providing about the system.

This questionnaire is neither a data sheet nor a survey, and so it does not have a number. It is not marked FOUO, so do not include information that could jeopardize the participants' identities.

A basic example is on the next page.

Operator Qualification/Experience Questionnaire

Name: _____ Date: _____

Rate/Rank: _____ Years in Service: _____

NEC: _____ Watch Station: _____

School/Training: (list schools and training, including date attained)

School/Training	Location	Date

Data Sheet D-1
Page 1 of 1

Data sheets are often in tabular form, and can be inserted as a graphic, especially if electronic data sheets in an existing data collection system will be used. (In such cases, obtain a screen shot from the application; it will need to be a good quality graphic, at least 300 dpi.)

Each data sheet will be numbered consecutively, at the top right corner of the page.

A basic example data sheet (Word table) is on the next page.

Remember to use **Next Page** section breaks to separate portrait-oriented data sheets and landscape-oriented data sheets. Do not put blank pages between data sheets, or between the data sheets and the surveys. Blank pages are only inserted between sections and appendices to keep them starting on an odd-numbered (restarting at 1 for each section/appendix) page.

Since surveys have an FOUO marking in the heading, you will need to use the section break between the last data sheet and the first survey (as shown on the next page.)

SYSTEM Hardware Failures CLASSIFICATION		
Description of Hardware Failure	**Time Failed/Time Corrected (hh:mm:ss/hh:mm:ss)**	**Description of Fix**

Surveys will begin numbering from the last data sheet (they will not start numbering at "1").

Surveys will not always carry the same classification as the rest of the document, and may be For Official Use Only when filled in. If this is not the case, remove the FOUO statement in the header. In some cases, it is classified when filled in. Insert the appropriate classification notification in the header.

Examples of header markings:
Unclassified, but Secret when filled in
Unclassified, but Confidential when filled in

Format 16 point, Courier New font, bold, in the center of the header. Using a text box (as in the header on this page) keeps the page numbers aligned with the rest of the pages in the document.

Surveys sheets can be formatted in many different ways, and often have rating scales for agreement to a provided statement. Please remember, if using a table to format your survey that the survey number and page information must be on each page of the survey, so you will need a separate table on each page.

Surveys will often include instructions for completing, either one general page/paragraph of instructions if consistent across all surveys, or one instruction page/paragraph for each survey.

An example of a basic survey is on the following page.

Reliability Survey

Describe purpose of survey. Explain rating scale (if one will be used).

This is an example of a rating scale. Yours may be like this or not.

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Not Applicable
1	2	3	4	5	N/A

Your questions would then be listed, something like the following:

1. System XYZ performed all the necessary tasks for my watch station.

1	2	3	4	5

2. System XYZ did not fail to operate during high tempo operations.

1	2	3	4	5

Your survey may also have short answer or multiple-choice questions. Just make sure that you explain how each type of question needs to be answered.

Make sure you provide space for comment after each question, or section of related questions.

-
-
-

BLANK PAGE

Use acronyms from the CAAL; if the correct acronym is not in the CAAL, request that it be added.

CHOOSE AN ITEM.

APPENDIX F - ACRONYMS AND ABBREVIATIONS

There is a 2-column table for entering acronyms and abbreviations. Insert rows as necessary. Leave a blank row between each alphabet grouping. In addition to acronyms in the body of the document, ensure that all acronyms used in the appendix B and C tables (which may not have been previously defined in the document) are captured here.

ASDS	Advanced SEAL Delivery System
CAAL	COMOPTEVFOR Acronym and Abbreviation List
CSRR	Common Submarine Radio Room
SEAL	Sea-Air-Land

BLANK PAGE

If the document is classified but the title isn't, place a (U) after all references listed with the complete title. If the title is classified, use the appropriate classification for it.

APPENDIX G - REFERENCES

List all references used in construction of this IEF. Include all documents called out as sources of SUT and SoS attributes. Also list anything used to create tasks and conditions including documentation on the kill chains. Include any IEFs used for comparison.

EXAMPLE

(a) COMOPTEVFORINST 3980.2, Operational Test Director's Manual of 1 Jun 12

(b) COMOPTEVFOR PIN 10-01, Operational Reporting Guidance and Procedures of 2 Mar 10

(c) Previous IEF version of date (if applicable) (list all)

(d) SYSTEM ACRONYM Test and Evaluation Master Plan (TEMP) No. XXXX of date (U)

(e) SYSTEM ACRONYM ORD/CDD/CPD of date

(f) COMOPTEVFOR ltr 3980 Ser 54/S231 of 23 Aug 11

In this example, reference (d) is a classified TEMP, but the title is unclassified. Reference (f) is a classified letter, which is depicted by the "S" in the serial #. References (a) and (b) are examples of unclassified references.

BLANK PAGE