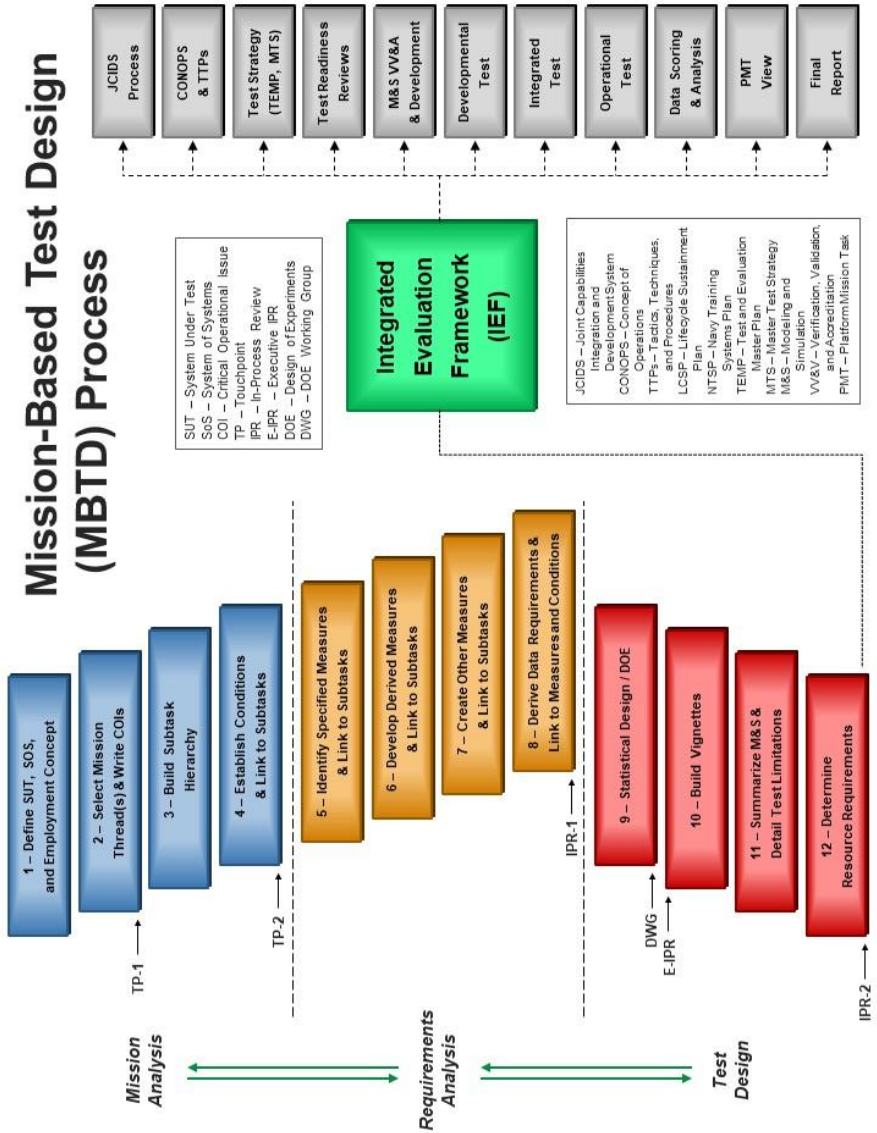
Operational Test and Evaluation Force

MBTD / IEF Checklist

Purpose: This step-by-step checklist leads the Operational Test Director (OTD) through the entire Mission-Based Test Design (MBTD) process. Details within the MBTD steps include development of sections in the Integrated Evaluation Framework (IEF), and use of the Mission-Based Test and Evaluation System (MBTES). The diagram below is the MBTD process. For any questions, **contact your Core Team Facilitator (CTF) first**.



The MBTD process relies heavily on the Subject Matter Experts (SME) in the warfare divisions. The OTD (and their test team) follows this checklist (to the extent appropriate) for each program, for each MBTD effort. Blue font indicates portions applicable to the Software Acquisition Pathway (SWP), though the entire checklist may apply by the time such programs complete Operational Test (OT).

The checklist is broken up into smaller checklists based on significant reviews:

- Touchpoint 1,
- Touchpoint 2,
- In-Process Review (IPR)-1,
- Design of Experiment (DOE) Working Group (DWG),
- Executive IPR (E-IPR),
- IPR-2,
- Several Annexes.

Tailored IEFs (T-IEF) are constructed using the applicable steps of this checklist, as directed by the Warfare division Assistant Chief of Staff (ACOS) and the 01B Director. E-IPR is done for TIEFs the director will sign.

NIPRNET resources available to the OTD beyond this checklist (mostly found at Y:\OT&E Production Library\IEF) include:

- OT and Evaluation (OT&E) Manual Chapter 4;
- IEF and T-IEF templates, as well as E-IPR briefing template;
- OT Analysis handbook, and OT M&S instruction;
- Suitability handbook appendices;
- Running Comment Resolution Matrix (CRM) (RCRM) template;
- MBTES training course, MBTES user Manual, and MBTES training slides.

SIPRNET also has copies of the templates. Others resources include:

- USN Required Operational Capability/Projected Operational Environment (ROC/POE)
- US Fleet Forces Warfare Capability Gap Tool (WCGT) (https://intelshare.intelink.sgov.gov/sites/WCGT/SiteAssets/Reference.aspx)

Process Admin

Purpose: Admin actions done before and during the process.

☐ 1. Read OT&E Manual Chapter 4 MBTD section.

It contains vital MBTD details beyond the contents of this checklist.

☐ 2. Get a CTF and a 01D representative assigned.

Contact the 01B deputy first. If not available, contact the lead CTF to determine who will facilitate your MBTD. Contact 01D deputy.
CTFs do not work exclusively within divisions. By assigning programs from many divisions to each of the CTFs, we hope to achieve commonality of execution across the building.

☐ 3. Get "tool" (MBTES) access.

MBTES is an integral part of all OPTEVFOR test operations. Many of the tables in IEFs and test plans are produced directly from MBTES. Any issues discovered in test will be tracked in the Issues Module.

a. Enter access request in MBTES, including the program(s) the user will work and the access needed. Non-OTDs/OTCs, notify OTD/OTC of need to endorse the request in MBTES.

An account can only be created using this path because MBTES access on NIPR/SIPR is based on the user's CAC/Token.
User group assignment gives different permission for MBTES operations. A single user can have different permissions for different databases. MBTES support will reject or modify any requests that are not accurate here.
MBTES Support must verify need to know. Personnel listed with the program in iBOSS are automatically approved. Access for contractors is granted through OTD request. If contractors associated with a program change, the OTD must inform their CTF or MBTES Admin.

b. MBTES Support grants appropriate permissions and emails the requester that access has been granted.

☐ 4. Get a database for your new program.

It is likely your program already has a database. If not, complete this step.
All database manipulation is completed by the MBTES Support per direction of the program CTF at the OTD's request.

a. Request a blank database, or

b. Request a copy of an existing database created for a different program.

Similar systems should have similar MBTDs. Copying an existing database for a similar system could save work, as many MBTD products you need would already be in the tool. However, it could create more work to identify and delete MBTD contents not related to your SUT. Consult with your CTF on the best option.

☐ 5. Manage your database.

The database management rules below cannot account for all situations. Consult with your CTF as required to clarify what you need and why.

Each TEIN gets one database. Issues are tracked/searched by TEIN. Having multiple databases breaks that paradigm. Programs with multiple databases must transition to a single database as soon as workload allows.

MBTES will have the ability to work one database for multiple TEINs. This function should be used rarely, and only with carful consideration of the implications to test design. A team might use this when they expect all/most test events to yield reporting on all/most included TEINs. MBTD efficiencies are created when the same DRs apply to different program measures, tasks/conditions are common across multiple programs etc. However, extra MBTD work will be needed/wasted if many test events will be done for single programs, forcing the test team to spend time excluding material. A TEIN included in a multi-TEIN database cannot have a separate database.

Some TEINs have databases on both the classified and unclassified network. This option is highly discouraged; only do it after careful consideration. More work will be needed to keep both MBTDs up to date. To minimize this work, one of the databases would be the full MBTD, while the second database only has the minimum needed (likely just a subtask hierarchy, and set of issues).

a. Backup databases often.

Non-read-only user groups can complete. The tool is not foolproof, and neither is the test team. Backups can be restored by MBTES Support if necessary. These backups only save for 30 days.

b. Archive databases that are no longer useful.

A database is useful as long as the system is in the fleet, not just for ongoing test. Archiving likely needs endorsement at the Warfare ACOS level.

MBTES Admin completes. These databases are no longer visible to MBTES users. Archived copies are saved indefinitely.

c. Do not delete databases.

MBTES Admin completes very rarely. Each database is the record of issues (or lack there-of). They will be retained in the active list or in the archive.

d. Do not lock IEF or Test Plan nodes.

In this context, "nodes" represent specific documents that will include MBTD items created in MBTES. IEF nodes are added/edited in the IEF Module. Test plan nodes are added/edited in the Test Plan Module. Locking these nodes prevents changes from being made in the node, and prevents changes in the Base Layer (described below) from affecting the node.

It is tempting to lock these nodes when the associated document is signed. In almost all cases, that would be a mistake. The IEF node needs to be unlocked in case MBTD during test planning leads to base layer changes that must flow into the test plan node. The test plan node must remain unlocked so base layer changes during reporting can populate as needed.

MBTD is the river; documents are the fish. The MBTD flows and changes with the current program/mission/sustainment/survivability information. The documents are only pulled out when needed. Locking stops the river's flow. Locking any node should be preceded by extensive discussion with the CTF.

MBTD does not stop when a document goes into the router. Often before the IEF is signed, changes are made to the MBTD through the test planning process. If the IEF is far along in the router, adding such changes may be cumbersome (e.g., giving measures to the front office that have not been reviewed by the division ACOS). A locked node would support re-export of some IEF products if changes were directed during the review, but only if those changes are not facilitated by the base layer. It may be simpler to work such changes outside the database (e.g., editing an old measures matrix in excel), while ensuring MBTES remains the definitive MBTD (e.g., capturing measures changes in MBTES at the same time as excel). Thus, nodes remain unlocked.

☐ 6. Work your database.

Documents (IEFs, test plans, reports, etc.) are merely snap-shots of the temporary status of a program's MBTD on the signature date. The MBTES database is the current MBTD, free to change through good process execution at any time, no matter the status of document signatures. Do not incorporate review comments solely into documents; capture them in MBTES too.

MBTES has many functions to help users create MBTD. "Standard" suitability and/or Cyber Survivability (CS) measures and DRs can be pulled from internal lists. Tasks, conditions, and measures can be copied from other program databases. The more a user knows about MBTES, the faster they work and the better they execute MBTD.

a. Develop the MBTD fundamentals in the base layer.

The MBTES "base layer" is the first four tabs, used to work COIs/tasks, conditions, measures, and Data Requirements (DR). These MBTD components are fundamental to the SUT components/functions, the intended capabilities, and the associated employment concept. Thus, they apply across the life of the program. Base layer inputs are pulled in to specific documents (nodes) by using the IEF Module and Test Plan Module.

b. Create new IEF nodes as needed.

The IEF module allows test teams to create as many nodes as required. An IEF node is likely needed for each MBTD iteration (e.g. drafting an IEF, providing TEMP inputs, and completing a test). Name the node for the program increment, phase, or similar convention that matches the node's scoping. Transitioning to the next phase of test (EOA to OA, OA to IOT&E, etc.) means a new TEMP. This almost always means a new MBTD effort, and a new node.

Parallel MBTD efforts can exist within a program (e.g. an incremental program writing a test plan for one increment, and working an IEF for the next increment). Thus, multiple IEF nodes may be worked at the same time.

c. Create new test plan nodes as needed.

The Test Planning Module enables creation of multiple test plan nodes within an IEF node. At minimum, a node is needed for each planned final report.

d. Keep the MBTD current.

MBTD does not stop; our design should evolve based on any new information. Thus, MBTES is not stagnant. For example, it does not matter that the test plan published a PV-0 with a bad measure. We should upgrade the measure, and the PV-1 in the test report should use the better measure version.

Be careful not to break prior agreements without proper coordination. A signed IEF captures the minimum-adequate OT scope (resourcing, data, etc.) to which stakeholders agreed. A significant test design change (e.g., reduction in resourcing) should be coordinated at the O6 level following established processes.

☐ 7. Store documents.

Despite being the official location for the living MBTD on any test program, MBTES does not have input fields to capture all of the work done to design a test. There is no place to enter a SUT description, a critical subsystems matrix, an employment concept exploration, a full DOE write-up, a test phase's limitations, etc. IEFs and test plans must be saved.

a. Verify signed documents are in iBOSS.

This is the official repository. When an IEF completes routing, it is saved here.

b. Upload documents to MBTES.

Duplication of iBOSS archiving is unnecessary, and risks version control issues.

Not all design documents get signed. Files such as SWP workbooks can be uploaded to a program database. Emails with sponsor approval of derived requirements are perhaps the most important to capture.

☐ 8. Talk with your CTF, LTE, and 01D representative often.

Don't complete the entire checklist for a meeting and only consult your CTF preceding the review meeting. Get clarification early. Keep your CTF involved, especially with big MBTD decisions.

Engagement with 01D during MBTD is critical and ensures the OTD develops an informed cyber strategy tailored to the SUT and enable early collaboration with the warfare division, program office, and other stakeholders to align cyber T&E strategies and support future test planning efforts.

01D: usn.hampton-roads.optevfor-norva.list.01d-mil-civ@us.navy.mil

☐ 9. Identify/contact members of the Core Team.

MBTD is done by the Core Team. These are OPTEVFOR people and any stake-holders whose input will be needed to create an accurate and comprehensive IEF. Members of the program's T&E WIPT should all be considered as possible Core Team membership. At minimum, contact the Program Manager (PM), the program T&E lead, Developmental Test (DT) rep (if independent from PM), resource sponsor, OPNAV T&E rep (N942), the Fleet user community, associated Navy Warfare Center (NWC), and DOT&E (oversight programs). For Joint programs, coordinate with the other involved OTAs to consult the equivalent personnel in the other services (Joint PMs, sponsors, user reps, etc.). The VXs are vital. Get them involved right away on MBTD.

☐ 10. Manage IEF schedule and status in iBOSS PM.

a. Determine the driver for IEF completion.

An IEF does not have an independent due date. The IEF timeline is driven by when MBTD products will be needed to support other test efforts. Most commonly, drivers are TEMPs and/or Test Plans. Other drivers include test considerations with long-lead times such as M&S identification/development, Test Resource Requirements (TRR) inputs for development, POM (budget) inputs, and targets/surrogates allocation.

b. Schedule MBTD steps in iBOSS PM.

Consult with your CTF. IEF timelines vary greatly depending on team workload, program complexity, inputs maturity, and IEF scope.

The primary scheduling method is starting at program entry to the building, and setting the review/signature dates based on the "standard" timeline for the program type, phase, and oversight status.

The secondary method should only be used with leadership approval based on an inability to start immediately. Set the start/review/signature dates by counting backwards from the driver date. By delaying the start, this approach avoids rework on MBTD products created from early/flawed information. However, it accepts risk of missing the due-date should the process be delayed.

c. Update iBOSS PM as required to reflect MBTD progress, IEF completion expectations, or driver changes.

☐ 11. Keep all IEF stakeholders in the loop.

IEF contents affect many other processes and documents owned by players outside OPTEVFOR. At minimum, the Core Team members should be given a chance to comment on our products.

a. Send MBTD products out for review/comment.

b. Adjudicate comments and/or prepare comments for discussion at IEF reviews.

c. Inform leadership of significant disagreements.

d. Invite stakeholders to IEF reviews as appropriate.

Major decisions that impact testing trade-space (duration, cost, quality) may be made at these review meetings. Completion of the three major reviews (IPR-1, DWG, IPR-2) along with routing of the associated RCRMs serves to confirm stakeholders agree to the MBTD produced so far. Stakeholders must be informed of this paradigm, and be encouraged to participate so that the opportunity agree becomes active agreement. If a stakeholder cannot attend, their concurrence with the design can still be gained through sending out the products and confirming they have been reviewed.

IEFs are approved by OPTEVFOR. Stakeholders do not "approve" these documents, making the review meeting agreement paradigm even more important.

☐ 12. Conduct IEF reviews.

Details on, and expectations for each meeting are provided in the related checklist of steps leading to that review meeting.

General principles and practices:

- ***Goal – Each meeting is intended to fully cover and approve the new MBTD products. Ultimately, the meeting is being held to receive Division ACOS/Deputy ACOS (DACOS) and 01B/01B1 approval for the new items. Approval does not preclude later changes.***
- ***Read-aheads – participants must have the chance to be ready at each meeting to provide official input on the MBTD. Get read-aheads out soon enough for them to complete a sufficient review.***
- ***Introductions – Make sure everyone in the meeting knows each other. Knowing the commenter can help understand the comment.***
- ***Discussion – These are working meetings. Going in, nothing is final. Feedback from various participants may have been received/adjudicated, but that does not remove their ability to make further comment. Discussion is guided by the OTD.***
- ***Disagreements – Ideally, all disagreements will be resolved before the meeting is complete. At minimum, the substance and reasoning behind the differing positions must be fully discussed. For oversite programs, the RCRM is used to document disagreements at IPR-1/DWG/IPR-2.***

(continued)

- ***Minutes and action items – The OTD is responsible for minutes/action items being recorded during the meeting and reviewing them at the end. Meeting minutes may be the only place that key agreements on MBTD products are recorded. Meeting minutes often include the CRM, updated to capture changes agreed-to at the meeting. CRM items (which include significant comments that come up during the meeting) that remain unresolved are also captured in the RCRM.***

Materials:

- ***Old material – All the prior products (previously reviewed) will be available for reference, but there should be no need to go back and review that material (comments still allowed on old stuff).***
- ***New items – Be ready to review all the MBTD material as listed in each specific meeting's directions.***
  - ***Print a set of hard copies for the top-level reviewers (Div ACOS, 01B, perhaps DOT&E). The Front-office will create the copies they need at the E-IPR; only print extras for non-front office reviewers if preferred.***
  - ***Bring the materials up electronically for other reviewers to see.***
  - ***Ensure all stakeholder comments are available and ready to be reviewed as that section is reached. If a CRM is used, the most up-to-date CRM is a meeting input (Note: the RCRM is only a meeting output).***
- ***Briefs – Briefs are discouraged for all MBTD reviews except the E-IPR. It is vital that OPTEVFOR and stakeholder leadership approve the actual MBTD materials.***
- ***References – Have the following available: Capabilities Document, CONOPS, previous IEF (if applicable), other significant source material.***
- ***IEF production metrics – Brief how the executed process timeline lines up with the "standard" timeline. If the timeline is "behind schedule" (moving slower than the "standard"), brief the specific delays/barriers.***

Overall Timeline:

- ***Scheduling***
  - ***Meet the minimum lead-time for each meeting (e.g., 2 weeks for IPR-1, DWG, and IPR-2), or get ACOS/DACOS approval for less, as this incurs risk to RCRM use. More lead time is better for in-person attendance.***
  - ***Consult the CTF on anticipated product readiness by the meeting date. Ideally, the CTF will agree the team's progress indicates success at the review, and will agree with scheduling. The meeting can be scheduled without CTF concurrence.***
  - ***Read-aheads can go out before, after, or with the calendar invite. Before is better, enabling better products to accompany the invite.***
- ***Duration – The scheduled duration of each meeting should account for the amount of material to be reviewed, and the complexity of possible discussions/disagreements.***
- ***Updated read-aheads – Review all stakeholder comments prior to the meeting. If any being accepted require significant product changes, it may be prudent to make those adjustment prior to the meeting and send updated read-aheads. Send these updates at least two days prior. Discuss***

(continued)

with CTF prior to sending, as this is a risky strategy. It may be easier to work through the old product, but capture intended adjudications in a new version or in a CRM (recommended approach).

- *Publish meeting minutes and action items – Send these items for stakeholder concurrence/comment within two working days after the meeting. Stakeholders agreements must be documented in meeting minutes so they are not lost over the long MBTD timeline. This helps ensure accountability throughout the process. Template is available at Y:\OT&E Production Library\IEF.*
- *Disseminate RCRM for oversite programs – The full RCRM process is described in the next step.*

IEF Production Metrics: Collect and record the information called for in the 01B metrics tracker. Do this leading up to the meeting, and after it's done.

☐ 13. For all programs, use the RCRM.

Not all stakeholder disagreements can be resolved at the warfare division level. Some must be elevated to senior leadership. The RCRM is the mechanism behind identifying, managing, and elevating such issues. It is required to be used for all IEFs. A blank RCRM is available at Y:\OT&E Production Library\IEF.

Agreements between stakeholders go in meeting minutes, not the RCRM. Internal OPTEVFOR comments do not go in the RCRM. Separate CRMs are used for these.

The RCRM enhances communication, enables early identification of issues, promotes accountability for OTA and DOT&E staff, and facilitates senior leader communication prior to formal end-game document routing. All OTAs have signed up to use RCRMs.

There are three RCRMs, one for each of the major MBTD reviews (IPR-1, DWG, IPR-2). Comments are added for the current meeting. New comments cannot be added after close-out of the RCRM.

A pillar of the RCRM process is the rule to get read-aheads out to stakeholders 2 weeks prior to major MBTD reviews (IPR-1, DWG, IPR-2). The ability to close out the RCRM hinges on stakeholders being accountable for the material covered at the meeting.

a. Document disagreements in RCRM

Comments are added when agreement cannot be reached at the O6 level (decisional authority for organizations) during MBTD reviews. Document in the "Comment" columns.

Inputs must be specific. The "Recommended Change" column must include a detailed solution that, if implemented, would fully address the issue from the perspective of the stakeholder providing the input (this can be refined post-meeting via stakeholder review).

b. Send draft RCRM for stakeholder review.

Sent within 2 work days after the MBTD meeting (starts a two-week review period). Initial documentation of comments must be reviewed by stakeholders to ensure technical accuracy of what is recorded, and awareness of the disagreement. Even if the RCRM is blank, send it out for review (stakeholders still acknowledge).

As part of this review, the stakeholder who made a specific comment has an opportunity to refine the contents of the "Comment" columns. Any time changes are made to these columns, the version number is increased (e.g. version column is blank, then becomes 2).

During the review period, DOT&E and the program office can add new comments they think were not captured in the draft RCRM.

After any edits/additions to the "Comments" section, the RCRM must immediately be sent to OPTEVFOR for further dissemination. These changes must be reviewed in the two-week window.

c. PM, DOT&E, and OPTEVFOR AOs establish their positions on each of the comments as part of stakeholder review.

These are the organization's O6 (and equivalent) stances on written comments. Columns only exist for OPTEVFOR, PM, DOT&E. Other stakeholders work with one of these three to input their positions on RCRM comments (e.g. IDA comments are owned and entered by the DOT&E AO, Sponsor comments go in with the PM). O6 positions are due for all comments within 2 weeks of first receiving the RCRM. For OPTEVFOR, the AO is the Warfare Division ACOS.

d. Close-out RCRM; start AO-level clock.

Two weeks after the RCRM is sent following the MBTD review, no additional comments can be added; the RCRM is finalized.

RCRM items have 90 days from close-out to be resolved at the O6 level. When the time limit is reached, any remaining items must be resolved at the executive leadership level (OTA Executive Director, DOT&E Deputy Director, PEO).

Working-level discussion continues during the 90-day time limit. Changes are not made to the Comment and Stakeholder sections.

e. Close out comments resolved at O6 level.

Once agreement is reached, the final resolution is input to the resolution section. These comments require no SES review/resolution.

f. Elevate comments that bust the AO-level adjudication time limit, or are obviously irresolvable.

Comments that cannot be resolved at the O6 level may be elevated sooner than the 90-day limit.

g. Update MBTD based on top-level agreements.

The IEF is updated to reflect the agreement(s).

h. OTDs provide weekly status updates to the CTF.

Weekly updates are consolidated by the CTFs. A status updated is sent to the Warfare ACOSs/DACOSs for the RCRMs with comments.

The blank RCRMs are not tracked at the O6 level, but OTD status updates to the CTF are vital. Blank RCRMs are just as important as the rest.

i. Notify leadership of critical comments that violate the principles of the RCRM.

Stakeholders should not have new comments on material already covered under the RCRMs. If they do, this is a party foul. Inform 01B if it happens.

☐ 14. Leverage a CRM for faster MBTD review and document routing.

A CRM can be provided as a meeting read-ahead (input), capturing concerns to be discussed during formal IEF reviews. It can then be updated for meeting minutes (output), documenting changes agreed-to during the formal review. It is not used to track unresolved disagreements from the review; that is done in the RCRM. A CRM template is available at Y:\OT&E Production Library.

Do not turn the CRM into a rolling history of all past comments. This could lead to rehashing the same arguments, violating the principles of the RCRM. The CRM should be empty at the start of review for each MBTD meeting.

The Warfare Division owns the product. Ultimately, the Warfare Division ACOS has final say on IEF content routed for approval.

01B will provide feedback to the OTD within two weeks of IPR-2. If the product reviewed at IPR-2 was a near-finalized IEF, 01B can be skipped in the document router.

☐ 15. Record lessons learned.

The MBTD process is not stagnant. This checklist is occasionally updated with lessons learned by core teams. The best time to record these and provide them to 01B is after the MBTD reviews.

Do not include lessons learned with meeting minutes or other documents published outside OPTEVFOR.

☐ 16. Disagree with Program documents as needed.

MBTD products can disagree with official program documents when OT understanding of that particular content (such as CDD CONOPS) is different. Ideally, creation of such divergent content early enough leads to clarification and agreement between T&E stakeholders before IEF signature and before testing. When IEF products include divergent content, it may be prudent to clearly state what is different, why, and where to find the official version. The underlying goal is to collaborate with all stakeholders to ensure understanding, and correct testing.

☐ 17. Acknowledge what is known and unknown.

The IEF is often signed before all of its contents are set in stone. It is important to acknowledge (clearly write down) what is known, expected, and unknown. Many MBTD items can simply be documented "TBD" if that is the best we know at time of signature.

☐ 18. Execute MBTD appropriate to the test phase.

EOA is unlikely to include testing of the actual SUT and/or employment by the Fleet. MBTD must reflect this. SUT/SoS, tasks, conditions, and measures are likely to be the same as any other test phase, as EOA conclusions speak to the likelihood of achieving acceptable performance in these areas. But DRs, DOE (if any), vignettes, limitations, and resources must be written to the actual expected testing. Likely vignettes include meeting attendance, publication review, and table-top discussions.

OA should offer similar data collection to IOT&E. IEFs to support OA may not have unique MBTD products like EOA. But other contents like events and limitations will certainly be different.

The IEF can recognize MBTD content that does not apply to the current phase (e.g. measure and tasks that will be covered by future testing may be added, but grayed out).

An IEF can cover multiple test phases. When this is the case, close attention must be paid to the parsing of MBTD products, such as applicability of DRs to vignettes (e.g. DRs only available for collection at IOT&E can't be listed in vignettes for OA).

Any observation of DT (leading to either a Letter of observation (LOO) or Assessment of Operational Capability (AOC)) needs no MBTD. These events would benefit from having a task structure to which insights could be associated, but have no need for measures/DRs. Preparation for the DT event would benefit from examination of capabilities and CONOPS, but doesn't need precise definition of the SUT. Thus, getting to a TP-2 is helpful, but going to an IPR-1 is excessive.

IEFs are built (and reviewed via the regular meetings) in order to support OT. The MBTD components (measures, DRs, etc.) needed to be of the highest quality possible (based on available information), as if test planning for IOT&E was imminent.

☐ 19. Consider impacts of previous testing.

All MBTD is informed by SME knowledge of the mission and the SUT. This includes knowledge gained in prior test. What DRs were not valid? What measures need to be added? What conditions need to be adjusted? How can DOE leverage existing data?

Often (especially in DOE) it will be appropriate to reference prior OT in the text of the IEF. Review all previous Response Variable (RV) results in the data appendices of previous reports.

☐ 20. Notify 01B1 when your program is using CBTE.

CBTE, completed by the program, is the DT partner to MBTD. OPTEVFOR needs to be aware when we can leverage this parallel process.

☐ 21. Participate in CBTE as required.

At minimum, OPTEVFOR will support the CBTE WIPTs.

CBTE shifts DT focus from specification compliance verification to assessing system capability to support missions in the larger context of the SoS. Among other things, proper coordination for a CBTE program should result in cost avoidance/savings. Cost avoidance is the reduction of up-front OT scope through the use of IT vice dedicated OT. Cost savings is the reduction in OT execution below that planned, based on the scoring of DT data for use in OT.

☐ 22. Execute steps applicable to Software Acquisition Pathway (SWP).

As noted earlier, SWP steps are colored blue. Only the top-level (numbered) checklist steps are colored blue. All of the content below those (lettered steps; amplifying info boxes) are black font, but still apply to SWP.

Because DOE is expected to be rarely included in SWP, the DWG section has no blue font. Employ these steps as appropriate.

TP-1, TP-2, DWG, and E-IPR meetings are not recommended for SWP. Teams can still do them if it will benefit the process.

a. Develop SWP TIEF in MBTES.

Producing the IEF Word file is not required for SWP. The MBTD is documented and approved via MBTES.

b. Document MBTD material not captured by the TIEF in other documents as appropriate.

Several steps (e.g., Define SUT/SOS) are done as part of SWP MBTD effort (colored blue), but not documented in MBTES. This material is usually written in the IEF. It should be captured in the Master Test Strategy (MTS), or some other program document to which OPTEVFOR is a contributor and/or signatory. If appropriate, these files should be reviewed at the IPRs in addition to MBTES outputs.

☐ 23. Leverage MBTD principles to complete Level-of-Test Determination (LTD).

LTD is described in the OT&E Manual, chapter 4. Further direction is provided in the LTD briefing template. But that policy direction for LTDs is not exhaustive. This is deliberate. The LTD process is meant to be flexible, allowing test teams to support their unique program and the specific decision's context as fast as possible.

a. Complete mission analysis steps.

Don't write an IEF document; don't work in a database; don't create MBTD specifics. Rather, work through all of the considerations of the first four MBTD steps to fill in the applicable LTD brief slides.

Because LTD supports an initial decision (for a program, an increment, an upgrade, etc.), there is often a limited amount of information to support the decision. Coordinate with stakeholders early to minimize the unknowns.

The system capabilities in-scope for the LTD are the most important consideration. Sometimes measure development aids in capabilities understanding, but this work is not required.

b. Draft LTD brief.

Only complete the slides that will aid decision-makers.

c. Schedule meeting and provide read-aheads.

If possible, follow the timeline and practices for scheduling an IPR/DWG. LTD leverages the RCRM process, which states the read-aheads must be out at least two weeks before the meeting. Because LTD is a rapid/flexible process, this read-ahead time can be collapsed to 2 work days, but doing so accepts the risk that stakeholder disagreements may not be resolved at the meeting or be immediately added to the RCRM.

d. Conduct the determination meeting.

Internally, ACOS/DACOS attendance is required for the warfare division and the 01B/01C/01D competency divisions. Briefers should consult with their CTF on expected meeting progression.

e. Issue the LTD letter.

☐ 24. Build the minimum-adequate test, even if stakeholder test strategy negotiations are expected to change that scope.

When OT input to a test strategy document is required, that input will be pulled from the IEF. Because initial strategy inputs are sometimes reduced through Working Integrated Process Team (WIPT) negotiations, the approved strategy document may eventually differ from the signed framework. For example, the IEF may list a test resource that the TEMP removes. The TEMP accounts for that removal by adding a limitation the IEF did not have.

Usually these negotiations happen during the MBTD process, but sometimes limited information or new circumstances make test design changes inevitable. Do not anticipate such changes. Write the IEF to ask for the test we want, including all the necessary resourcing. For example, a vignette may need a target surrogate that does not yet exist. Call out the target in the resources table if it is expected to be resourced prior to OT; do not add a limitation. If the surrogate is never developed, the new limitation

(continued)
can be added to the strategy, the plan, and/or even the report. On the other hand, when there is no funding to deliver the surrogate, add the limitation to the IEF, thus driving TRR inputs.

Touchpoint 1 Checklist

Purpose: Building the reference library, scoping the IEF purpose, defining System Under Test (SUT)/System of Systems (SoS), initial mission analysis, and COI selection.

☐ 1. Warfare division A Code invite O6 counterparts (PM, Sponsor, Fleet SME, NWC, OPNAV T&E AO, DOT&E AO) to participate in the MBTD process as members of the core team.

☐ 2. Gather and review the applicable reference documents for the SUT.

a. ORD, CDD and/or CPD

These are the overarching OPNAV and Joint Requirement Oversight Council (JROC)-approved capabilities documents. System requirements are not always approved in this form. Make sure you have the most current specified requirements.

OTDs need to be aware of the schedule and plan for updates to the CDD or CPD. The program office and/or OPNAV resource sponsor can provide this info. All OTDs must ensure OPTEVFOR is aware of and is included in the review process. OTDs shall review their JCIDS requirement documents and submit recommended changes to OPNAV and the Program Office when they are routed for review.

b. Test & Evaluation Master Plan (TEMP)/MTS.

A program office document that requires OPTEVFOR concurrence. The program office T&E IPT lead manages all TEMP/MTS updates and is the primary source for updates and status of this document.

c. Concept of Operations (CONOPS)

May also be referred to as a Concept of Employment (CONEMP). Usually provided by the user community and/or the resource sponsor (OPNAV). Brand new programs may not have one or may rely on an older one for a legacy system.

Evaluation of the Acquisition CONOPS is a new focus for OT, especially early in the process (MS A TEMP). If there is no Acquisition CONOPS, a minor limitation to test must be added to the IEF and the TEMP.

d. Information Support Plan (ISP)

A program office document that captures information technology requirements and interfaces in sufficient detail to enable testing and verification of those requirements. Some may contain useful mission task breakdowns for the SUT.

e. Program Protection Plan

This is another Cyber reference. It will be vital to have this as the Cyber-triad does supports the MBTD effort.

f. ROC/POE

Defines the primary mission areas and expected environment for the overarching platform. Not all programs will have a platform or system specific ROC/POE. These should align with the overarching USN ROC/POE – OPNAVINST C3501.2K CH-1 (classified document).

i. Platform specific ROC/POE

ii. USN ROC/POE

Each ROC/POE mission has an OPTEVFOR-approved mission thread made up of standard/required 1st-level subtasks. All 1st-level subtasks for each mission COI must be included in the IEF. The mission thread repository is available at Y:\T&E\Mission Thread Repository.

g. Functional Requirement Documents (FRDs)

Typically, a document written by the program office, for the program office to identify requirements for their testing; typically, at the sub-system level. May also be used to capture requirements for upgrades/modifications.

h. DODAF Architectures and Systems Command Integrated Capability Framework (ICF)

DODAF architectures can be gathered through the program office or the resource sponsor (OPNAV). The most useful will likely be the OV-1, OV-5, and OV-3/SV-6. These documents will be key to drafting SUT/SoS paragraphs. Thus, a description of all DODAF products relevant to MBTD is provided in the SUT/SoS Annex at the back of this checklist.

i. Review the CDD References and gather those deemed appropriate. References cited in a CDD/ CPD are sources for specified requirements and should not be overlooked. (i.e. VOLT, TTPs, SEP, TTVRs, etc.)

Even if these Fleet documents are not listed references in the CDD, they still may be vital to MBTD. TTPs, VOLTs, TTVRs, and other Fleet documents often levy requirements on mission accomplishment, detail task execution, and identify conditions in the mission environment that must be included in a comprehensive MBTD.

j. Security Classification Guide

The Security Classification Guide is a required reference for every OPTEVFOR test document (Y:\OT&E Reference Library\Security Classification Guides). Contact the vault if further help is needed.

k. The Warfare Capability Gap Tool

l. Tactical Situations (TACSIT)

TACSITs provide Red Order of Battle (OOB), doctrine and TTPs, Blue OOB, doctrine and TTPs, environmental details, C2, ROE, and more based on current OPLANs. They are Fleet documents.

m. Mission Technical Baselines (MTB), and Initial Capability Technical Baselines (ICTB).

These documents come from the acquisition community. MTBs describe the threat (Red OOB, TTPs, doctrine, etc.). ICTBs cover neutralizing threats (Blue OOB, TTPs, doctrine, etc.).

n. Previously signed IEFs

These are not formal references. However, consistent MBTD across related or similar systems is vital to test. For example, subsystem IEFs must line up with their platform IEFs in many ways.

o. Operational Availability Handbook (OPNAVIST 3000.12A), and the MOA on MOT&E and Operational Suitability Terminology and Definitions

There is a long list of references for effectiveness. It is equally important that suitability is consistent with approved guidance.

☐ 3. Establish the purpose of this IEF.

The purpose paragraph informs all MBTD steps by scoping the IEF. An IEF can cover multiple test phases and include multiple TEINs; or be much simpler. Discuss support for TEMPs, test plans, VCD, etc. For TIEFs, it should be easily understood from the purpose why a full IEF is not required.

If this is an IEF Revision, be sure that is clear, along with why the IEF is being revised (new CPD, next increment, etc.).

☐ 4. Draft the IEF purpose section.

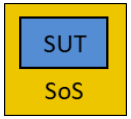
This step begins drafting of IEF section 1, which provides necessary context for all other MBTD material (capabilities explain measures, CONOPS explain tasks and conditions, etc.). Little (often none) of section 1 is input to the TEMP. Only a portion is input to test plans. Therefore, this section should include the minimum-adequate information necessary to provide that MBTD context.

☐ 5. Define the SUT.

Defining the SUT is the process of categorizing elements of a system designed to meet a set of requirements and provide the capabilities needed by the Fleet. Correct SUT and SoS definitions are vital to proper scoping of test designs and detailed test plans, as well as correct determination of deficiency assignment between blue and gold sheets during the post-test reporting process. Defining SUT/SoS requires close collaboration with the program office and resource sponsor for proper interpretation of the mission requirements and operational architectures of the SUT. Focus on the appropriate test scope leading to proper test execution.

The SUT is the hardware and/or software, as well as the logistics support, technical manuals, training, and trainer devices being developed/delivered to meet the requirements set by the resource sponsor and to provide the capabilities needed by the Fleet. This definition is supported by the Concept of Operations (CONOPS), program Concept(s) of Employment (CONEMP), Tactics, Techniques, and Procedures (TTP), and manpower requirements associated with the system.

The SUT is almost always equivalent to what is described with the TEIN. The figure depicts the typical SUT and SoS relationship up through IOT&E.



Guiding principles for development of SUT descriptions:

- ***A platform SUT may employ many subsystems (some with their own TEIN) to complete a mission (weapons, fire control, etc.). These subsystems are part of the defined platform SUT.***
- ***The SUT definition is influenced, shaped, and clarified by requirements documents (CDD/CPD/FRD/ORD, gap analysis, etc.), the developmental system CONOPS/program CONEMP, and kill/ effects chains.***
- ***Ideally, the SUT definition will match the program office definition. However, full agreement on SUT definition is not required.***
- ***The SUT description should tie the procurement of the system to the overarching capability gap/enhancement it is planned to address within the kill/effects chain and SoS.***
- ***The description should include a simple graphic to help clearly define the boundaries of the SUT.***
- ***If multiple test phases with different configurations are covered by the IEF, explain those phases and configuration changes in the SUT definition (IEF only). The test plan SUT description will include the SUT configuration for***

(continued)
that specific phase of test.

a. Define the in-scope SUT (always applicable).

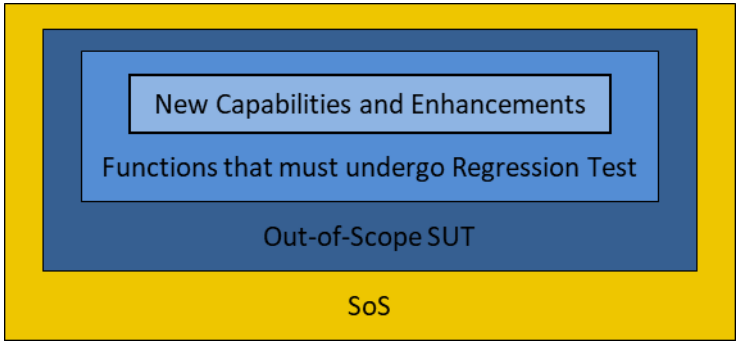
SUT description for the purposes of determining test scope may not remain the same throughout the testing lifecycle. At or before IOT&E, the entire SUT is in-scope (do not use the term in-scope until FOT&E). The full SUT write-up for the MBTD phase includes Components and Capabilities.

i. Components (what it is) - Identify the final/fielding configuration of the SUT, to include major hardware and software components plus any support infrastructure and other non-material elements procured with the system.

ii. Capabilities (what it does) - Describe the capabilities the SUT provides, the capability gaps it addresses, and the system's desired effects.

As the SUT is upgraded, testing must focus on the impact of those upgrades. After IOT&E (FOT&E, follow-on increment, VCD, etc.), the SUT may therefore be further divided between In-Scope and Out-of-Scope SUT. For FOT&E (and the like), the fielding configuration discussion focuses on the new/upgraded/changed hardware and/or software. Also, the capability discussion is broken into three subsections: new capabilities, capability enhancements and regression confirmation. These will inform the Purpose of Test section within the Test Plan.

The figure below depicts a modified view of SUT and SOS as testing shifts from the earliest program OT to subsequent test phases. The SUT is captured within the three blue boxes. The In-Scope SUT is comprised of the innermost two boxes.



Changing SUT and SoS definitions impact all of MBTD. Missions, tasks, conditions, measures, and more could change as the focus of test shifts to the upgrades. Notes throughout the checklist will remind the OTD to consider in-scope SUT.

The in-scope SUT concept affects test reporting, but those implications are not covered in this checklist.

System upgrades, deferred capabilities from earlier test periods, Verification of Correction of Deficiencies (VCDs) and regression confirmation are tested during VCD/subsequent FOT&E phases. The scope of test (the minimum-acceptable data set needed from the test) beyond IOT&E should be limited to the new capabilities, enhancements, or regression confirmations in the new test phase. The new phase of test is being conducted to validate funded capabilities and should focus on those new capabilities/enhancements (includes VCD confirmations requested by the PM) and not on re-testing legacy capability proven in IOT&E, (with the exception of regression testing). Thus, the In-Scope SUT definition is typically limited to the impacts of the upgrade. Upgrades to the system include hardware and software providing/affecting:

- *New capabilities – capabilities/functions never tested because they were not previously provided by the system (includes deferred capabilities from earlier test periods);*
- *Enhancements – existing capabilities/functions previously tested but intentionally targeted for improvement in the upgrade (may include VCD for deficiencies from IOT&E);*
- *Regression confirmations – existing capabilities/functions previously tested and not targeted for improvement in the upgrade, but tested again to confirm performance was not adversely impacted.*

Note: Regression confirmation testing seeks to uncover issues in existing functional and non-functional areas of a system after updates have been made. The intent of regression testing is to ensure that updates such as those mentioned above have not introduced new fault or failure modes in the system. One of the main reasons for regression testing is to determine whether a change in one part of the SUT affects other parts of the SUT (e.g., software functionality changes inadvertently affected functionality not intended to be changed). Common methods of regression testing include rerunning previously completed tests and checking whether program behavior has changed and whether previously corrected faults have re-emerged. Determination of required regression testing should not be undertaken without a thorough understanding of the interfaces associated with the capabilities and enhancements. Consultation with the system SMEs (program systems engineers, developmental test engineers, etc.) as well as a review of system architecture documents may be required to establish the scope of regression testing required.

Suitability is always evaluated as a part of IOT&E and subsequent FOT&E phases. Suitability testing conducted during FOT&E should focus on the new capabilities, enhancements and regression confirmation of the updated SUT, to include updated/upgraded hardware to bring those new capabilities and enhancements to bear (the In-Scope SUT). For FOT&E with In-Scope/Out-of-Scope determination, there are two scenarios that define the strategy for suitability data collection:

- *No SUT Change: In this case, the test is being conducted due to a test limitation during IOT&E that precluded completion of the full scope of testing and required shifting test requirements to FOT&E phases. As the SUT is the same as that tested during IOT&E, suitability data should be collected on the full system and combined with the previous data from IOT&E.*
- *SUT Change: In the case where the SUT is changed for FOT&E phases and In-Scope determination is made, suitability data collection will focus on the In-Scope portion of the system. This will result in a suitability COI determination for the changes that were made to the system and may require updates to critical suitability measures. Suitability COIs should be modified to reflect the focus on the In-Scope SUT components. Suitability measures will need to be updated to reflect component-level requirements, which should be coordinated and agreed to by the requirements officer prior to the E-IPR. In many cases, the testing needed to resolve the reliability, maintainability, and logistic supportability of the upgraded hardware/software will drive the scope of test. Consistent with OPTEVFOR policy, identification of qualifiable data from other test phases may reduce the scope of test required for suitability COI resolution.*

A comparison of the updated requirements documentation, system CONOPS/CONEMP, kill/effects chains, and DoDAF products with those for previously tested increments may help clarify the In-Scope SUT.

b. Define the Out-of-Scope SUT (if required).

The Out-of-Scope SUT includes hardware/software not included in the In-Scope SUT that is responsible for legacy functions/capabilities not impacted by the upgrade. These legacy components and functions are not specifically intended as the focus of test, but are part of the SUT for reporting purposes. Tasks, measures, and data requirements supporting performance of the Out-of-Scope SUT may be included in the test but will not drive scope of test.

For ACAT/BCAT paths, out-of-scope SUT is only defined after IOT&E. Before that, the entire SUT is relevant to test. If a capability is not yet ready for test, it is in-scope so that a limitation can be acknowledged.

On the SWP, MTA, and UONs pathways, out-of-scope can be designated as part of any test effort. These pathways allow fielding that does not have to deliver the full-scope, IOT&E capability. The in-scope SUT is the intended delivery. The out-of-scope SUT is the remainder of the program to eventually be fielded.

The in-scope SUT concept does not limit, impede, prevent, or in any way restrict the OTD's ability to observe tasks, data, and measures; determine mission relation; and ultimately make COI resolution call.

☐ 6. Draft the IEF SUT paragraphs.

<u>Be concise</u>. Only include material that helps to explain the rest of the MBTD. The SUT is the fielding configuration and final capabilities, not the test configuration and capabilities at test.

- ***The SUT written in the test plan is the same as the SUT in the IEF. The test configuration needs to be discussed only in the test plan, and identifies the differences between what will be tested and what will be fielded. These differences include hardware, software, and capabilities.***
- ***When the test configuration differs from the fielding configuration, test conclusions are limited to/by the data collected. Thus, limitations are written to account for the fielding configuration available. The IEF may include these limitations if it is known this will occur (at OA, for example).***

☐ 7. Define the SoS (this may be defined in requirements documents as Family of Systems).

The SoS is the existing (or updated) infrastructure not procured with the SUT, but within which the SUT will function to support mission accomplishment. The SUT falls within, and functions as a part of the SoS. For the purpose of the test, it is important to think of SoS in terms of the <u>SoS supporting the SUT</u>.

a. Determine what other systems the SUT will interface and interact with.

The SoS description must include:

- ***SoS hardware/software components required for test execution;***
- ***Non-material SoS elements such as logistics support and technical manuals;***
- ***Specific functions supporting the test that are required to fully exercise the SUT such as direct interfaces or interactions.***

The SoS as defined for the test is normally expected to include the critical interfaces with the SoS within which the system functions (or that are required to prove the entire SoS kill/effects chain) but may not encapsulate all possible systems as it is unlikely sufficient resources will be available to address the totality of the SoS. Therefore, at a minimum, the SoS interfaces critical to SUT mission accomplishment and COI resolution must be identified (allows for proper resourcing). For instance, a system designed to support a strike group may be limited to interaction with a single ship of a particular class during the test event but is deemed satisfactory to demonstrate SUT integration with the SoS for a test phase.

b. Identify how the SUT impacts the SoS.

OTDs must be aware of the impact SUT deficiencies may have on the SOS. Understanding the SUT/SoS definition/relationship will support categorizing deficiencies as Blue or Gold Sheets.

c. Consider SUT/SoS integration.

Integration of the SUT into the SoS is key to understanding the boundary between SUT and SoS. Integration concerns include physical (components), functional (capabilities), and behavioral (CONOPS). Integration is important to understand the allocation of components between the SUT and SoS.

d. Consider SoS changes throughout the testing lifecycle, and update as required.

For testing beyond IOT&E, the SoS is still thought of as supporting the SUT, but the interactions between SUT and SoS may have been changed or evolved over time.

☐ 8. Draft the IEF SoS paragraph.

Similar to the SUT, SoS determination is informed by requirements documents, CONOPS/CONEMP, and kill/effect chains but also must consider the DoDAF architecture. As such, program DoDAF architecture and SYSCOM ICF mission products are key to understanding of SUT connections that form the SoS, thus helping to define the SoS. For further direction on DODAF products, see the SUT/SoS Annex at the back of this booklet.

☐ 9. Define the SUT employment Concept (how you use it).

a. Examine the SUT CONOPS.

Summarize the acquisition CONOPS (if applicable). Identify the system's operating environment and the end user (operator, maintainer, etc.). This section aids in understanding the COIs, tasks, and conditions. A description for each mission may be appropriate.

b. Determine how the SUT is operated/sustained.

Aids in understanding suitability testing scope. SUT operation translates to duty cycle (treatment of the SUT as continuously operated, on-demand, or impulse). Will the SUT be maintained at O-level or not? How will the SUT be logistically supported? For FOT&E (and the like), focus sustainment on the upgraded SUT.

i. If appropriate, clarify suitability terms in the context of SUT specifics.

Scoring of suitability is based on definitions agreed to in the TEMP. Develop these unique definitions, or clarify those from the CDD.

When does operating time start and stop? Can the system operate in standby? What constitutes neutral time. How are Operational Mission Failures (OMF) defined for the SUT, and does this lead to the use of Essential Function Failures (EFF)? Be comprehensive.

Do not write in generic suitability definitions (e.g. operating time is the time in which the system is on, and being stressed at mission levels). These words add no clarity for the SUT.

ii. Generate the Mission-Critical Subsystem Matrix (MCSM) and/or Mission-Critical Software Function Matrix (MCSFM).

In the MCSM, list the critical components within the SUT. These are the hardware items whose failure are expected to cause an OMF.

In the MCSFM, list the SUT critical software functions. Do not include software dedicated to the operation of hardware. Criticality of this software is covered by listing the hardware in the MCSM.

For each component/function, identify the redundancy and duty cycle. If there are three units, and only one is needed for mission execution, redundancy is "1 of 3".

Within a column for each mission (identified in subsequent step 11), mark each of the rows as appropriate to identify criticality by mission area. If appropriate, include a column for CS.

☐ 10. Draft the IEF Employment Concept paragraphs, excluding the Cyber Concept.

The CS concept is presented at TP-2. Subsequent steps direct the preparation of that material. The CS column of the MCSM and/or MCSFM, are not completed yet either (presented at TP-2).

☐ 11. Determine the Effectiveness COIs.

a. Define the mission-based COIs.

i. Review the Operational Capabilities for each Mission Area in the ROC/POE that could apply.

ii. Identify the Mission Areas that capture the operational capabilities affected by the SUT.

CAUTION: If review of the operational capabilities does not result in mapping to ROC/POE mission areas, and a functional COI is considered, approval must be received from 00 or 00D. 01B CTFs can assist in this decision. Mission and functional COIs can be used together.

iii. Refine selected mission areas into COIs.

If the operational capabilities supported by the SUT are captured under multiple mission areas, but are similar such that there is little difference in how the SUT is used (the tasks that operators perform are the same, with similar success criteria and conditions), then select the most stressing mission area(s) as your COI(s).

☐ 12. Determine the Suitability COIs.

a. Tailor the standard suitability COIs as needed.

Reliability, Maintainability, Logistics Supportability & Availability are the standard COIs. Excluding one or more of these COIs is possible (e.g. a SUT without maintenance has no Maintainability COI).

b. Create additional suitability COIs if appropriate.

Items previously captured in Suitability COIs like Training and Documentation should be evaluated under the appropriate effectiveness or suitability COIs. For example, compatibility aspects like sufficient cooling and power supply can best be addressed through reliability.

☐ 13. Refine the CS COIs.

The CS COIs will mirror all the effectiveness COIs and, if appropriate, some of the suitability COIs (e.g., a training COI is used because the SUT includes a large training infrastructure, and those training systems have significant CS considerations; or the maintenance hardware/software provided with a SUT needs CS testing separate from that of the rest of the SUT). Reliability and availability cannot have CS COIs.

☐ 14. Draft the COI questions.

Effectiveness COI wording should be appropriate to SUT involvement in mission completion. For example, the COI may ask if the SUT supports the mission (e.g. MIW) completion, if the SUT supports a smaller sub-set of the mission (e.g. MCM within MIW), if the SUT supports a specific task (e.g. self-defense within ASW mission), etc.

For IOT&E and prior, the COI accounts for full SUT contribution. With an in-scope SUT, COI language should be modified to align with evaluation of the In-Scope SUT. For example, if the In-Scope SUT consists of the hardware and software components required for a major radar upgrade, the COI may be stated as "Will the radar upgrade to the [SUT] support execution of the [assigned] mission?".

Also considers in-scope SUT when drafting suitability COIs.

☐ 15. For TIEFs supporting Joint COIs, identify how Navy COIs map to Joint COIs.

Other Lead OTAs will not always use our COIs. Complete MBTD using our missions/process. Map the MBTD results to their COIs.

☐ 16. OTD, LTE, and CTF review all Touchpoint-1 products and discuss (ideally agree on) readiness for leadership review.

Readiness for review does not mean all products are 100% correct and agreed to by all participants. MBTD reviews are informal meetings where any issue (ideally within meeting scope) can be discussed. MBTD creates constructive conflict within our matrix organization. Don't let disagreement delay progress.

☐ 17. Request stakeholder feedback on products.

Ideally, continuous core team collaboration should make this step redundant. However, a final/formal request should go out at least 2 work days before the review.

Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

☐ 18. Meet with 01D rep.

This meeting ensures the 01D rep is well-aware of the SUT, and has the necessary information to support upcoming test design. By the end of this meeting, the test team must (through 01D help) understand what is expected for each IEF CS section. The warfare division CTE should develop all initial CS inputs to the IEF, and provide to 01D for review and maturation as required.

When scheduling the meeting, consult with 01D on whether it will be appropriate for any outside stakeholders to attend. The answer will depend on available planning time, program coordination so far, and ability to line up everyone's schedules to meet.

01D is consulted after TP-1 products are refined, and out for stakeholder review (give 01D your TP-1 products at the same time), thus ensuring the material is ready to enable an efficient and productive discussion. Schedule for 1-2hr, depending on program size and CS concept complexity. Discuss/develop the following:

a. Review defined SUT/SoS.

b. Review SUT CONEMP/CONOPS.

c. Review MCSM/MCSFM.

d. Review effectiveness COIs and associated tasks.

e. Review system VOLT.

The test team should already have the VOLT as part of TP-1 preps. If not in-hand, make request and provide to 01D on receipt. Cyber contents of the VOLT must be understood for T&E planning.

f. Discuss cybersecurity concept (threats/defense).

Reviewed at TP-2. This section is vital to understand why CS matters to the SUT, as well as what the SUT/SOS brings to the fight. If possible, prepare a draft of the section prior to the meeting.

g. Refine cyber T&E system information on-hand and additional requirements.

The boilerplate list (CS OT&E documentation support; reviewed at TP-2) in the IEF template is a good place to start. 01D will help reduced/expanded the list as needed, so the OTD can make an accurate request for CS documentation delivery from the program. CS T&E plans/execution will be flawed without the proper references.

h. Discuss cyber DT/OT alignment strategy.

Reviewed at DWG. The cyber T&E roadmap should exist, or be in development. What cyber DT is planned (e.g. 1553 testing), and what will these tests provide to OT? When are they planned? OT can leverage data and refine test given enough lead time.

If they exist, bring a copy of the T&E WIPT charter, the M&S WIPT charter, and the cyber T&E WIPT charter. At minimum, come to the 01D meeting having at least once discussed the cyber T&E strategy with the appropriate program office personnel.

i. Discuss anticipated cyber OT scope.

Reviewed at DWG (see DWG directions). With an understanding of cyber concept, a general scope of all cyber test can be anticipated. Then, the data expected from DT can reduce the scope of events needed for OT. Identify the IT/OT vignettes expected for cyber.

j. Discuss cyber T&E capability augmentation requirements, and other T&E considerations.

Reviewed at DWG. The necessary augmentations are based on the details of the SUT. More complex CS capabilities and/or threats require more complex test. As with any special T&E resource, sufficient lead time may be needed to procure these (e.g. cyber M&S).

☐ 19. Adjudicate/incorporate stakeholder feedback.

If there isn't time or agreement to incorporate the comments, ensure the items are ready to be discussed at the meeting.

20. Schedule TP-1 and provide read-aheads.

TP-1 and TP-2 can be conducted separately, or as a combined brief (preferred).

Deliver the official read-aheads to all stakeholders at least 2 work days prior to the meeting. More than one week of lead time is preferred.

The greater the scheduling lead-time, the easier it is to deconflict calendars. In-person attendance requires more warning. Send the invite at least 1 (but not more than 6) weeks ahead. The meeting can be scheduled before stakeholder feedback is returned.

Products to provide (exported from appropriate MBTES layer, likely IEF node):
- *IEF section 1 (minus CS concept and CS column of MCSM).*
- *Stakeholder feedback.*
- *CRM, if used.*

Personnel to attend:
- *Division – ACOS or DACOS, Section Head, OTD, Contract Support.*
- *Support – 01B or 01B1, CTF, LTE. 01D rep participation at TP-1 is optional (01D rep's decision); still send them read-aheads/invite.*
- *Outside agencies (Including OPNAV T&E rep (N942)) – TP1 should be easy to achieve email agreement. Dial-in may be appropriate, especially for other OTAs. In-person attendance is acceptable.*

Meeting time:
- *Large systems may take 1-2 hours. The smallest systems may take 15 minutes (schedule minimum of a half-hour).*

☐ 21. Conduct Touchpoint-1.

Begin the meeting by setting the classification level and noting if electronic communication means (e.g., telecon) are used. Then do quick introductions. Then summarize how the meeting will progress, and the goal at completion. Encourage attendees to bring up any unresolved comments as that section is reached in the review, and state that the test team will be doing that for anyone not in attendance.

Put section 1 up on the screen and out on the table to be reviewed. Have a folder of reference documents ready to be accessed electronically. Don't read any of the products. Talk to them, and allow for questions. Do more than just announce a section and ask for feedback.

Meeting key concerns and progression:
- *Director's letter – Only talk about this if needed for unique content.*
- *Purpose paragraph – The entire MBTD effort hinges on this content. Discuss it in detail.*
- *SUT configuration – The system makeup and functions must establish a clear dividing line exists between SUT and SoS. Provide a brief voice-over on how well that division is understood.*
  SUT capabilities – The major benefits to mission success provided to fleet must support future measure development. Do not give an overall

☐ *(continued)*

summary; sum up each bullet.
- *SOS – This will potentially drive assets for test. Mention major parts and their relevance to the mission(s).*
- *Mission CONOPS and Effectiveness COIs – Who uses the system, and how, must be understood to support future subtask hierarchy development. Flip to the E-COIs and verify those align with the capabilities/concept. Discuss how missions were combined to create COIs (if applicable). If non-ROC/POE E-COIs are used, justify why the standard missions were not used.*
- *Sustainment concept and Suitability COIs – The details behind keeping the system available must be understood to support future subtask hierarchy development. Flip to the suitability COIs. If any standard COIs were omitted, discuss why. If any non-standard were added, justify.*
- *Cyber – Do not cover the document content. Discuss coordination meeting held with 01D CS analysts.*
- *TIEF Joint COIs – Review planned mapping of Navy MBTD to Joint COIs (if required). Highlight any expected difficulties in translation.*
- *Old TEMP COIs – If the IEF COIs do not match those in an earlier version of the TEMP, this must be covered and approved. A TEMP page change may need to be initiated.*
- *Action Items – Review action items, and who has each for action.*

☐ 22. Disseminate meeting minutes/action items for stakeholder review.

Completed within 2 work days of meeting. Include a list of attendees.

☐ 23. Update stakeholders on action item completion.

Completed within 1 week of meeting.

☐ 24. Close action items; send post-review MBTD products for stakeholder awareness.

This is especially important to informing stakeholder who could not attend the meeting of MBTD changes.

☐ 25. Update iBOSS PM (TP-1 complete).

☐ 26. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

Touchpoint 2 Checklist

Purpose: Developing the subtask hierarchy, defining conditions & tracing conditions to subtasks.

☐ 1. Select and add USN ROC/POE-mission-based COIs from the IEF database standard list.

☐ 2. Refine the effectiveness 1st-level subtasks.

a. Review the 1st-level subtasks provided for each COI. Identify the 1st-level subtasks that are not relevant to the SUT. These will be retained in your hierarchy, but grayed out.

Graying is a mechanism by which MBTD products are listed, but shown as not applicable. 1st-level subtasks cannot be deleted, and thus, are grayed at the base layer when they do not apply. All other graying is done in the IEF node. Graying in IEFs is done to support approval of the choice to exclude MBTD components. Deletion is discouraged because it would bypass such discussion and approval. Consult your CTF before simply deleting items from the IEF node. Gray items are deleted from the test plan nodes.

b. For existing MBTDs, remove non-standard 1st-level subtasks if previously created, and adjust any child tasks as needed.

The 1st-level subtasks in ROC/POE-based COIs were set by OPTEVFOR and cannot be changed for a program MBTD. Changing them would take a policy initiative coordinated with outside stakeholders.

The most commonly added 1st-level subtasks were subtasks from other missions that were not chosen as COIs (C3, MOB, etc.). These concerns can be moved to the 2nd-level, changed to measures, or broken out as COIs.

This policy is based on the need to search standard mission threads for deficiencies by 1st-level subtask. The mission threads must be static.

☐ 3. Decompose effectiveness 1st-level subtasks as needed.

The intent of subtask decomposition is to define the missions in terms of specific user tasks, which reflect the section 1 CONOPS and capabilities (new capabilities, enhancements, and regression at FOT&E). Task decomposition should provide enough detail so that all affected major components of mission accomplishment are accounted for. The OV-5, the CONOPS, and/or the TTPs may be useful resources for this step.

Final report conclusions on SUT effectiveness are written to subtask performance (informed by measures and data requirements). Keep this in mind during subtask hierarchy creation.

a. Review the lower-level subtasks automatically included with the selected mission COIs. Delete any that do not apply. Reword and/or move others if they don't precisely convey the meaning of SUT user tasks.

01B developed these optional subtasks to help Core Teams brainstorm on mission breakdown. Multiple past IEFs were referenced, but these tasks are by no means comprehensive.

b. Review subtask hierarchies of any closely related systems. Ensure commonality as appropriate.

Platform and subsystem (and other similarly related) hierarchies for the same mission should match. Some flexibility to this rule is allowed based on the fact that these related IEFs will likely be worked at different times. New MBTDs are not beholden to what is now understood as poor choices by a prior related SUT.

c. Create additional subtasks to comprehensively break down mission execution. Verify subtask organization is logical versus SUT CONOPS.

MBTD is a systems engineering process. Subtasks form the base structure upon which all other MBTD products are built. The extent of task decomposition is not always intuitively obvious. How far to go is often informed by other steps. Subtasks support identifying/assigning conditions and measures. They should also support creating vignettes (and test events) that will comprehensively exercise mission execution.

MBTD is an iterative process. Any IEF products can be reexamined and rewritten based on what is learned in later steps. Don't feel your subtask hierarchy must be correct/perfect on the first try, and cannot be changed.

Once a subtask is decomposed, the scope of that 'parent' subtask must be fully covered by the next level (multiple child subtasks). Conditions and measures do not apply to the parent. They apply to the child. Thus, the children must encompass the parent's full meaning. Having one child subtask is rarely acceptable, as it is simply re-naming the parent.

Tasks are written for operator actions. However, system tasks are allowed in rare cases when complex system operations and/or logic specifically enable mission completion and cannot be expressed as an operator using the SUT to complete that task.

MBTD is completed to scope test for the SUT. Subtasks are written for operator use of the SUT, not the SoS. SoS–related tasks are rarely allowed, and only added if vital to test design.

d. Verify (with program systems engineers) subtask hierarchy reflects actual system functions.

Actual function of the system is vital to mission execution, and thus is vital to the subtask hierarchy.

CAUTION: Do not state mundane/obvious operator functions, but focus on meaningful events. Avoid decomposing tasks beyond these meaningful events. Recommend proceeding beyond a 3rd- level only by exception. Do not include details of the operating environment that should instead be captured as conditions.

☐ 4. Develop the Suitability task hierarchy.

Suitability subtasks are not required. Reliability and Availability subtasks are not allowed. Decomposing the other suitability COIs follows a similar process to effectiveness, though there are no required 1st-level subtasks.

a. Review the suitability subtasks automatically included with the standard suitability COIs. Delete any that do not apply. Reword others if they don't precisely convey the meaning of SUT user tasks

Preventative and corrective maintenance are the most common suitability tasks. The choice to include these tasks hinges on how robust the O-level maintenance is.

The actions (fault identification, troubleshooting, repair, retest) to complete maintenance (including PMS) belong in maintainability. The actions (part procurement, help-lines, I/D-level repair) and items (documentation, on-board supplies) supporting maintenance belong in logistics.

b. Review subtask hierarchies of any closely related systems. Ensure commonality as appropriate.

c. Create additional subtasks to comprehensively break down user tasks completed to support SUT availability, etc.

☐ 5. Refine the CS subtask hierarchy.

The CS subtasks within each CS COI will match all the 1st-level subtasks for the COIs (including suitability, if chosen) being mirrored. Lower-level subtasks are not recommended.

☐ 6. Review completed task hierarchy.

A second look is always a good idea. Now that you have the trees, re-examine the forest. This is a prudent action throughout MBTD. If applicable, consider in-scope SUT impact on subtasks.

If applicable, consider in-scope SUT impact on subtasks. Remove any that no longer apply within the new test scope.
Subtasks at the base layer must be those for the full SUT. Removal of tasks that do not apply to subsets of the SUT is done in the IEF node.

a. Verify any overlap of tasks between and within effectiveness COIs is appropriate.

Missions can easily share similar preparation, conclusion, or supporting tasks. Just be sure this is deliberate.

b. Verify that no tasks appear in both effectiveness and suitability.

Tasks common to both Effectiveness and Suitability are not allowed.

c. Verify no task overlap between suitability COIs.

☐ 7. Create an initial conditions list.

Conditions define the operating environment; aspects of the real world that affect SUT performance or influence operator actions. They are broken into four categories: physical environment, military environment, civil environment (rarely used), & custom. The first three categories are already populated in the IEF database from the Universal Naval Task List (UNTL). Custom conditions are created by the core team and added to the IEF database.

a. Review SUT reference documents.

i. The CDD may identify conditions that define the SUT's operating envelope.

ii. Review CONOPs.

b. Review conditions directories of any closely related systems. Ensure commonality as appropriate.

c. Review the 01B database of conditions used in previous signed IEFs (category #5).

d. Identify standard physical, military, and civil conditions that apply.

The database conditions directory only includes conditions traced to subtasks, and conditions only apply to lowest level tasks.
It may be simplest to start by brainstorming a list of potential conditions outside the database.

e. Create custom conditions; input in database.

Do not duplicate, or replace standard conditions. Use the standard conditions (vice creating a custom) if possible. There is one standard CS condition, cyber threat. Additional CS conditions can be created like any others.

☐ 8. Complete conditions traceability.

The resulting linkage should identify what things can influence the operator's actions and/or performance of the SUT. Having this traceability supports building a vignette, and subsequent planning associated with design of experiments (DOE).

a. Associate conditions from the initial conditions list with the appropriate lowest level tasks.

The initial conditions list created above will make working traceability go much faster. The standard CS condition traces to all CS tasks.

b. Examine each lowest–level subtask one-by-one.

It is easy to create a quick list of conditions applicable to the SUT. But the core team cannot be certain all conditions have been defined until every subtask is examined, and conditions (plus their associated descriptors) applicable to each subtask are comprehensively traced.

i. Determine if any more conditions need to be traced to fully define the environment that affects performance of that subtask.

Conditions are not required for every subtask. Conditions can be traced to suitability, including Reliability/Availability (very rare).

ii. Create more custom conditions as needed.

CAUTION: Do not attempt to document every conceivable condition affecting the SUT (e.g. sun spots). Using subject matter expertise, identify conditions most likely to impact performance or those of most interest to the testers.

iii. Trace these new conditions to subtasks.

☐ 9. Examine the conditions directory. Verify the output is consistent with expectations following traceability.

If applicable, consider in-scope SUT impact on conditions. Remove any that no longer apply within the new test scope.
Conditions at the base layer must be those for the full SUT. Removal of conditions that do not apply to subsets of the SUT is done in the IEF node.

a. Verify conditions are appropriate for test.

b. Consider SoS conditions. Identify any that are vital for test and trace to the appropriate tasks.

c. Only if appropriate, add conditions unique to test.

Conditions can include aspects of the test environment (ranges, M&S, etc.) that define variability needed for comprehensive testing and act as surrogates for the real-world conditions.

☐ 10. Begin thinking about whether conditions are controllable during test and how the OTD might control them.

Controlled conditions are items that can be set/adjusted as desired by the OTD during an operational test event (e.g. presence of jamming). These controllable conditions will be used to identify different conditional variations required to test the SUT (e.g. day/high alt/EO mode vs night/low alt/IR, etc.). If a condition is not controllable, it is still recordable. Recordable conditions are items that can't be specifically controlled (e.g. sea state) but are critical to capture for post-test analysis purposes. The identification of controllable vs. recordable conditions will impact DOE, test resource requirements and data collection requirements.
Identification of conditions as controllable or recordable in the IEF does not prevent OTDs from modifying or updating their test design at a later date. For example, data collected and analyzed during IT events may drive OTDs to a different conditions list which will be documented in the appropriate test plan. Don't document the choices yet. That's done later. This is just a preliminary evaluation to help define conditions/descriptors.

☐ 11. Establish descriptors (levels) for each condition.

The table below provides examples of three conditions with their descriptors. The first is a continuous condition for which levels do not need to be defined. It is enough to define the operating envelope, or even to just state the units of measure (if envelope is unknown) like the third example. The second is a categorical factor for which two distinct levels exist. Descriptors must match the behavior of a condition, and the intended use during test.

Condition/Factor	Descriptors/Level
Altitude	50-1000 ft
Target	Small (Drone) / Large (Aircraft)
Wind Speed	knots

a. Review the reference material used to identify conditions; now use to establish descriptors.

b. Modify descriptors for standard UNTL conditions.

The standard descriptors in the tool are unlikely to be appropriate to your SUT. Add, remove, and edit as needed.

c. Create descriptors for custom conditions.

d. Verify descriptors match expectations for the condition type (continuous, discrete, categorical).

Do not use "Recordable" in any descriptors. This designation is made for specific designs, not within the conditions directory.

e. Verify the definition of each level is operationally relevant, and clearly bounds the purposes of data collection and post-test analysis for the SUT.

If possible, descriptors should be identified in quantifiable terms vice simply "easy medium/hard". Testers and reviewers need to know how each of those qualitative terms are defined. Recognition of the full variability of the environment does not mean the test must cover all of that.

Adjust descriptors for the in-scope SUT if needed. Descriptors at the base layer must be those for the full SUT. Adjustment (add/delete) of descriptors that do not apply to subsets of the SUT is done in the IEF node.

Descriptors in the conditions directory should almost never perfectly match the levels chosen for the test design. For a categorical aspect, the test design likely only uses a subset of the levels. All the relevant levels should be in the directory. For a continuous aspect, the test design should include specific numbers or intervals that do not span the full operating envelope. None of these levels or intervals should be set as descriptors in the directory.

☐ 12. Verify the conditions and descriptors chosen account for the environment within the priority TACSIT or OPLAN/CONPLAN geographic areas.

MCO 1/2/3 and other mission areas have been studied thoroughly to understand the weather and/or other conditions that will affect mission operations. Test can only be properly planned if the threat area and environmental impacts on the SUT are fully understood.

Reference: Fleet Numerical Meteorology and Oceanography Center (FNMOC) offers a web-based portal to its extensive databases (https://portal.fnmoc.navy.mil/climoportal/index.htm), covering many environmental parameters over the period 1970-2011 for practically everywhere on the earth.

This is the first step in selecting proper test locations/times and understanding any limitations that result from those selections. The checklist continues the process with later direction. OPTEVFOR should have a meteorologist on staff to help with this process.

Ensure that all times of year for the threat area are considered. This data will not specifically be documented in the IEF, though a short summary could be included in the Mission CONOPS portion of the SUT Employment Concept paragraphs.

☐ 13. Draft the IEF Cyber Concept paragraphs; include threats and defense.

Detail CS with the same rigor as a mission, including users, networks, threats, etc. (cover the "why" of CS for your program). This includes explaining that there are no CS concerns for your SUT, if applicable. For FOT&E (and the like), focus CS on the system deltas that may introduce new operational concerns regarding insider, nearsider, and outsider threats without re-testing items that haven't changed.

This is not a summary of cyber testing. That appears in IEF sections 2 and 3, and is completed in the DWG and IPR-2 checklist.

☐ 14. Complete the CS column of the MCSM & MCSFM.

Identifying the SUT components that if successfully attacked, will result in critical negative mission impacts.

☐ 15. OTD, LTE, and CTF review all Touchpoint-2 products and discuss (ideally agree on) readiness for leadership review.

☐ 16. Request stakeholder feedback on products.

Ideally, continuous core team collaboration should make this step redundant. However, a final/formal request should go out at least 2 work days before the review.

Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

☐ 17. Adjudicate/incorporate stakeholder feedback.

If there isn't time or agreement to incorporate the comments, ensure the items are ready to be discussed at the meeting.

☐ 18. Schedule TP-2 and provide read-aheads.

Deliver the official read-aheads to all stakeholders at least 2 work days prior to the meeting. More than one week of lead time is preferred.

The greater the scheduling lead-time, the easier it is to deconflict calendars. In-person attendance requires more warning. Send the invite at least 1 (but not more than 6) weeks ahead. The meeting can be scheduled before stakeholder feedback is returned.

Products to provide (exported from appropriate MBTES layer, likely IEF node):
- *IEF Section 1 updated with cyber content.*
- *Table A-1 (Conditions Directory).*
- *Table A-3 (Traceability Matrix) – Measures columns are removed.*
- *Stakeholder feedback.*
- *CRM, if used.*

Personnel to attend:
- *Division – ACOS or DACOS, Section Head, OTD, Contract Support.*
- *Support – 01B or 01B1, CTF, LTE, 01D rep.*
- *Outside agencies – TP-2 has more T&E implications than TP-1. Stakeholder attendance (including OPNAV N942) is becoming more important. TP-2 may just require emails. Dial-in may be appropriate, especially for other OTAs. In-person is acceptable.*

Meeting time:
- *Large systems may take 1-3 hours. The smallest systems may take 30 minutes (schedule a minimum of 1 hour).*

☐ 19. Conduct Touchpoint-2.

Begin the meeting by setting the classification level and noting if electronic communication means (e.g., telecon) are used. Then do quick introductions. Then summarize how the meeting will progress, and the goal at completion. Encourage attendees to bring up any unresolved comments as that section is reached in the review, and state that the test team will be doing that for anyone not in attendance.

After looking at section 1 for changes and new material, display table A-3 on the screen. Put table A-1 and A-3 side-by-side on the table. Have a folder of reference documents ready to be accessed electronically. Don't read any of the products. Talk to them, and allow for questions.

Meeting key concerns and progression:
- *Cybersecurity concept and CS COIs – The cyber threats and defenses must be understood to support future cyber test planning. Compare the COIs to those of effectiveness and suitability.*
- *Task Hierarchy – Proceed COI-by-COI confirming adherence to, and full coverage of, the CONOPS/TTPs for each specific mission and suitability concern. Show that CS matches. Do not discuss task criticality.*
- *Conditions Traceability and Conditions Directory – With mission execution outlined, proceed task-by-task reviewing the linked conditions. Ensure that each subtask has the right ones. As each condition is covered, review the descriptors.*
- *Action Items – Review action items, and who has each for action.*
- *Checksum – State whether any checksums relevant to TP-2 show errors or warnings. Be prepared to explain any not cleared.*

☐ 20. Disseminate meeting minutes/action items for stakeholder review.

Completed within 2 work days of meeting. Include a list of attendees.

☐ 21. Update stakeholders on action item completion.

Completed within 2 work days of meeting.

☐ 22. Close action items; send post-review MBTD products for stakeholder awareness.

This is especially important to informing stakeholder who could not attend the meeting of MBTD changes.

☐ 23. Update iBOSS PM (TP-2 complete).

☐ 24. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

IPR-1 Checklist

Purpose: Building the measures matrix, linking measures to the task hierarchy, creating data requirements for measures and conditions, and identifying critical tasks/measures.

☐ 1. Review the sponsor-approved requirements document (CDD/CPD/ORD, etc.).

For purposes of this checklist, the requirements document will be referred to as the CDD.

☐ 2. Highlight all operationally relevant requirements in the CDD for potential inclusion as specified measures.

Specified measures are those measures directly translated from requirements clearly identified in the sponsor-approved document. Not all programs will have such a document. It is possible you may have no specified measures (rare). Other programs have sponsor-approved requirements written up in non-JCIDS documents. Special cases like these should be explained in the IEF.
This is just parsing the CDD. What to do with those requirements is a decision for later.

a. Identify KPPs, KSAs, APAs, etc. for possible inclusion as measures.

All KPPs must be measures in the IEF, even if they are not relevant to OT (and thus the measure is orphaned).
KPPs and KSAs have been called out in the CDD by the sponsor as important. Exclusion from use as OT measures is a significant decision. By including requirements as measures, and then orphaning those measures, we clearly communicate our intention to not report on the requirement.

b. Identify "will/shall/must" statements for possible inclusion as measures.

Not every one of these statements constitutes a measure. The OTD must decide if each statement reflects a capability to be evaluated during OT that the system delivers in support of mission execution.

☐ 3. Document the specified measures in the IEF Database.

Measures are the specific metrics used to assess performance of capabilities the SUT is supposed to provide. Not all requirements are measures. Not all measures are requirements.
Quantitative measures are preferred to qualitative measures, but both will

(continued)
be necessary for OT. For all measures, the core team must be able to develop a clear path to analyzing/resolving the measure. If this is not possible, it is not a measure.

a. Add the measure.

Measures are statements (not questions) or simple metric names (e.g. detection range). Specified measures should be entered as close as possible to verbatim from the CDD (especially KPPs). Some variation from CDD wording is allowed if necessary to make the measure fit our format and best express how we intend to test and resolve it. Too much variation and it is no longer specified (it would be type other).
Do not write measures that imply a specific analysis. For example, a measure of "mean detection range" implies that the distribution of detection ranges is approximately normal, and the mean of the distribution is the best summary statistic. Often this is not the case. It is better to write the measure as "detection range". Then the analysis plan can be more flexible, and post-test examination of the distribution will indicate the best way to convey the result. If "mean detection range" is a specified requirement, the measure is written "detection range", and the criterion is written "mean ≥ x".

b. Write the criterion.

Criteria is the performance level that must be achieved. For specified measures, these are the thresholds set by the sponsor. Some CDD requirements are objective-only. These can be added as specified measures, but with criterion of "No Threshold".
Ensure that the answer (criterion) matches the question (measure). Examples: a "Probability of kill" measure has a criterion of "≥0.XX"; a measure stated "The system kills the threat" has a criterion of "Yes". Obviously, P_k does not get the "Yes" criterion.
All quantitative criteria should be accompanied by inequalities (e.g. an A_O threshold of 0.80 yields a criterion input of "≥0.80").

c. Categorize the measure as MOE, MOS, MOCS, or SOS.

SUT measures only fit in one category amongst effectiveness (MOE), suitability (MOS), or CS (MOCS). If you have a measure you think applies to multiple categories, you actually have multiple measures. Document them as such.
MOEs, MOSs, and MOCSs apply to the SUT. A SOS measure is any capability or issue not previously captured that is needed for SUT mission accomplishment from a system-of-systems perspective (e.g. when the SUT is a weapon system that relies on the accurate input of a radar track passed to it by a radar outside the SUT. While the SUT may perform effectively on its own, when the accuracy of the track provided to it is taken into account, the overall SOS may not perform effectively). Determining what metrics should be categorized as SOS depends on a well-defined SUT & SOS.
Specified MOCSs are possible. When any specified MOCS is used, determine

(continued)
whether that measure replaces any of the six standard CS measures that can be included.

M#	Measure	Criterion	Source	Para	Type
M56	Flank speed endurance	28 kt for 12 hr	CDD	Table 5.1	Specified
M57	Transit speed endurance	18 kt for 18 hr	CDD	Table 5.1	Specified
M58	Reverse thrust possible	Yes	CDD	Table 5.2	Specified
M59	Maximum reverse speed	5 kt	FRD	3.1.2	Derived
M60	Stopping distance from flank using reverse thrust	No Threshold	OTA Create		Other

d. Set the measure type to Specified.

The measure type establishes basic groupings of where each measure comes from. Specified measures only come from the Sponsor.

e. Write the source and paragraph.

Must be a sponsor-approved document.

f. Mark unique aspects of the measure:

i. KPP

ii. DT Only

This is best done later, when identifying DRs for measures. DT only measures are those that remain relevant to OT, but rely on data collected solely during DT. Often, OT does not have the ability or expertise to test these, and relies on DT conclusions.

iii. Needs Clarification

OT should resolve specified measures as they were intended to be resolved. Not all specified measures are clear. If clarification by an outside agency is needed, this marking notifies readers that the measure is intentionally incomplete/unclear (for the moment), and reminds the OTD to continue pursuing the information needed.

iv. Gray

Graying is an important MBTD concept already discussed briefly on tasks. To gray an item is to say it is not relevant to this MBTD. Perhaps it will be in the future, or was in the past. If you want to keep it, but not test it this time, then gray it out. Measures parsed from a current CDD that apply to a future increment are often grayed. Graying only applies in IEFs, not test plans, etc.

v. Orphaned, including why it is orphaned.

Orphan measures are not relevant to OT. We will not look at them as part of our effectiveness/suitability assessment. Only specified measures are marked as orphaned (usually just KPPs). We keep them in the IEF to show full parsing of the CDD. We mark them orphaned to show we don't care about the requirement. Mark them now if it's obvious, but this can also be done later when tracing measures to tasks.

Be careful when marking a measure orphaned because it is "untestable". Doing so means that the measure truly cannot be tested. Consider a detection range requirement for a specific threat that cannot be tested when we lack the proper surrogate for that threat. The measure would be testable if we were not limited by resources. Thus, orphaning this measure would be wrong. Gray-out the measure if it is deferred to future testing. Or keep the measure, but associate a limitation. Don't orphan relevant metrics.

☐ 4. Review the specified measures.

OTDs should scrub the resulting measure matrix for duplicate or unnecessary measures, and ensure the matrix is as expected.

If applicable, consider in-scope SUT impact on specified measures. Remove any that no longer apply within the new test scope.

☐ 5. Trace specified measures to subtasks.

OTDs should approach this using the question "What measures do I need to evaluate the ability of the SUT to perform this task?" and then select measures that answer that question from the measure matrix. If the measure does not help answer that question, don't link it to that task.

a. Trace MOEs to lowest-level effectiveness subtasks.

b. Trace MOSs to suitability COIs and/or lowest-level subtasks.

c. Trace MOCSs to CS COIs and/or lowest-level subtasks.

d. Trace SOS measures like MOEs or MOSs.

Each SOS measure either applies to SOS effectiveness, or SOS suitability, not both (just like with SUT).

☐ 6. Review specified measures traceability.

This task is iterated several times during the identification of measures. Reviewing traceability after linking just specified measures allows the core team to know how far they need to go in developing Derived measures and perhaps creating Other measures. Plus, going over traceability several times is always prudent. However, reviewing the comprehensive nature of traceability can be delayed until after Derived and Other measures are thought to be mostly done. It is a style choice for the core team.

a. Identify measures that did not link to subtasks.

i. If they are relevant to OT, go back and link them to the appropriate task. If the appropriate task does not exist, add tasks as needed.

ii. Orphan measures that do not trace.

b. Examine each lowest–level subtask (or COI) one-by-one to ensure task (or COI) success is fully defined by the associated measures.

The focus of OT is task execution and success. The purpose of measures is to define (qualitatively and quantitatively) that success. If the measures traced to a task (or COI) do not accomplish this, there is a hole in the traceability matrix that must be filled. Derived and other measures will be used to fill such holes.

☐ 7. Review the other applicable reference documents. Extract measures from those alternate sources.

<u>Derived</u> measures are not explicitly stated in the CDD, but are stated clearly in other SUT source documents (such as: Fleet TTPs, applicable CONOPS, system specifications, platform level ROC/POE, military standards, CDD references, subsystem CDDs (for platforms), OPNAV instructions (like the Availability Handbook), a UONS, etc.).

When a CDD states "SUT will meet the requirements of document XYZ", measures from that source are still considered derived.

☐ 8. Document the derived measures in the IEF Database.

a. Add the measure.

b. Write the criterion.

For now, input the quantitative or qualitative value provided by the source. Later, if the sponsor does not agree the criterion is a relevant threshold for system capability, this will be replaced with "No Threshold". This approval is usually requested after IPR-1 (covered by a later step), when a more refined set of measures is ready for stakeholder agreement.

c. Mark the measure as MOE, MOS, MOCS, or SOS.

d. Set the measure type to Derived.

e. Write the source and paragraph.

f. Mark unique aspects of the measure:

i. DT Only

ii. Needs Clarification

iii. Gray

☐ 9. Review the specified and derived measures.

Look for duplicate or conflicting measures. There is no guarantee the CDD will agree with all other SUT documents on measures.

If applicable, consider in-scope SUT impact on derived measures. Remove any that no longer apply within the new test scope.

☐ 10. Trace derived measures to subtasks.

a. Trace MOEs to lowest-level effectiveness subtasks.

b. Trace MOSs to suitability COIs and lowest-level subtasks.

c. Trace MOCSs to CS COIs and/or lowest-level subtasks.

d. Traced SOS measures like MOEs or MOSs.

☐ 11. Review specified and derived measures traceability.

a. Identify measures that did not link to tasks.

i. If they are relevant to OT, go back and link them to the appropriate task. If the appropriate task does not exist, add tasks as needed.

ii. Delete measures that do not trace.

b. Examine each lowest–level subtask one-by-one to ensure task success is fully defined by the associated measures.

Tracing derived measures to subtasks should fill some of the holes in the traceability matrix left by specified measures. But there is not guarantee holes do not still exist.

☐ 12. Create other measures as needed.

<u>Other</u> measures do not have a SUT source document (non-SUT documents, like the CDD of a similar system, yield other measures). Every effort should be made to find SUT documented measures. When they cannot be found, the OTA must create measures. This is not the creation of requirements, only measures for test.

a. Add the measure.

Suitability metrics from the Suitability Handbook or the MOA on suitability are excellent OTA Created measures. The standard CS measures are all OTA Created.

CDD metrics rewritten for OT applicability (poor CDD wording), and deviate greatly from CDD wording are also type Other.

b. Write the criterion.

Because these measures cannot be system requirements, OTA Created measures have criterion of "No Threshold" unless a threshold value is added by the sponsor.
This restriction applies to both quantitative and qualitative criterion, as both types should create a performance burden to be met. If a measure (through the application of any criterion) cannot be interpreted as setting a performance standard for the system, it is not a measure.

c. Mark the measure as MOE, MOS, MOCS, or SOS.

d. Set the measure type to Other.

e. Write the source as OTA Created and leave the paragraph blank.

This is not always the case. The Suitability Handbook and similar sources are understood to be OTA Created, so document that specific source and fill in a paragraph. Stating "OTA Created from…" and adding a source is also allowed if the inspiration for the Other measure needs to be communicated (a paragraph can be added in this case). When the genesis of the measure is bad wording in the CDD, the source is "OTA Created from CDD".

f. Mark unique aspects of the measure:

i. DT Only

ii. Gray

☐ 13. Review the full measures matrix.

Look for duplicate or conflicting measures. OTA Created measures should not replace specified/derived measures.
If applicable, consider in-scope SUT impact on other measures. Remove any that no longer apply within the new test scope.
Measures at the base layer must be those for the full SUT. Removal of measures that do not apply to subsets of the SUT is done in the IEF node.

☐ 14. Trace other measures to subtask/COI for which they were created.

Subtasks can have many associated specified/derived measures, and still need OTA Created measures to cover the full meaning of success for that task.

☐ 15. Review the complete traceability matrix.

a. Verify all measures properly linked to tasks.

b. Verify all tasks/COIs fully covered by measures.

c. Ensure, through traceability, that all measure are in-fact MOEs or MOSs.

d. Verify different subtasks do not share all of the same measures.

Subtasks are written differently for a reason. They cannot have the exact same measures. If they are the same subtask (just found in separate missions) they can have the exact same measures.

e. Look for measures that trace in too many places.

Measures can apply to many tasks, but applying in too many places can complicate data collection and/or analysis. When this happens, the measures may need to be more specific, or traceability may need to be refined. Training measures, for example, apply everywhere. But tracing them to the tasks that depend most on having proper training is the best approach.

i. Edit measure and add new measures to be more specific.

Sometimes broad measures are used with the intention of being more specific in the DRs. This is a dangerous approach. Broad measures can trace to very different tasks. Testing may focus on just one of the tasks. But the vignette pulls in these broad measures with many DRs that do not apply. Now all the DRs that do not apply must be hidden, one-by-one, vignette-by-vignette. It is much simpler to pay the price up front and make specific measures.

ii. Adjust traceability of edited measure and new measures created for specificity.

f. Verify that graying of measures and tasks is consistent and correct.

An active (not gray) task needs at least one active measure. An active measure must be traced to at least one active task.

g. Request clarification of measures as needed.

☐ 16. Identify Data Requirements (DR) for each measure.

DRs consist of three primary items: (1) Element (i.e. temperature, position, time, opinion, screen shot), (2) Unit of measure (e.g. °F, lat/long in degrees, min:sec), & (3) Data source (i.e. the specific mission computer log, operator survey, OTD observation). Each new combination of element, UoM, and source forms a new DR.

***Good** DRs **state** exactly what data needs to be collected and how that data will be available/presented during the test event and collected by testers. Suggested suitability DRs are presented in the Suitability **Handbook**.*

Do not write the obvious. Create DRs that guide those at test to bring back the specific data needed (e.g., "OTD Notes" is not a data element; of course, the OTD will take notes).

DRs must encompass all the info needed for evaluating the measure. For example, the suitability handbook includes more DRs for MTBOMF = Operating Time / # of OMFs than those that directly translate to the numerator and denominator. Data is collected for context; in this case, to support determination of which faults/failures are OMFs.

DRs needed for full measure resolution are not influenced by whether that data can be gathered at test. We include measures that might never actually be tested (due to real-world restrictions) and then we write limitations acknowledging the measure won't be tested. We can do the same for DRs that we need but will never actually get. In fact, such DRs must be included to justify existence of associated limitations.

a. Select from existing DRs, if appropriate.

A single DR can apply to many measures. Duplicate DRs are not allowed, so using existing DRs is often necessary. If the DR you need already exists in the database, simply link it to the measure.

b. Define Data Element.

Be specific to the exact piece of data that needs to be collected. "System operating time" is not a data element. Nor is "number of OMFs". Neither of those items can be written down during test. Operating time comes from recording when the system is turned on, goes down, is restored, put in standby, etc. Number of OMFs comes from post-test analysis where all qualitative failure data is looked at through scoring boards and OMFs are assigned/counted. Individual failure times are an example of 'instance DRs'.

On the other hand, operating time could be a data element if that exact number can be pull off an automated log. If the log contains the individual times (not the total), those are still the DRs.

i. Create a new element, or use an existing one.

New data will need a new data element. A data element will have only one applicable UoM (it should only be quantified one way), but may have multiple sources. Thus, the data element you need may already exist in another DR. You can't pick it from a list, but make sure you use the exact same element wording for the same data.

ii. Include data frequency, if needed.

Some data must be collected at a periodicity (e.g. every 30sec). For example: to understand ship's track, position is collected at a periodicity appropriate to the track DR, to the measure, and to the SUT (missiles likely need higher frequency than ships).

c. Define the Unit of Measure (UoM).

DRs have one UoM. If more than one UoM is listed, you may have multiple DRs. Break them up as appropriate. It is acceptable to use "Various" if the Data Element does in-fact have several UoMs.

The UoM for qualitative DRs is often "Qualitative". The UoM for survey DRs is often "5-point Likert Scale", or similar.

i. Select from list or create a new UoM.

The database keeps a list of UoMs already used. Using this list is a powerful aid in tracking and refining UoMs.

ii. Include the data accuracy, if needed.

Some data collection has an acceptable tolerance (e.g. ±1m). When the data collector gathers the data element, how accurate do they need to be? For example: an MCMTOMF of 5min needs times collected to the second, not to the minute.

All stochastic results have multiple sources of noise. One source we cannot avoid is natural variability in the population. Sources we must minimize/avoid are types of error. For data accuracy, we are referring to measurement error. If data is measured inaccurately, this adds to noise and negatively impacts analysis.

d. Select Source from list or create a new Source.

Each DR has one source. If more than one source is listed within a DR, you actually have multiple DRs. Break them up as appropriate.

For each DR, the IEF database also has optional fields of 'where recorded', 'collector position', and 'other items needed'. These are used for the test plan, not the IEF. Thus, the data source is not any of these options (data sheets and OTD log are not sources).

Source is exactly that, where the data originates. This can be an OTD observation, a system display, an automated log, an operator interview, a DT report, etc. Use SME knowledge and be specific.

Surveys are governed by rules and best practices. Thus, DRs listing a survey as their source must follow those guidelines. 01C should be consulted to aid in constructing surveys for specific purposes. The survey DRs must match that survey.

"Post-test analysis" is not a source (it's not collected at test).

i. Verify the data source will provide the data element at the required periodicity and accuracy.

If it won't, the DR is invalid and you need a different source.

ii. Verify the source has the required precision.

The data can be recorded from the source very accurately, but it is only precise if the source is giving the correct value. This is instrument error, another source of noise that negatively impacts analysis and we must minimize/avoid. Imprecise sources invalidate DRs.

For example, a GPS might provide ship's positions within +/-10ft. But where is the GPS located in relation to the true point of interest (the ship's centroid)? Separation between these two points may invalidate the source. Listing a source means you understand that source fully, and are certain it will yield the required data.

A simpler example of lacking required precision is recording times from a clock that is wrong. But that has a simple fix prior to test. This verification focuses on precision problems that cannot be fixed or need major fixes.

☐ 17. Review the DRs for each measure.

DRs are the most extensive and often the most time-consuming part of MBTD. But they are also the most important. Execute this process with the same rigor as ensuring the measures traced to each task are appropriate for that task.

a. Ensure DRs for each measure are comprehensive.

Test is data collection. Analysis relies on data. The DRs for each measure must represent everything that will be needed to analyze the result post-test. They must also not be excessive (don't link DRs that are not needed to resolve that measure).

Anticipating some DRs being unavailable at test is not a reason to omit those DRs. They are required data. Write the DRs. Then later, write the limitation acknowledging we won't get those DRs.

DRs at the base layer must be those for the full SUT. Removal of DRs that do not apply to subsets of the SUT or to this test phase is done in the IEF node.

b. Ensure the DRs for each measure are appropriate.

DT Only measures have DT DRs only. But DT DRs can apply to any measure. Survey DRs won't align with some measures. Look at the DRs again to verify they can actually be collected, and are useful for OT resolution of that measure specific measure.

c. Consider the need for data redundancy.

What happens if data from a source is garbled, and you only have one source? It may be prudent (especially for critical measures) to collect data from backup sources. These redundant sources might involve different collection methods (manual, vice automatic).

☐ 18. Identify the DRs for each condition.

DRs for measures and conditions should be separate. The conditions under which tasks are completed are also those under which measures are tested. By writing DRs for conditions, the conditional aspects of measures are covered. Don't link measures DRs to conditions, don't link conditions DRs to measures, and don't duplicate.

a. Define the Data Element.

It may be enough to simply repeat the condition name. It may not.

For example, Ocean Acoustics is a common condition in ASW. But a DR of that name is so broad as to be meaningless. The more specific aspects of ocean acoustics should either be other conditions in the MBTD with their own specific DRs, or that specific data needs to be written for this condition. Post-test analysis (especially in DOE) often relies on recording this precise data.

Your test design may have light levels of day/night, but perhaps the precise data of local lumens will be more meaningful.

b. Define the UoM.

For categorical DRs, "Nominal" is a good choice. It means you are picking a name out of a list (your list of descriptors).

Don't forget to include the required data accuracy, if applicable.

c. Define the Source.

"Run plan" is a good source for some controlled conditions. But listing this as a source before DOE is completed has some risk. It indicates intention to control the condition. It may also need to be changed after writing the test design.

Don't forget to verify acceptable source precision, if applicable.

☐ 19. Review the DRs for each condition.

Like those for measures, the DRs collected for conditions are vital to post-test analysis. If any are not vital, delete them. If you don't need the data, perhaps you should simply delete the condition.

a. Ensure DRs for each condition are comprehensive.

Conditions data may require multiple sources. Conditions data may cover both what is planned, and what is actually executed.

As with measures, DRs at the base layer must be those for the full SUT.

b. Ensure the DRs for each condition are appropriate.

OTD observations are rarely accurate for weather data. Is there a NOAA buoy? What about a METOC report? Get the best data.

☐ 20. If advantageous to the MBTD, define the categories of DRs.

It is best to define categories based on how data will be collected during test, (keeping like-DRs together). Common categories are automated data, manual data, conditional data, and DT data. But many others can be created. If many DRs come from a source, that can be a category. If many DRs will be recorded in one place (e.g. OTD Log or Data Sheet X), that can be a category.

☐ 21. If advantageous to the MBTD, categorize DRs.

DRs will only populate in the vignette DRTM (not required to be created) if they are categorized. Start by looking at the DR's source. Every DR from a common source should belong in the same category, though there may be exceptions here. Conversely, not every DR in a category is expected to come from the same source.

☐ 22. If necessary, review DR categorization.

a. Use all categories you create. Delete any extra.

A few of the categories in the tool are standard (because they should almost always be used), and cannot be deleted.

b. If used, ensure all DRs are categorized.

c. Ensure DRs fit the category to which they are assigned, especially DT DRs.

Categorizing a DR as DT Data indicates that DT will collect that data. It also means that OT does not need to qualify that data for use in OT purposes. Thus, this categorization needs careful thought.

☐ 23. Identify the critical tasks.

Critical tasks are those tasks essential to mission accomplishment, and thus drive COI resolution. If the operator is not able to successfully accomplish the critical task, this could potentially result in finding a COI to be UNSAT, and/or the SUT not effective or suitable.

a. Select at least one critical task per COI.

Only lowest-level subtasks can be critical; parent subtasks are not critical, except through the associated child subtasks. COIs without tasks are critical. There is no need to identify them as such in the IEF database or document.
Try not to select too many subtasks as critical. Non-critical subtasks are still important to mission completion, and can still have major impacts on COI resolution.

b. Ensure the critical tasks are consistent with both the In-scope SUT and the current CONOPS.

Critical tasks can change as the system is upgraded. Proving upgraded performance and doing regression confirmation is often the focus when testing after IOT&E and the SUT is broken down between In-scope and Out-of-scope. Consider what is critical to that.

☐ 24. Identify the critical measures.

a. Select at least one critical measure for each critical subtask (or COI if there are no critical subtasks).

Candidates include KPPs (if operationally relevant), significant MOEs, MOSs, other measures deemed important to OT, etc. Critical measures define success of critical tasks (failure of the measure likely constitutes failure of the task).
Try not to select too many measures as critical. Non-critical measures are still tested/reported, can still require resourcing, and can still have major impacts on task completion and COI resolution.
DT Only measures can be critical; so can SoS measures.

b. Ensure the critical measures are consistent with the In-scope SUT.

For FOT&E, critical measures define success of SUT upgraded capability (new or enhanced) or regression performance.

☐ 25. Review the "critical tasks to critical measures" tables for every COI. Compare against the KPPs, MOEs, & MOSs in the CDD, previous TEMP, etc. and verify that all critical measures are accounted for.

Requirements documents, TEMPs, past IEFs, and past test plans are not perfect. There may be KPPs, MOEs, or MOSs that are not of critical value in determining effectiveness and suitability of the SUT. What was critical in past tests may not be now.

☐ 26. For TIEFs ending work at IPR-1, draft IEF section 2.

The TIEF template directs the minimum content needed. Statistical design or other test scoping language cannot be added in this case, as that would mean the MBTD is proceeding to a DWG.
IPR-1 is the final review meeting for these TIEFs. Like IPR-2 for full IEFs, the document needs to be essentially ready for routing. See the IEF routing checklist after IPR-2.

☐ 27. Review critical measures again, this time in the context of DOE.

The critical measures (some anyway) will be those upon which statistical test will be constructed, and thus how testing will be sized. It is important to consult your CTF as you consider DOE.

a. Consider the nature of each critical measure.

Is the measure deterministic or stochastic?
- ***"Deterministic" implies that system performance, as measured, is nearly 100-percent predictable, understood, repeatable, and essentially non-variable over multiple measurement trials. These are also called "diagnostic measures."***
- ***"Stochastic" implies that measurements of SUT performance vary as a function of known and unknown conditions, and are not predictable, understood, and repeatable. Measurements are expected to vary from one measurement trial to the next. These are also called "response variables." Note: here at OPTEVFOR the term response variable (RV) is only applied to stochastic measures for which DOE is completed to examine factor effects.***

b. Consider whether the chosen measures are appropriate for statistical design.

c. Change critical measures as needed.

☐ 28. Update IEF section 1 SUT capabilities description with annotation of how measures match up to the key capabilities being delivered.

Every effectiveness critical measure must be associated with a capability listed in the SUT description. Not every capability listed in the SUT description needs a critical measure. Every capability needs at least one measure. Do not list all measures.

☐ 29. Run IEF database checksum. Fix all issues.

The checksum report includes many mistakes commonly made in MBTD. It can help you catch them throughout the process, not just leading up to a meeting. Not everything flagged by the checksum is guaranteed to be wrong. Consult your CTF.

☐ 30. Ensure measures are in numerical order before IPR-1.

This is a must. Review of all IPR-1 products is made far more difficult if these are not in numerical order. Having DRs in order by appearance in measures is in no way required.

☐ 31. OTD, LTE, and CTF review all IPR-1 products and discuss (ideally agree on) readiness for leadership/stakeholder review.

☐ 32. Request stakeholder feedback on products.

Sponsor approval of non-specified criteria (covered by a later step) can be worked here, even though the expectation is to seek approval after IPR-1.

Ideally, continuous core team collaboration should make this step redundant. Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

Deliver formal read-aheads to all stakeholders at least 2 weeks prior to the meeting. DOT&E has agreed (per the RCRM policy) this lead-time is the minimum they need to provide feedback and to participate in the review with final decisions authority on the material. DOT&E AOs can sometimes support meetings with shorter notice, but are under no obligation to have final feedback sent in prior to the meeting.

☐ 33. Adjudicate/incorporate stakeholder feedback.

Significant disagreements with outside agencies are a vital discussion point. If these exist, it is prudent to pre-brief leadership on this before conducting the meeting.
If there isn't time or agreement to incorporate the comments, ensure the items are ready to be discussed at the meeting.

☐ 34. Schedule IPR-1 and provide read-aheads.

The greater the scheduling lead-time, the easier it is to deconflict calendars. In-person attendance requires more warning. Send the invite at least 2 (but not more than 6) weeks ahead. The meeting can be scheduled before stakeholder feedback is returned.

Products to provide (exported from appropriate MBTES layer, likely IEF node):
- ***All TP-1/2 products (2 hard copies available for reference, 3 for oversight).***
- ***Table A-2 (Measure Matrix).***
- ***Table A-3 (Traceability Matrix).***
- ***Table B-1 (Measures-to-Data Requirements).***
- ***Table B-2 (Conditions-to-Data Requirements).***
- ***Section 2 (only for TIEFs ending review at IPR-1).***
- ***Stakeholder feedback.***
- ***CRM, if used.***

Personnel to attend:
- ***Division – ACOS or DACOS, Section Head, OTD, VX representation, Contract support.***
- ***Support – 01B or 01B1, CTF, LTE, 01C Rep, VX CNA rep.***
  Outside agencies – Dial-in is acceptable, but face-to-face is more appropriate. DOT&E is a must for oversight programs. PM, program T&E lead, DT rep (if program has an associated independent tester), resource sponsor, Fleet user community, associated NWC, and OPNAV T&E rep (N942) are optional; make sure to inform them it is happening and send them read-aheads. Invite SME/analyst support (NUWC/CORONA/etc.) if involved. For Joint programs, coordinate with the other OTAs to consult the equivalent personnel in the other services (Joint PMs, sponsors, user reps, etc.).

(continued)
Meeting time:
- *Large systems may take 4+hr. The smallest systems may take 1hr (schedule a minimum of 2 hours).*

☐ 35. For TIEFs ending at IPR-1, review the Platform Mission Tasks (PMT) View Shell (PV-0). Verify formatting is correct per the PMT View Annex to this checklist.

PV-0 is an automated output of MBTES, and is embedded in TIEF appendix A. The PV-0 is uncolored, but it provides the format to be used to create the other views as the test program progresses. Thus, the PV-0 is a baseline depiction of the OT requirements for a given SUT. It graphically displays the MBTD-developed test design.

CAUTION: The PV-0 may be handled on NIPR and/or SIPR. All other PMT Views contain test results, and should normally be handled only on SIPR.

☐ 36. Conduct IPR-1.

Begin the meeting by setting the classification level and noting if electronic communication means are used. Then do quick introductions. Then summarize how the meeting will progress, and the goal at completion. Include how the RCRM works. Encourage attendees to bring up any unresolved comments as that section is reached in the review, and state that the test team will be doing that for anyone not in attendance.

After reviewing any changes to TP-1/2 material, display table A-3 on the screen. Put table A-3 and B-1 side-by-side on the table. Underneath those, have Table A-2 and B-2 ready. Have a folder of reference documents ready to be accessed electronically. Don't read any of the products. Talk to them, and allow for questions. Begin by stating if PM feedback on DT data and/or Sponsor feedback on non-specified criterion has been received.

Meeting steps and key concerns:
- *Prior products—Move quickly through all MBTD products previously discussed at TP-1 and TP-2 (they're already approved). Cover comments.*
- *Traceability Matrix, Measures, and Measures DRs – With mission execution still understood from TP-1, proceed task-by-task (using Table A-3) reviewing the linked measures (in Table B-1). Ensure that each subtask has the right ones.*
  - *Each measure will be read/evaluated in the context of the current subtask. For derived or OTA created measures with sponsor-approved criteria, provide the approval evidence. If the sponsor rejected any criteria point that out with the associated measure.*
    DRs for the measures will be quickly reviewed in tandem with each measure. This will definitely be the case for critical measures. If the program office has rejected collection of any DT data, point that out.
  - *Critical task and measure designations are able to be discussed throughout this process.*
  - *With all measures confirmed to fit (or de-traced), reexamine the subtask holistically to ensure the traced measures define the full scope of task performance.*
- *Orphaned measures – Discuss.*
- *Conditions DRs – Discuss. Do not discuss intended test location comparison to expected OPAREAs. That comes in IPR-2.*
- *Checksum – State whether any checksums relevant to IPR-1 show errors or warnings. Be prepared to explain any not cleared.*
- *Action Items – Review action items, and who has each for action.*
- *RCRM comments – Establish whether any RCRM comments have come out of the meeting.*

☐ 37. Disseminate meeting minutes/action items and draft RCRM for stakeholder review.

Completed within 2 work days of meeting. Include a list of attendees.
Follow the RCRM policies from earlier in this checklist to clarify comments, gather inputs, close the RCRM to further inputs, adjudicate any remain comment, and elevate the issues as needed.

☐ 38. Update stakeholders on action item completion.

Completed within 1 week of meeting.

☐ 39. Request (if not already done) written confirmation from the resource sponsor that criterion associated with derived measures are acceptable for SUT evaluation.

This is completed via ACOS-to-Sponsor O6-level email. If the sponsor does not agree, measures will be retained and still used for test, but will have "No Threshold" as the criteria. Sponsor feedback will be briefed at the E-IPR.
When a non-specified criterion is approved by the Sponsor, the sponsor is essentially establishing a new requirement by agreeing the measure must perform up to a certain standard. That standard can now be considered a threshold. The measure language can be read as a derived requirement.

☐ 40. Request (if not already done) program office agreement to collect all DT DRs (for measures and conditions).

Completed via ACOS email to PM. This feedback ensures the program office is fully aware of what we need, and will collect the data we need (under the associated conditions).
- ***DT should already be doing this data collection for their test. It should***

(continued)
rarely present any additional burden for the program.

- *If DT will not collect this data, OT must resource for, and collect it.*
- *PM feedback should confirm DT will report on all DT Only measures.*
- *PM feedback should also confirm all DRs gathered from DT sources will be collected. This includes DT DRs linked to measures which are not DT Only and DRs linked to conditions that apply to DT events.*
- *This feedback will be briefed at E-IPR.*
- *Talk to your CTF about how best to create products from the database to support this effort.*

☐ 41. Close action items; send post-review MBTD products for stakeholder awareness.

This is especially important to informing stakeholder who could not attend the meeting of MBTD changes.

☐ 42. Update iBOSS PM (IPR-1 complete).

☐ 43. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

DWG Checklist

Purpose: Selecting response variables (RV), conditions/factors, developing the statistical design needed to generate a run matrix and determine the appropriate sample size to achieve a satisfactory statistical power associated with the factors.

Statistical Design of Experiments varies from one program to the next. One size does not fit all. OTDs need to work closely with their divisional analysts and 01B to arrive at a statistical design that is defendable and useful. The following checklist touches on the basics – each program will be unique. Communication amongst all parties, including with DOT&E and IDA reps is critical.

Work with DOT&E in all of the steps taken to develop the test design. DOE is constructed collaborating with all stakeholders (DOT&E, program office, resource sponsor, Fleet user community, associated NWC, OPNAV T&E rep (N942), SME/analyst support (NUWC/CORONA/etc.), other OTAs). Deltas amongst the working group members will be adjudicated at the DWG.

If the OTD does not consider design changes proposed by stakeholders to be value-added, do not change the design. These inputs will be adjudicated at the DWG.

☐ 1. Identify test objectives for the SUT.

Up to this point, the MBTD process has been proceeding down the "system's engineering V", adding more detail at lower levels. With the creation of test design, move up the "V" by combining the various MBTD items into an executable test. It is beneficial to work the DWG checklist and the first few steps of the IPR-2 checklist in parallel. Both sections are experimental design efforts that will directly relate through run matrices, data gathering, and test execution.

a. The format of testing will hinge on goals set for test. Some can be met at-sea. Others can be met through modeling and simulation. Still others can be fully satisfied with DT data. This process impacts all test design, not just statistics.

b. Test must focus on characterizing performance of critical tasks across the operational environment (main effects and interactions).

c. Ensure that critical measures and their associated data requirements cover these objectives.

d. Begin thinking about the statistical designs that will be needed to meet these objectives.

☐ 2. Draft the IEF "Critical Tasks & Measures" paragraph and table for each COI (see IEF template).

This paragraph should provide an objective statement (characterize across the operational envelope) and identify the overall approach for evaluating each COI by pointing the reader to the critical tasks and measures used to evaluate that COI. The identification of critical tasks and measures does not imply that other measures mapped to that COI in the traceability matrix won't be looked at; just that the critical ones carry more weight in the assessment.

☐ 3. Export critical tasks to critical measures tables from IEF dB and insert into section 2.

☐ 4. Identify potential response variables from the list of remaining critical measures. Identify as many as required to cover the mission(s).

Response variables are critical measures that must be analyzed with statistical methods to support conclusions in the report and whose result may be influenced by controlled conditions (factors). They are used in planning to ensure a minimum-adequate sample size (number of runs and/or data points) and proper design (factor variations).

a. Response variables should be:

i. Testable (i.e. practical, able to collect data on)

ii. Reliable (i.e., relatively free of random noise)

iii. Valid (i.e., represent an essential aspect of SUT performance)

iv. Meaningful – a direct measure of the mission performance we are interested in; overall mission performance or key elements of a mission task breakdown (one or multiple critical tasks)

Ideally, response variables are explicitly identified and have a threshold in a requirement document. In some cases, OTD's may create response variables to better capture the SUT performance (criterion is usually "No Threshold").

v. Quantifiable (i.e. either a numerical performance measure or some qualitative characteristic for which a numerical scale can be directly developed)

vi. Discriminating – should distinguish levels of effectiveness

vii. Preferably continuous vice discrete

CAUTION: Using binomial/discrete response variables should be avoided, in favor of continuous variables if at all possible. While binomial variables can provide just as much info regarding system performance, they require significantly more data.

The type of statistical test for an RV is dictated by the distribution of the RV.

- ***Continuous variables can be plotted along a range of values on a numerical scale (e.g. time, range, speed). These are often normally distributed, meaning the frequency of occurrence of values follows the bell-curve, and allow for the use of a variety of statistical analysis techniques. Non-normal distributions may require conversion to normal values prior to analysis, or may rely on different statistical tests.***
- ***Binomial variables are discrete yes/no, probabilities, proportions, etc. and do not provide operators with as much insight into the performance of a SUT in the intended environment. There are also other types of discrete variables (e.g. count data such as number of false alarms which usually follows the Poisson distribution), etc.***

b. For each response variable, determine if historical data from previous evaluations exist. This includes both previous OT and DT data. If available, review with divisional and 01B analysts.

Historical data are extremely useful in the subsequent DOE planning steps. They can provide a baseline for performance of a legacy system, validate assumptions in the numerical behavior of the measure (type of distribution, standard deviation, etc.), serve as the basis for screening of factors impacting the RV, or justify an effect size for expected test data.

c. Identify the statistical test objectives for each RV. Common objectives include:

i. Characterizing performance across the operational envelope through main effects and interactions (developing a design that supports ANOVA or logistic regression, if applicable)

ii. Verifying performance is above a criterion across all conditions

iii. Verifying performance is above a criterion in a specific subset of conditions

iv. Verifying new system is as good as a legacy system

The objective statement varies with chosen test objectives. Consult with your CTF to be certain your objective statement is accurate to the statistical test goals. Also ensure the objective for this RV is consistent with the overall objective written for the COI in the critical tasks and measures paragraph.

☐ 5. Identify the conditions that are associated with the selected critical tasks and measures and the selected response variables. These should already be linked to the relevant tasks/subtasks and can be found by reviewing the Traceability Matrix.

This also serves as a sanity check of the conditions associated with the task. Previously, conditions were linked to tasks, and then measures were linked to tasks. Those conditions should also logically affect the measures if this was done properly. When thinking about conditions/factors that affect the critical measures, if there are conditions that affect those measures but weren't previously identified or linked to the parent task, then add them to the conditions directory and update the linkage. Do not overlook conditions that can't be controlled (recordable) but are important to collect data on to understand and analyze system performance.

☐ 6. Prioritize the conditions associated with each response variable by the anticipated impact they will have on SUT or operator performance. Determine factors for the test design.

☐ 7. Prioritize the levels of each condition (which were previously identified as part of the Touchpoint 2 checklist) as they apply to each critical measure.

a. Estimate the effect that different levels of conditions have on the performance of the SUT as the condition changes between descriptors/levels (i.e., significant/moderate/low effect).

b. Estimate the likelihood of encountering the different levels in an operational environment (i.e., all levels are equally encountered, some are seen more than others, etc.).

This step focuses the test design on the most operationally relevant environment/scenarios. Include outside organizations in this process. The goal is to ensure that the test design includes and focuses on the conditions that are most operationally relevant. This should be done for both controlled and recorded conditions.

c. Use the below table as a guide in assessing the levels of each condition.

		Likelihood of Encountering level During Operations		
		Multiple levels occur at balanced frequencies (e.g., 1/3, 1/3, 1/3)	Some levels are balanced, others are infrequent (e.g., 5/10, 4/10, 1/10)	One level dominates (e.g., 4/5, 1/10, 1/10)
Effect of Changing Level on Performance		**Balanced**	**Mixed**	**Dominant**
Significant Effect on Performance	**High**	Vary all	Vary balanced levels. Demonstrate infrequent levels	Fix dominant level. Demonstrate others.
Moderate Effect on Performance	**Medium**	Vary all	Vary balanced levels. Demonstrate others	Fix dominant level. Demonstrate others.
Low Effect on Performance	**Low**	Fix levels or record level used	Fix levels or record level used	Fix dominant level

d. The result of this exercise is the identification of levels of controlled conditions that have an important effect on the performance of the SUT and are likely to be encountered by the operator. They will be used to design a test with statistical power and confidence. The descriptors that have a low effect or are encountered infrequently may only be demonstrated.

e. When a single level dominates, testing may focus on the dominant level, with demonstrations for the other levels, if appropriate.

f. Efforts should be made to define factors as continuous vice categorical. Continuous factors often afford greater power, but not always.

☐ 8. Meet and discuss proposed response variables, prioritized conditions (factors) and selected descriptors with divisional analyst and 01B CTF.

a. Define the objective of the test

b. Identify response variables and their associated criteria

c. Prioritize conditions with selected factors/levels using the above matrix

d. Develop list of all variations used for designing a statistical test. These variations will be used to create the run matrix.

See below example:

Conditions		
Altitude	**Airspeed**	**Target**
0-5K	1-100	RCS A
5-10K	100-200	
10-15K	200-300	RCS B
15K+	300+	

In this example, the OTD may have determined that low altitude, low airspeed, and airspeeds above 300kts will have a low impact on performance or are not likely to be encountered by the system. They may be considered for demonstration runs, but won't be included in the statistical DOE. The remaining levels will be used as factors in the design to meet statistical power.

e. Identify disallowed combinations by reviewing the list of factors/levels for combinations that are not testable or realistic (i.e. arctic terrain & hot temperatures, etc.).

If there are a significant number of disallowed combinations, consideration should be given to splitting the test design into separate stand-alone designed experiments. It is important to get operator feedback during this step (i.e., those who will execute the test).

f. Review historical data

g. Review known limitations to test

h. Review statistical objectives for the RV.

☐ 9. For the response variables selected, estimate the following with 01B CTF and divisional analysts:

a. Anticipated distribution of data (continuous normal, continuous but skewed such as $\chi 2$, binomial (pass/fail), Poisson (small discrete integer values), etc.)

Terms/parameters of the distribution can change with factor effects. Often defining the distribution means defining the entire model for the response.

b. Standard deviation (variability) of anticipated data for continuous variables.

Preferably the standard deviation is estimated from historical or DT data. Sigma can be roughly estimated based on the expected range of data, if required. Subtract the minimum from the maximum anticipated value to derive the range. Divide this range by 4 to get sigma. This approach is valid for normally distributed variables, and becomes riskier as distributions depart further from normality.

c. Anticipated factor effects and effect sizes

i. Determine main effects and interactions to be investigated. Identify the most limiting/important factor.

This analysis may include main effects only, or may extend to 2-way and 3-way interactions, or even quadratic terms. More complex interactions often have lower power because of a lower effect (paucity of effects concept).

Different effects will have different powers based on the type of factor (continuous, discrete) the number of levels, correlations with other factors, etc. The most limiting factor will have the lowest power, thus determining the minimum test size. Discovery of the most limiting factor may take several design iterations. Sometimes test is sized for the most important factor, rather than the most limiting.

ii. Determine appropriate effect sizes.

Effect size is related to the sensitivity of the test and can be thought of as the difference in performance that the warfighter will care about and that a statistical analysis needs to be able to detect in the data (if that difference is actually present).

This effect size may be the difference in performance between factor levels (e.g. high vs. low altitude bombing accuracy) or from a specific value (margin above or below where performance is critical) like a threshold.

For binomial tests, the upper and lower value on either side of the effect size must be known for analysis. The same is true for Poisson tests. Continuous metrics require values that can be expressed as a Signal-to-Noise Ratio (SNR). SNR is the effect size divided by the standard deviation.

iii. Determine any limitations on the design, including correlation and confounding.

d. Appropriate confidence and power levels.

See the IEF template for definitions of power and confidence and their related terms. 80% is often the target value for both, with alpha set to 0.2 as an entering argument to the test. However, lower values of alpha may be chosen for systems where there is significant need to avoid type I errors. Lower/higher power levels may be acceptable. Choices of these terms will require justification in the IEF.

☐ 10. Meet with the divisional analysts, 01B CTF/DOE support to discuss inputs into the DOE calculations.

a. Following this meeting, 01B will provide the following for each response variable:

i. Recommended type of statistical test/analysis method

for each response variable

(1) DOE Analyses: analysis of variance, regression, response surface modeling, logistic regression, etc.

(2) Non-DOE Analyses: one-sample t-test, binomial test of proportions, etc.

The analysis method will be consistent with the objectives for the response variable.

ii. Proposed run matrix

The run matrix will be drafted based on the objectives the disallowed combinations, and the anticipated factor effects. Ensure these are understood before calculation.

Descriptors specific to the design can be added to MBTES.

- ***Select from list associated with categorical factors. If level added, add to conditions directory.***
- ***Refine values or intervals for continuous/discrete factors. Do not make changes to conditions directory.***

iii. Power/sample size calculations including confidence, power, sample size, effect size, and any other amplifying notes and assumptions (to be incorporated as notes within the DOE run matrix table, Table C-1 Vignette to Subtasks to Conditions matrix).

b. Additionally, 01B will draft tables for the IEF section 2 presenting the test power.

i. The first table shows power versus varying sample sizes and effect sizes for the most limiting and/or critical factor effect or interaction

ii. The second table shows power versus varying effect sizes at the chosen sample size for all the other main effects and interactions consistent with the anticipated analysis. Note any correlations here. Also note any significant terms that cannot be estimated

c. Review run tables to ensure the selected design is executable.

☐ 11. Draft the relevant statistical design paragraphs of the IEF for each response variable.

a. Explain the objective. Describe the response variable, why it is critical, and identify the criterion value. Detail the expected distribution and associated assumptions.

Include rationale for your assumptions. Articulate the underlying physical/engineering justifications behind the distribution, or detail the data used to support. Any transformations of data (i.e. log-normal to normal) used for continuous variables must be explained here.

b. Discuss conditions/factors chosen and their operational relevance.

i. For the controlled conditions, list the levels for the condition applicable to this response variable, explain why they are controlled and why the levels were chosen.

Anticipated effects can be explained here. Factors can affect several parameters (mean, sigma, etc.) within a model. Detail those effects here as required.

ii. For constant conditions, list the constant level and explain why they are constant

"Because the CDD said so" is not an acceptable reason. All justifications must be operationally relevant.

iii. For recordable conditions, explain why they are important enough to record; but not able to be controlled, or why it was chosen not to be controlled.

Often a recordable condition can be the most important factor affecting the RV, but just cannot be controlled because of physical limitations (weather), or expense.

Depending on the importance of the recordable condition, the range or levels of the condition can be added to adequately anticipate covariate regression analysis.

c. Explain the test design (full factorial, CCD, etc.). All assumptions should be addressed

i. Include any disallowed combinations

ii. Consider inclusion of a summary table of the test points (see below) for complex designs that require more clarity on how the test space is covered

		Mission	Attack				Defend			
Illumin	OPFOR	Terrain	Urban	Mixed	Forest	Desert	Urban	Mixed	Forest	Desert
Day	Low		1	1		1	1		1	
Day	Med		1		1	1		1		1
Day	High			1	1		1			1
Night	Low		1	1	1			1		1
Night	Med			1	1	1	1		1	
Night	High		1			1		1	1	

d. Provide the statistics associated with the design. Explain the operational relevance of the effect size and how it was chosen.

☐ 12. Review the critical measures table for any critical measures (non-response variables; unaffected by controlled conditions) that require amplifying information not contained in the MBTD appendices.

The majority of the testing should be addressed using experimental design for the response variables (with factors). The sections on simple experiments and one-sample hypothesis tests against a criterion are included for completeness, but are not the emphasis of OT&E.

a. Draft paragraphs for any critical measures requiring associated confidence intervals. Explain the assumptions (distribution, parameters, etc.) like you would for a response variable. Follow the IEF template format and best practice #2.

The objective is to characterize overall performance using a summary statistic (e.g. mean or median) and an associated confidence interval. A 1-sided interval is often used for comparison to a criterion. A 2-sided interval shows that sufficient accuracy can be achieved in testing this measure. Both forms can be used to justify or confirm a sample size.

b. Identify critical measures that clearly should be evaluated via demonstrations. Analysis by demonstration usually applies to measures that will be evaluated qualitatively, under multiple conditions (multivariate), or under recordable and/or constant conditions.

There are several types of demonstrations: (1) Some measures deemed critical can be evaluated quickly via demonstrations and do not need multiple data points to evaluate (i.e. the ability to load the SUT onto a C-130 may be critical and can be verified by demoing it once). (2) Alternatively, demonstrating it several times using different support equipment would be a multi-variate demonstration.

i. Paragraphs in section 2 are not required

ii. Vignette DOE notes will identify demonstration runs as part of vignettes that are demonstration only, or within vignettes that are primarily run for DOE completion.

☐ 13. Verify M&S resourcing is apparent in test designs.

Test scoping for M&S is reviewed at the DWG so that the associated resourcing can be approved at the E-IPR. For each RV with M&S runs, the M&S test design description includes OT Runs-for-the-Record (RFR) designed for performance characterization, excursion runs, and runs for replication of live events to support VV&A.

Different simulations run at different speeds (e.g., real time, thousands of run/hour). Provide an estimation of the time required to conduct M&S runs so that the trade-offs between factor choices, design size, test accuracy, etc. can be understood.

Finally, the RV test design must include a separate paragraph listing live events that are required to support VV&S of the M&S.

Do not duplicate OT resourcing details (IEF section 4). Simply provide the understanding necessary to approve those resources.

☐ 14. Summarize time required to complete design runs.

All test design paragraphs (not just M&S) need a brief statement on how much test time results from the STAT/scoping choices made.

☐ 15. Provide DRs related to RVs for analyst review.

Analyst will verify the DRs contain the clear detail required so the delivered data is human-readable (format, fidelity, and source) to support analysis. Without this check, extraction of workable information from raw data may require unsupportable man-hours to complete. Test teams should already try to meet this intent in DRs. For example, latitudes and longitudes supporting a range metric should already be translated into individual range samples.

☐ 16. Ensure all COIs are covered in section 2.

It is beneficial to work the first few steps of the IPR-2 checklist in parallel with the DWG checklist. Both sections are experimental design efforts that will directly relate through run matrices, data gathering, and test execution.

☐ 17. Review the PMT View Shell (PV-0). Verify formatting is correct per the PMT View Annex to this checklist.

PV-0 is an automated output of MBTES, and is embedded in TIEF appendix A. The PV-0 is uncolored, but it provides the format to be used to create the other views as the test program progresses. Thus, the PV-0 is a baseline depiction of the OT requirements for a given SUT. It graphically displays the MBTD-developed test design.

CAUTION: The PV-0 may be handled on NIPR and/or SIPR. All other PMT Views contain test results, and should normally be handled only on SIPR.

☐ 18. OTD, LTE, and CTF review all DWG products and discuss (ideally agree on) readiness for leadership/stakeholder review.

Products still don't have to be 100% final, but it is very important that the experimental design be fully understood and agreed-upon within the building before external review.

☐ 19. Request stakeholder feedback on final products (DOT&E and the program office are most important).

Ideally, continuous core team collaboration should make this step redundant.
Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

ONI feedback must be requested on any IEF involving scenarios, threats, or other MBTD contents that involve current intelligence.

Deliver formal read-aheads to all stakeholders at least 2 weeks prior to the meeting. DOT&E has agreed (per the RCRM policy) this lead-time is the minimum they need to provide feedback and to participate in the review with final decisions authority on the material. DOT&E AOs can sometimes support meetings with shorter notice, but are under no obligation to have final feedback sent in prior to the meeting.
The cognizant IDA Warfare Lead (Air, Naval, or Expeditionary Warfare) must also get copies of the read-aheads. This step is best completed in concert with scheduling the meeting (see below).

☐ 20. Adjudicate/incorporate stakeholder feedback.

Significant disagreements with outside agencies are a vital discussion point. If these exist, it is prudent to pre-brief leadership on this before conducting the meeting.
If there isn't time or agreement to incorporate the comments, ensure the items are ready to be discussed at the meeting.

☐ 21. Schedule DWG and provide read-aheads.

The greater the scheduling lead-time, the easier it is to deconflict calendars. In-person attendance requires more warning. Send the invite at least 2 (but not more than 6) weeks ahead. The meeting can be scheduled before stakeholder feedback is returned.

Products to provide (exported from appropriate MBTES layer, likely IEF node):
- ***All TP-1/TP-2/IPR-1 MBTD products (2 hard copies available for reference, 3 for oversight).***
- ***IEF section 2.***
- ***Draft run matrices (optional; created in excel or using IEF tool).***
- ***PMT View.***
- ***Stakeholder feedback.***
- ***CRM, if used.***

Personnel to attend:
- ***Division – ACOS or DACOS, Section Head, OTD, VX representation, Contract support, Division Analyst.***
- ***Support – 01B or 01B1, CTF, LTE, 01C Rep, VX CNA rep.***
- ***Outside agencies – Dial-in is acceptable for some, but in-person attendance is critical for the key players. The same stakeholders for IPR-1 (see this block in IPR-1 checklist) should be invited, including N942 and ONI. The cognizant IDA Warfare Lead must also be invited for oversight programs.***

Meeting time:
- ***Large systems may take 4+hr. The smallest systems may take 1hr (schedule a minimum of 2 hours).***

DOT&E is mandatory for oversight programs. The PM, resource sponsor, and associated NWC are also high-priority attendees. Without these players, full discussion of the proper test design is jeopardized. Consideration should be given to rescheduling the DWG if key stakeholders cannot attend.
Program T&E and tech leads, Fleet user community, and OPNAV T&E rep (N942) are optional; inform them it's happening. Invite SME/analyst support (NUWC/CORONA/etc.) if involved. For Joint programs, coordinate with other OTAs to consult the equivalent personnel in other services (Joint PMs, sponsors, user reps, etc.).
If key outside agencies listed above cannot attend or dial in, the warfare division ACOS/DACOS must contact them by phone and discuss the value of their attendance/feedback at the DWG. They must understand that the DWG is a forum in which all players can provide vital insight (on SUT functions/design, mission execution, etc.) and defend their positions, and that this meeting often has a significant impact on test resourcing. Confirmation that this was done will be discussed at the E-IPR.
If the stakeholder still chooses not to attend, an email will be sent at the Front Office level from OPTEVOFOR (likely 00D) to the stakeholder requesting formal negative reply to our invite. OTD, provide the information to support.

☐ 22. Conduct DWG.

Begin the meeting by setting the classification level and noting if electronic communication means are used. Then do quick introductions. Then summarize how the meeting will progress, and the goal at completion. Include how the RCRM works. Encourage attendees to bring up any unresolved comments as that section is reached in the review, and state that the test team will be doing that for anyone not in attendance.

After reviewing any changes to IPR-1 material, display section 2 on the screen. Put section 2 on the table. Have a folder of reference documents ready to be accessed electronically. Don't read any of the products. Talk to them, and allow for questions. Do more than just announce a section and ask for feedback.

Meeting steps and key concerns:

- ***Prior products—Move quickly through all MBTD products previously discussed at IPR-1 (they're already approved). Cover comments.***
- ***DOE write-up – discuss section 2. The group will move COI by COI, measure by measure, reaching agreement on the design/scoping for each RV and each critical measure before proceeding to the next. Within each COI:***
  - ***First review table 2-1 (etc.) for an overall perspective on selection of critical tasks and critical measures. Discuss whether the selection of response variables, confidence intervals, and demonstrations is correct and supports proper sizing of test through statistical methods.***
  - ***Then review each measure for adequacy of that design.***
- ***Run Matrices (if provided) – Review these in parallel with section 2. Present at the same time as the Test Design paragraph for the RV.***
- ***Action Items – Review action items, and who has each for action.***
- ***RCRM comments – Establish whether any RCRM comments have come out of the meeting.***
- ***Checksum – State whether any checksums concerns from IPR-1 remain.***

☐ 23. Disseminate meeting minutes/action items and draft RCRM for stakeholder review.

Completed within 2 work days of meeting. Include a list of attendees.

Follow the RCRM policies from earlier in this checklist to clarify comments, gather inputs, close the RCRM to further inputs, adjudicate any remain comment, and elevate the issues as needed.

☐ 24. Update stakeholders on action item completion.

Completed within 1 week of meeting.

☐ 25. Close action items and OPTEVFOR-internal CRM; send post-meet MBTD products for stakeholder awareness.

This is especially important to informing stakeholder who could not attend the meeting of MBTD changes.

☐ 26. Document (in the E-IPR brief) how the test design changed at the DWG as a result of external agency inputs to reduce testing risks.

This is not documented in the IEF, only prepared for E-IPR.

There is risk in determining the balance point between test rigor (more test) and test cost (less test). Our initial design sent out before DWG reflects the minimum-acceptable test establish by OPTEVFOR accounting for the risks we're willing to accept. External stakeholder inputs (based on the risk they are willing to accept) may reduce our risk. The deltas (additional resources, etc.) between initial design and final (post-DWG) design must be understood. The front office must know how much of test sizing is being driven by OPTEVFOR, and how much is externally driven.

☐ 27. Update iBOSS PM (DWG complete).

☐ 28. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

E-IPR Checklist

Purpose: Brief the front office on the MBTD developed so far to ensure it meets expectations/standards.

☐ 1. Incorporate all corrections from IPR-1 and DWG.

☐ 2. Resolve stakeholder concerns with MBTD products.

a. When agreement cannot be reached, ensure the concerns are fully understood.

Any outstanding issues are briefed at E-IPR. Disagreements that cannot be resolved at ACOS level will need front office attention.

b. When OPTEVFOR agrees to test design changes driven by stakeholders, document those impacts.

These are already documented following the DWG, but additional changes may occur before E-IPR. Record resource agreed to above the initial IEF design in 'Design Deltas' slide.

☐ 3. Run IEF database checksum. Fix all issues.

☐ 4. Gather M&S information to support the brief.

If M&S is applicable to the test, it has appeared in the MBTD via the DR sources, and perhaps the conditions and the test design. More detailed work will be done as part of IPR-2 preparation. However, M&S status and expectations related to the test will be discussed at the E-IPR.

Include a list of all models (likely the generic requirement, though it can be the specific expected model), scope of M&S resourcing (e.g., length of time to complete M&S runs), live data planned for validation, model funding status, any previous accreditation results, readiness to support OT, and any additional concerns (such as expected OT limitations).

☐ 5. Create E-IPR brief.

The E-IPR is the only MBTD review for which a brief is required. Most of the briefing material comes straight from the IEF, but some is E-IPR specific. See the briefing template for additional guidance.

☐ 6. OTD, LTE, and CTF review all E-IPR products and discuss (ideally agree on) readiness for meeting.

Vignettes are completed after E-IPR. However, it is a good idea to know the basic layout of your test for E-IPR, in case you are asked. A discussion of vignettes (without actually writing them) prior to the E-IPR can give you this.

☐ 7. Schedule E-IPR and provide read-aheads to Warfare Division and 01B leadership, and to the external stakeholders.

The E-IPR brief should be provided to 01B (actual) 72 hours before being sent to the front office (no meeting required).

The front office calendar is busy. Schedule brief early enough to prevent MBTD delay, but not so early as to be unprepared and have to reschedule. See the Flag LT and/or Flag YN for assistance.

PM (O6 level) attendance (telcon or in person) is desired. Schedule the E-IPR early enough to invite the PM and give them a chance to de-conflict schedule/arrange travel. Other stakeholders can attend if they want to. Invite OPNAV N942.

Products to provide to all other stakeholders:

- ***PowerPoint brief – Use E-IPR template.***
- ***IEF sections 1 and 2.***
- ***Appendix A workbook (Tables A-1 thru A-3).***
- ***Sponsor feedback on Derived and Other measures (likely email).***
- ***Table B-1 (Measures-to-Data Requirements).***
- ***Table B-2 (Conditions-to-Data Requirements).***
- ***Draft run matrices (created in excel or using IEF tool).***
- ***RCRMs.***

Personnel to attend:

- ***Division – ACOS and/or DACOS, Section Head, OTD, VX representation, Contract Support, Division Analyst.***
- ***Support – 01B or 01B1, CTF, LTE, 01C Rep, VX CNA rep, 01D or 01DB.***
- ***Outside agencies—Invite same stakeholders as IPR-1.***

Meeting time:

- ***Large systems may take 2hr. Small systems may take 30 minutes.***

☐ 8. Update E-IPR brief.

Review of the brief by warfare and support division leadership can often require product update. Have this update completed before sending read-aheads to the front office.

☐ 9. Provide read-aheads to Front Office.

The only read-ahead to be sent to the front office is the PowerPoint brief.

Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

See the Directors' Executive Assistant for expectation (quantity, format, etc.) on read-aheads delivery to the front office.

☐ 10. Conduct E-IPR.

Begin the meeting by setting the classification level and noting if electronic communication means are used. Then do quick introductions. Unlike other meetings, don't summarize the goals or talk about RCRM use. There should be no new comments to bring up. Specific template slides offer the opportunity to discuss unresolved issues from prior reviews.

Meeting steps and key concerns:

- ***Brief – Present the brief. If the director is moving faster/slower through the slides than you are, move at the director's pace. Do not expect to open any tables/documents except the PMT View and the RCRMs.***
- ***The Cyber, PMT View, M&S, KPPs, Stakeholder concerns, and RCRM slides are very important.***

☐ 11. Close action items and OPTEVFOR-internal CRM; disseminate any changes to MBTD products and/or the RCRM to stakeholders.

☐ 12. Update iBOSS PM (E-IPR complete).

☐ 13. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

IPR-2 Checklist

Purpose: Refine cyber T&E strategy, developing vignettes, refining data collection by vignette, finalizing run matrices, writing test methods, approximating a test schedule, describing modeling and simulation, identifying resource requirements, detailing limitations, and drafting data sheets.

☐ 1. Develop the CS testing approach.

Scope the Cyber effort (justify the OT cyber vignettes/resourcing.
Although this section is approved at DWG, the test team should be creating this content throughout the MBTD effort. For additional information, see they cybersecurity handbook.

a. Detail the cyber T&E Strategy.

Summarize the cyber efforts prior to OT that inform the scoping of Cyber OT. This section will provide far less detail than how DT will describe their testing in the TEMP. Write enough to explain why the planned OT will be "enough". This effort is similar to estimating the amount of maintainability data that will be available for OT analysis, in order to determine if an MDEMO will/may be required. Another parallel is establishing that DT data will be collected by the PM, thus eliminating the need for OT collection. Do not duplicate cyber concept content (section 1).

b. Outline the resulting cyber OT Scope.

List the planned events, and overall expectations for each. Test planning is likely to make changes to these; say so. Major CS data requirements may be included. Do not duplicate OT execution details (vignettes paragraphs) or resourcing details (section 4).

c. If possible, detail the System Re-baselining, Recertification, and/or Off-limits Components.

Operationally realistic cyber test poses risk. Some attacks may not be allowed on real and/or live systems due to dangers they pose to equipment and/or personnel. Some attacks may not be allowed due to sensitivity level, or the rework that might follow. Engage the program office to understand how far OT can go in attacking the system, and the justifications behind the limits.

Detail the following:
System re-baselining: The impacts of cyber test must be corrected, restoring the SUT to a state of operational readiness. What minimum re-baselining is required? What capability does the program office have to

(continued)
support this? Is that time and resourcing affordable?

<u>System recertification</u>: The system likely needs a certification to operate within the SOS. Certifications are difficult and expensive. Will it need to be redone? Is that time and resourcing affordable?

<u>Off-limits Components</u>: Identify the portions of the SUT that cannot be attacked/accessed in a fully realistic manner, and why. This list is likely more nuanced/granular than the MCSM.

It is possible the program office may not know this information, or who to ask. However, it is their responsibility to seek out the necessary authorities for engagement and to resource any additional items that are required beyond what the test team is responsible for (i.e., test artifact removal). The following bullets are for discussion with the program office regarding "authority" types to begin the conversation; there could be other technical authorities that become apparent during the discussion:

- *System technical warrant holder*
- *SYSCOM cyber technical warrant holder*
- *Combat element/system or warfare certification official (NAVSEA systems)*
- *TYCOM (subsurface systems)*
- *Ship Technical Authority*
- *Flight Safety Certification (NAVAIR systems)*
- *System Security Engineer*
- *Anti-Tamper (AT) Tech Warrant Holder, or AT Executive Agent*
- *Information System Security Manager.*

It may only be possible (now, during IEF work) to acknowledge that the above information is not yet know, and to summarize plans for cyber authority engagement to obtain the information.

Ideally, test teams should not wait until test planning to ascertain and document this material. For example, an inaccurate off-limits components list will likely lead to significant re-work of attack plans when the real list becomes know. It is better to agree on the accurate list early, ensuring plans are correct the first time.

Do not simply take the off-limits list at face value. If enough critical systems are on the list, the restrictions may ultimately prevent a Cyber Survivability determination. The list is to be debated and agreed-to in the same way restrictions to effectiveness testing are debated. There should be an operational reason for limits, such as fleet procedures.

Cyber OT scope is potentially bound by the above considerations, leading to test limitations. Acknowledge these limitations (the full details go later in section 3). If mitigating any limitation will be done using cyber T&E augmentation, acknowledge that plan.

d. Identify necessary cyber T&E Augmentation(s).

Some T&E may require capabilities beyond the normal routine. If applicable, detail the augmentations needed. Cyber M&S is one form of augmentation. As with effectiveness M&S, this content must enable the test team to start drafting the M&S requirements letter, by identifying the scope of inputs and outputs to the model (leave M&S details for later in section 3).

e. Summarize the cyber Post-Test Analysis plans.

Document any cyber analysis plans that should not be lost between now and test planning. General statements are acceptable. Merging of data, or anticipated scoring rules are two potential inputs.

☐ 2. Consider all material with an impact on test design.

a. Review the task hierarchy, conditions, measures, data requirements, and the linkages between all of these items.

b. Review the experimental design in IEF section 2.

Often, the statistical/experimental design associated with the critical measures will dictate how the testing must be executed to gather the data for those measures. The design efforts in this checklist must match up with the outputs of the DWG.

c. Review the draft TEMP section 3 (if available), or the T&E WIPT outputs for test scheduling.

The TEMP is the official schedule for all test, including OT. The IEF test design helps create this schedule, and must be consistent with the eventual signed TEMP. Ideally, IEF creation precedes TEMP review. If so, T&E WIPT efforts to produce a coordinated schedule may provide what is needed here.

Plan test with an eye toward reducing data collection required in dedicated OT events. If OT data can be obtained during DT or IT, the test design must account for those opportunities.

d. Obtain DT test plans and schedule (if available), and review them for specific test events that could support OT data gathering.

The details of DT events (procedures, data collection methods, etc.) dictates whether that data can be used for OT. Will Fleet-representative users operate the system with no input from system technicians? Will finalized tactics be used against relevant threats?

e. Understand the status of system development.

What SUT functionality will be available when? Scheduled testing must match up with the delivery timeline for tested functions.

f. Understand the testing formats and/or locations available.

The data that can be gathered in a lab, or using M&S, may be different from data gathered in a real operational environment. The data gathered on an instrumented range may be more accurate or comprehensive than from other test locations.

g. Understand the availability of resources.

Test design must be realistic for resourcing that will support test.

h. Consider the phase(s) of test being covered in the MBTD effort.

OA test events and IOT&E test events are often different in format, scope, etc. That being the case, test design covering both phases must account for these differences.

☐ 3. Outline vignettes that will support data collection.

The creation of vignettes is a subjective process that can be done in many different ways. It requires the test team have a good plan to collect all OT data. Each vignette is just a part of that plan. If the overall plan is not comprehensive/efficient/logical, the vignettes used to outline the plan will be flawed. Review of prior/related programs can help formulate the vignette outline.

Vignettes are logical groupings of subtasks that serve to organize data gathering and support executing test events. While not in-and-of themselves a detailed test plan, they establish the type of testing OT may execute, the data collection plan, and the resources needed to gather OT data.

"Vignette" is not an equivalent term to "test event". Vignettes can stand alone to support an event, or can be combined to create an event. However, vignettes cannot be sub-divided to create test events. If a smaller set of data collection will be needed as a test event, create that smaller vignette.

Knowing the outline of data collection informs the detailed construction of each separate vignette. The DOE written in section 2 is a key reference in outlining a vignette strategy.

☐ 4. Document mission-based vignettes in the IEF tool.

a. Determine the vignette type.

This is based on the type of test to which the vignette will be associated (IT or OT). If the vignette is only to be completed during independent OT, select the OT radial button. Otherwise, select IT.

Vignettes are not created to cover DT execution. Data collection by DT is governed by the program office. It requires no organization by us.

b. Number the vignette.

Vignette numbering is based on the primary COI (number) it relates to, and the numerical order of this vignette within those associated with that COI. Examples: IT S1-1, OT E2-1, OT S2-3, IT E1-7, etc.

All effectiveness COIs require associated vignettes. Do not skip any integers when numbering vignettes for a COI. The same number pair (X-Y) cannot be used, no matter if one vignette is IT and the other is OT.

c. Name the vignette.

The name should be short, but still provide q quick way to recognize what is being covered by the vignette. For example, M&S vignettes should be named to identify they are M&S. Vignettes covering whole plots (or specific factor levels) within a statistical design should probably be named for those factor levels.

d. Select the tasks making up the vignette.

Look for logical groupings of related tasks that flow together and can be combined into an executable test event. Only add lowest-level subtasks. Parent subtasks have no linked measures or conditions, and serve no purpose in vignette construction.

In the simplest form, a vignette could be constructed around one subtask. On the other end of the spectrum, one vignette may be created to describe an end-to-end test event that captures every task under a COI.

Vignettes may cross COIs. Tasks for multiple COIs that are performed simultaneously or tasks that are common to multiple COIs may be combined into one data collection opportunity (e.g. if the prepare/configure task is exactly the same for all missions, just one vignette may be needed to gather all that data).

e. Set conditions to controllable or recordable.

These selections are often driven by DOE written in IEF section 2. Controlled and constant conditions for response variables are both "controllable" (constant conditions are controlled at a single level). All other conditions are recordable (there may be some conditions associated with the vignette that did not apply to the response variable and weren't used for the DOE, but still apply to the vignette).

Even when there is no associated DOE, conditions may still need to be controlled for the purpose of multi-variate demonstrations. Set the conditions accordingly.

In rare cases, some conditions may remain unassigned. But this is only when these conditions simply do not apply. Consider a M&S vignette containing the same subtasks as an at-sea vignette for which the model does not incorporate all of the real-world aspects of the tasks and the associated mission environment. The IEF dB will pull in the same conditions for both. But

(continued)
for M&S, several conditions could be neither controlled, nor recordable. Thus, they remain unassigned in the tool.

f. Build the vignette run matrix.

The controllable conditions will form the columns of the run matrix under which descriptor levels should be set run-by-run (row-by-row). Use one of the methods described below:

- *Use the IEF dB internal run building function to create the matrix.*
- *Export an Excel template for the run matrix from the IEF dB, create the run matrix in Excel, and import the file back into the dB.*
- *Create the run matrix in JMP and import the file into the IEF dB.*

Note: Importing an excel file to a vignette automatically over-writes any existing run matrix for that vignette in the IEF dB. Edits can be made to matrices following import. Conditions and descriptors in the import file must match those used in the IEF dB, or the import will not succeed.

The run matrix must be consistent with the DOE. If necessary, add runs to the matrix for demonstrations outside the DOE. Present the runs in standard, left-to-right order (not randomized) with the factorial design organized by levels. This allows confirmation the design is properly balanced across factor combinations.

CAUTION: Depending on the number of controlled conditions, the run matrix can be difficult to set up for printing. Seek CTF assistance in building these tables.

g. Set measures to test and non-test.

The measures traced to subtasks in the vignette are automatically pulled in, and will initially be unassigned in the tool. Depending on the intentions of the vignette, not all measures pulled in with the subtasks will be tested during that vignette. Thus, some measures are set to non-test.

All measures must be set to test or non-test. Gray measures are set to non-test. DT Only measures must be set to test somewhere, but should really only be set to test once (you'll only need to collect that data once from DT).

h. Refine data requirements being collected.

The DRs linked to measures that are set to test in the vignette are automatically pulled in to be collected. But not all these DRs are guaranteed to be gathered. Consider a measure that is tested in both real-world and M&S vignettes. DRs specific to real-world and to M&S testing will both link to the measure. But the M&S data cannot be collected in the real-world event (and vice versa). The inappropriate DRs must be removed from the vignette to prevent confusion in collection efforts.

DRs that do not apply to a vignette should be hidden. Based on current

(continued)
understanding of planned test execution, examine each DR to confirm that the element will be produced from the associated source at the required accuracy and frequency. If it will not, hide the DR, refine the DR (rather than hiding it), and/or create a new DR to represent the data that will be collected. A good place to start this effort is DR sources, and then categories (if used). The tool (and DRTM) lists DRs by subtask, and then by category.

Hidden DRs can be added back to the vignette by repopulating a particular task, or a particular category of DRs in that task. Taking this action will return the associated hidden DRs to the vignette.

i. Summarize vignette execution.

In the MBTES Test Methods 'Summary' field, provide a top-level description of what will happen during the test to yield the DRs grouped under this vignette. Write such that the data collection process can be understood and approved. This content will also appear in IEF section 3.

Don't draft the Detailed Method of Test (DMOT), but don't leave open questions about what's going to happen. Include (as appropriate) unit tasking, basic run progression, how long each run takes, test restrictions, major OTD actions, test format/location, etc. If the vignette will be combined with others to create an event, that is also vital to know in understanding the data gathering. If the vignette has a run matrix driven by a DOE, say so.

j. If advantageous to the MBTD, establish detailed test methods for the vignette.

Detailed test methods are not necessary for the IEF because detailed test execution will be created during test planning. Vignette test methods do not transfer in MBTES to test events. Nor are they referenced when creating the test plan document.

The Test Method fields are free text. Be sure to save changes before navigating away. Format the inputs as needed for best utility during test. For example, execution direction can be divided up into the numbered actions for each participating unit and for each data collector. Test methods for each vignette fall into these categories:

- *Introduction—do not use.*
- *Pre-test—includes all actions required to prepare for the vignette, both back at the office and immediately prior to execution. Common content for this field is briefing involved units, distributing data sheets, starting automated data collection, and confirming event conditions match those planned for the DOE run.*
- *Test Execution—gives direction on how to conduct the vignette, what testers must do to support execution, and identifies where/how the required data will be collected. Test methods should directly relate to the data requirements presented in the vignette.*
- *Post-test—includes all actions after the vignette, both immediately following execution and back at the office. Common content for this field*

(continued)

is administering surveys, verifying data was collected, and reporting significant results/deviations. DT Data is collected as a part of pre– or post–test. Ensure test method steps include the collection of DT reports, etc.

- *Summary — completed in the prior step.*

If vignette execution is not fully understood at the time of the IEF drafting, consider writing the test methods as 'notes' (at a minimum under pre-test, test execution, and post-test), retaining knowledge from the time of development. OTDs are highly encouraged to maximize use of test methods while the original plan is fresh in their mind. Writing test methods can help identify data requirements and other key test design components that were potentially overlooked during prior MBTD steps. Test methods used in vignettes in similar programs might be a good starting point.

Teams choosing to put a great deal of rigor into IEF test methods should read the test planning handbook touchpoint C checklist, which provides directions for creating the DMOT for test events. This process is very similar to drafting vignette test methods, as the DMOT originates from the intended execution of vignettes within that test event.

k. Draft DOE notes for the vignette.

The DOE Notes field is free text. Save changes before navigating away. These notes must be consistent with IEF section 2, but are not a copy of section 2. They should be much shorter.

Identify the critical measure (M# / name) driving specific runs. Then, summarize design/statistics for that measure. For RVs (DOE), include the type of test, a summary of the design, effect size, confidence, power for the driving

(continued)

factor, and which runs apply. For confidence intervals, include the type of interval, interval values, and which runs apply. For simple hypothesis tests, include the type of test, effect size, confidence, power, and which runs apply.

DOE notes can also explain that a vignette is "demonstration only" if statistics don't apply. Runs added for demonstration (including excursions from any DOE) must be identified as such.

l. Input resources for the vignette.

For each vignette, identify the following:

- *Test articles – number of test articles (full or partial systems) with any specific configuration (include length of time required).*
- *Test sites and instrumentation – specific range, test site, lab, special instrumentation, and/or data collection mechanisms (include length of time required).*
- *Test support equipment – diagnostic or calibration equipment, and/or other devices that do not fit other resource categories.*
- *Threat representation - threat type, number, availability, and fidelity*

(continued)

requirements (include length of time required).

- *Test targets and expendables – type, number, and availability requirements for all targets, weapons, flares, chaff, sonobouys, countermeasures, etc. (excluding threat targets).*
- *Operational force test support – specific aircraft, ship, submarine, unit, or exercise support requirements including flight hours, at-sea time, or system operating time; includes the OTD, analyst, and any additional required contractor support (travel days, underway days, analysis days).*
- *Models, simulations, and test beds – any M&S requirements, including labs (include length of time required), software models/applications, pre-faulted modules for M-DEMO, etc.; plus accreditation agency/personnel and time requirements (typically 1/4 man-year of accreditation effort is required).*
- *Manpower & personnel training – type and number of personnel required with the associated specific operator or maintainer training/expertise (include length of time required).*
- *Special requirements – specific non-instrumentation capabilities, such as special data processing, databases, unique charting, extreme or restricted environmental conditions.*

Some resources could potentially fit in multiple categories. Do not list the same resources twice.

For a vignette with multiple runs, the listed resources are the total for all runs, not the resources for a single run.

☐ 5. Document CS vignettes in the IEF tool.

When CS applies, write a Cooperative Vulnerability Penetration Assessment (CVPA) vignette that looks at the system's protect, mitigate, recover capabilities (as applicable). An Adversarial Assessment (AA) vignette may also be required. A cyber table-top vignette may also be needed as part of IT.

a. Determine the vignette type.

b. Number the vignette.

c. Name the vignette.

d. Select the tasks making up the vignette.

e. Set conditions to controllable or recordable.

f. Build the vignette run matrix (if applicable).

g. Set measures to test and non-test.

h. Refine data requirements being collected.

i. If advantageous to the MBTD, establish detailed test methods for the vignette.

j. Draft DOE notes for the vignette (“None”).

k. Input resources for the vignette.

☐ 6. Document suitability vignettes in the IEF tool.

There are two options for including the Reliability and Availability (R/A) COIs in vignettes in order to gather that data:

1. ***Add the COIs to all vignettes based on other tasks (usually effectiveness) in which R/A data will be available for collection (essentially all vignettes).***
2. ***Create a Reliability and Availability vignette that is not task-based, but serves as a data gathering procedure to be run in parallel with all vignettes making R/A data available. This option reduces duplication of DRs across the task-based vignettes.***

See the IEF template section 3 for more information on these.

Vignettes related to the Maintainability COI can be task-based if there are maintenance tasks, but maintenance vignettes can be created solely from the COI. They generally take two forms:

1. ***The maintenance action vignette governs data collection when corrective or preventive maintenance is performed as required by the Fleet during real-world test.***
2. ***The M-DEMO vignette is written for a staged maintenance event that is completed to supplement maintenance data collected (or not collected) in the maintenance action vignette.***

See the IEF template section 3 for more information on these.

Unlike reliability and availability, maintenance data collection (usually) cannot be added to effectiveness vignettes without changing the vignette execution. Corrective maintenance is not a regular part of the effectiveness task flow, and this data is gathered when missions (vignettes) pause due to failures. Regular preventative/pre-mission maintenance can be added.

Logistic Supportability data can be collected in several forms:

1. ***Logistics subtasks call for a specific vignette to be created (i.e. loading weapons, hooking up pier-side support).***
2. ***The maintenance action and/or M-DEMO vignette provide logistics data. Add logistics COI/subtasks to these vignettes.***

a. Determine the vignette type.

b. Number the vignette.

c. Name the vignette.

d. Select the tasks/COIs making up the vignette.

e. Set conditions to controllable or recordable.

f. Build the vignette run matrix (if applicable).

g. Set measures to test and non-test.

h. Refine data requirements being collected.

i. If advantageous to the MBTD, establish detailed test methods for the vignette.

j. Draft DOE notes for the vignette (“None”).

k. Input resources for the vignette.

The R/A vignette should not have any resources. If it does, these are only instrumentation and/or personnel unique to collecting DRs for R/A measures.

The maintenance action vignette follows the same convention as the R/A vignette. Test should not need extra resources to observe real-world repairs.

The MDEMO vignette must have resources. This is the primary reasons it is separate from the maintenance action vignette.

☐ 7. Audit vignettes.

The IEF tool checksum is a powerful mechanism assisting in completing this step. Start your audit by fixing any errors in the checksum related to vignettes. Determine if any warnings point to required corrections.

a. Verify each vignette is logical and executable.

Confirm the subtasks the OTD selected actually apply to the vignette. Incorrectly associating a task to a vignette can bring along measures, conditions and data requirements that do not actually apply to that vignette.

Confirm the data requirements are comprehensive for the vignette, and all the listed data can be collected. Review test methods for logical test event flow.

b. Verify all active subtasks/COIs are linked to vignettes.

Every non-gray lowest-level subtask (or COI) must be associated with at least one vignette. This won’t tell the OTD if the linkage makes sense, but it will identify any subtasks that have been overlooked. Gray subtasks do not link to vignettes.

If a subtask does not fit logically in any vignette, it may not be valid. Or a new vignette may be needed to exercise the subtask.

c. Verify all active measures are tested.

All non-gray, non-orphaned measures must be set to test in at least one vignette. If the measure cannot be set to test in any vignette, it may not be a valid measure. Or a new vignette may be needed to test the measure.

d. Verify all active data requirements are collected.

Every DR associated with an active measure must be collected in at least one vignette. If a DR is hidden from all applicable vignettes, the DR might need to be deleted. Or a new vignette may be needed to collect that data.

e. Verify the DOE and/or major test objectives for the SUT are logically addressed in the vignettes.

Ensure vignettes correctly include RVs, and are consistent with data collection for the critical measures and statistical designs in section 2 of the IEF. Compare the run matrix to the section 2 test design.

f. Verify that the vignettes cover all major test events, formats, and/or locations.

The IEF produces vignettes to inform the test schedule/strategy in the TEMP. If there are multiple phases in the TEMP, different vignettes for those phases may need to be created. Start by creating the full/comprehensive version of each IOT&E vignette. Then, determine if other test phases will use those vignettes or will need other vignettes to facilitate data collection.

The tasks that can be executed in a vignette conducted in a lab or using M&S may differ slightly from those executed in a real operational environment. The data collected and test methods used will certainly be different. Additionally, the resources required for the two different formats will be very different. This requires creating separate vignettes for each test format.

The test methods for one location, or using one major resource (e.g. threat surrogate) may be different from another location or major resource. The vignette should not have two sets of test methods. Rather, every vignette must be constructed so that the test methods to be executed are (almost) the same every time. If they're different, you have two vignettes.

Consistent with all of the above, any major difference in current test planning or eventual test execution is likely best accounted for through separation between vignettes.

g. Verify vignette organization will take best advantage of test resourcing.

For example, an AW engagement has a clear Detect to Engage sequence of tasks, but due to missile availability, many executions of this full kill-chain vignette are not possible. The tasks can be broken into a vignette covering the Detect to ID sequence that could be run numerous times under a broad set of conditions, while a second vignette captures the Engage task associated with an actual missile shot under a narrower subset of conditions and fewer runs.

The following steps will detail drafting of section 3 in the IEF. Ideally, this section will provide a cut/paste input for the OT portion of TEMP Part III. The level of detail in each item should be on par with that required by a TEMP.

☐ 8. Draft the IEF operational evaluation approach section.

Review the planned phases of test (DT/IT/OT) in the integrated master schedule being developed by the program office through coordination with the T&E WIPT. Describe the major phases and test periods relevant to OT data gathering, and provide a top-level summary of how OT will participate in each. Then (as needed), adjust the template paragraphs covering operationally representative testing and actual test conditions.

☐ 9. Draft the IEF vignette descriptions.

Pull the IEF Vignette Summaries from MBTES and input to IEF section 3. Make any appropriate edits, such as pointing to the DOE paragraph in section 2 that relates to the vignette, which can simplify the explanation.

Even if the IEF contains multiple test phases in section 3, do not describe the vignette in two different test phases/periods. Only one description is needed. If there is different execution, you have a different vignette.

☐ 10. Draft the IEF schedule of events.

Outline when the vignettes are expected to be executed (during which test period and/or test phase). Creation of a table is recommended. Account for every run of every vignette.

Exact order of execution, or scheduling within test periods is not required. Assignment of runs by period is only necessary to identify resources by period and provide that TEMP input. Combination of vignettes into events is necessary detail if that affects resourcing (vignettes share resources).

If it appears (based on the schedule of events) that the test design written in IEF section 2 is not executable, work with your CTF and 01B analyst to update the design.

☐ 11. Determine Resource Requirements by phase.

Enough granularity needs to be provided in the IEF to support the generation of OT funding requirements documented in the TEMP.

a. Using the resource requirements for each vignette and the proposed schedule of vignettes in the Vignette Execution Strategy developed earlier, identify the test resource requirements for each test period (DT-B1, IT-B2, OT-B2, OT-C1, etc.).

b. Create the IEF Test Event Resource Matrix.

The test event resource matrix (table 4-1) comes directly from the vignette resource matrix (table 4-2). Be certain they are consistent.

c. Consider the need to officially schedule resources now,

rather than waiting for TEMP inputs.

CS Testing is conducted by OPTEVFOR CS Test Team. Resources must be scheduled 1 year prior to testing, and funded properly.
Fleet service requests must be submitted to fleet schedulers NLT 9 months prior to test.

☐ 12. Compare the historic (SUT relevant) environmental data of planned test locations/dates to the historic environmental data of the priority TACSIT or OPLAN/ CONPLAN geographic areas.

Verify the desired test conditions can be expected in the test location(s) and date(s). Identify limitations that will result from condition deltas between test areas and real-world OPAREAs. Clearly articulate any limitations to test in terms of the environmental condition space not explored, and the significance in relation to expected performance in priority OPLAN/CONPLAN area(s).
Any limitation regarding environmental conditions during test must be specific. It is obvious that the test area won't present all conditions, so a generic limitation adds no value.

☐ 13. Write the IEF Modeling and Simulation section.

Every model/simulation called for in the vignette resources needs a separate paragraph. Remember that target emulations/augmentations and threat surrogates are also M&S.
Start with target/threat models used in live test. They are easy to grasp, and require less detail. Then discuss models being used to produce OT data. Cyber M&S may be relevant.

a. Describe the M&S.

This is similar to describing your SUT (but shorter). What constitutes the model (components, software, etc.)? How does it function (assumptions, inputs, calculations, outputs, etc.)? What does it simulate (units, tasks, environment, etc.)? Who runs it (if applicable)? What would a model "run" look like (timeline, operator involvement, etc.)? What are the limitations?

b. Explain the intended use for the M&S.

State the COI(s) and critical measure(s) being supported. What operationally relevant output from the M&S will be assessed for OT? What operational space of interest (factors and levels) will be seen in the M&S? How many runs will be completed? What data requirements for measure evaluation are being filled? Point to the vignette paragraph(s) to which the model applies. How will these outputs be used in forming OT conclusions? Point to the section 2 paragraph explaining M&S test design and post-test statistical analysis of model outputs (if applicable).

c. Summarize the planned accreditation effort.

What metric(s) will be used to determine whether the M&S should be validated (methodology for comparing M&S to live data)? What number of live data points will be required? How will the necessary live and simulation data for validation be collected? Can existing data from validation for a previous test be used (reference those data and summarize how they are relevant for the intended use in the current test)? Will live data cover the entire operational envelope to be explored with M&S or only a portion of the envelope (range of conditions over which the M&S will be validated)?
The explanation may be complex. Accreditation may apply to sub-components of the model, and also focus on the full system or environment being emulated. But do not provide the same level of detail as an accreditation letter. This is just a summary.
What methods will be employed in the accreditation effort (documentation review, face validation, SME evaluation, comparison to other models, statistical evaluation)? DOE methods should be employed as part of the process of determining what live data are needed for model validation, and in the process of determining how well the model reflects reality. Point to the applicable section 2 paragraph containing measures of merit (power, confidence, calculated uncertainties). If statistics are not used, explain why.
Note: the statistical design for accreditation is not the same as the statistical design for model data to be used in OT reporting.

d. On oversight programs, verify the M&S write-ups support briefing the accreditation concept.

The accreditation concept will need to be accepted before use of the model is accepted in the TEMP. The accreditation plan will likely not be approved when this needs to happen, but the IEF can provide a description of the model sufficient to produce the accreditation concept. The DOT&E memo "Guidance on the Validation of Models and Simulation used in Operational Test" from 14Mar16 applies.

e. Verify the M&S write-ups are comprehensive to all phases of test covered by the IEF.

Models are accredited for specific purposes. The uses for each model and the related accreditation efforts are likely different between phases. Create a paragraph for each phase, of each model.

☐ 14. Document IEF test limitations.

List the affected COI(s) in the title of the limitation. 01D may be consulted on which /whether CS limitations apply.

a. Select the limitation category.

The limitation category assumes the mitigation is implemented. For example: limited operating time during OT for a continuous system calls for a limitation on resolving MTBOMF. The limitation is severe/major/minor if the hours being collected are <0.5xMTOMF/ 0.5-1.0xMTBOMF/1.0-3.0xMTBOMF, respectively. But if the mitigation will yield additional test hours (perhaps from Fleet ops) such that the total test time changes the limitation severity, place the limitation in the category of that reduced severity.

There are no severe limitations prior to IOT&E. EOA/OA are not resolving COIs. Thus, severe limitations precluding COI resolution cannot exist.

b. Describe the limitation.

Identify the realities of test which deviate from the ideal test, making it impossible to collect specific data required for OT. These realities include unavailable resources (assets, threats, targets, personnel, data collection mechanisms), insufficient time, limited samples, data collection inaccuracies, unrepresentative test (conditions, procedures, personnel, training), and more. If any DRs are not being collected (in part, or in full), identify that limitation.

c. Outline the steps taken to mitigate the limitation.

Dealing with a limitation most often takes the form of reduction. Discuss how data gathering will compensate for the realities of limited OT, or how OT is attempting to gather sufficient data. However, some limitations have no mitigation. The strategy in these cases is mere acceptance. This must be acknowledged.

d. Describe the ultimate impact of the limitation.

This is the final impact, after a mitigation is successfully implemented. There must be an impact. If there isn't, it's not a limitation.

Focus on the <u>data</u> that will not be gathered, and then describe the implications (analysis/reporting) of not gathering that data. The impact wording is slightly different depending on the phase of test. Limitations for OAs relate to assessing risk, while those for IOT&E relate to resolving COIs.

e. Verify limitations are comprehensive to all phases of test covered by the IEF.

Limitations at OA are not likely to be the same as those at IOT&E. The vignettes for each phase should reflect the testing realities of that phase. Knowing the differences in vignette execution between phases illustrates the differences in limitations by phase.

☐ 15. Write event cards, data sheets, surveys, and interviews (optional).

IEF Appendix C is optional, and should be deleted if not used. Like test methods, it's ideal for the test team to document current test execution and data gathering intentions. However, the additions of such detail can wait until the test planning process. If a survey is included, it must be consistent with survey best practices available at Y:\OT&E Production Library\Test Plan and DCP\Surveys.

☐ 16. Run IEF database checksum. Fix all issues.

☐ 17. OTD, LTE, and CTF review all IPR-2 products and discuss (ideally agree on) readiness for leadership review.

CAUTION: Prior to generating the final vignette tables from the IEF dB, check each vignette for discrepancies. If MBTD items from earlier steps are modified, this may impact the vignettes if they're already built or drafted. For example, if a vignette is built and formatted and is considered complete, but an additional measure is linked to a task, that measure is also added to all vignettes that cover that task. The measure should be set to test or no test for each vignette or the measure and data requirements associated with it won't be displayed when the vignette is produced.

☐ 18. Request stakeholder feedback on products.

Ideally, continuous core team collaboration should make this step redundant.

Ensure classification markings are correct before sending. Classified material needs correct markings at all times, not just at signature. This includes derivative classification material on IEF cover page.

Deliver formal read-aheads to all stakeholders at least 2 weeks prior to the meeting. DOT&E has agreed (per the RCRM policy) this lead-time is the minimum they need to provide feedback and to participate in the review with final decisions authority on the material. DOT&E AOs can sometimes support meetings with shorter notice, but are under no obligation to have final feedback sent in prior to the meeting.

☐ 19. Adjudicate/incorporate stakeholder feedback.

Significant disagreements with outside agencies are a vital discussion point. If these exist, it is prudent to pre-brief leadership on this before conducting the meeting.

If there isn't time or agreement to incorporate the comments, ensure the items are ready to be discussed at the meeting.

☐ 20. Schedule IPR-2 and provide read-aheads.

The greater the scheduling lead-time, the easier it is to deconflict calendars. In-person attendance requires more warning. Send the invite at least 2 (but not more than 6) weeks ahead. The meeting can be scheduled before stakeholder feedback is returned.

Products to provide (exported from appropriate MBTES layer, likely IEF node):

- *All TP-1/TP-2/IPR-1/DWG MBTD products (2 hard copies available for reference, 3 for oversight).*
- *IEF section 3 and 4.*
- *Appendix B Tables B-3 thru B-X (Run Matrices)*
- *Data sheets and Survey sheets, if created.*
- *CRM, if used.*
- *Stakeholder feedback.*

Personnel to attend:

- *Division – ACOS or DACOS, Section Head, OTD, Contract Support.*
- *Support – 01B or 01B1, LTE, CTF, VX CNA rep, 01D rep.*
- *Outside Agencies – The meeting can be supported through email agreement or dial-in, but face-to-face is preferable. DOT&E is a must for oversight programs. PM, program T&E lead, resource sponsor, Fleet user community, associated NWC, and OPNAV T&E rep (N942) are optional; make sure to inform them it is happening and send them read-aheads. Invite SME/analyst support (NUWC/CORONA/etc.) if involved. For Joint programs, coordinate with the other OTAs to consult the equivalent personnel in the other services (Joint PMs, sponsors, user reps, etc.).*

Meeting time:

- *Large systems may take 3hr. Smallest systems may take 30 min (schedule a minimum of 1 hour).*

☐ 21. Conduct IPR-2.

Begin the meeting by setting the classification level and noting if electronic communication means are used. Then do quick introductions. Then summarize how the meeting will progress, and the goal at completion. Include how the RCRM works. Encourage attendees to bring up any unresolved comments as that section is reached in the review, and state that the test team will be doing that for anyone not in attendance.

After reviewing any changes to IPR-1/DWG material, display section 3 on the screen. Put section 3 and section 4 side-by-side on the table. Underneath those, have the Appendix B run matrices. Have a folder of reference documents ready to be accessed electronically. Don't read any of the products. Talk to them, and allow for questions. Do more than just announce a section and ask for feedback.

Meeting steps and key concerns:

- *Review Operational Evaluation Approach, including cyber.*
- *Vignettes – Walk through vignette-by-vignette; for each one, review:*
  - *The vignette description in Section 3.2,*
  - *The vignette run matrix from appendix B,*
  - *The vignette resources in table 4-2.*
  - *Note: from these products, everyone should understand how the vignette is being executed (without additional verbal explanation). If that is not the case, more detail is required in these three items.*
- *Schedule – Walk through period/phase-by-period/phase; for each one, review:*
  - *The notional schedule of events (made up of vignettes/runs).*

(continued)

  - *Table 4-1 resources by phase, referencing the schedule (vignettes within each phase), and table 4-2 (resources within each vignette).*
- *M&S – The M&S section must be understood to support M&S Requirements Letter creation.*
- *Limitations –Limitations to test must cover the gap between the minimum-adequate test and the test that is resourced.*
- *Checksum – State whether any checksums show errors or warnings. Be prepared to explain any not cleared.*
- *Action Items – Review action items, and who has each for action.*
- *RCRM comments – Establish whether any RCRM comments have come out of the meeting.*

☐ 22. Disseminate meeting minutes/action items and draft RCRM for stakeholder review.

Completed within 2 work days of meeting. Include a list of attendees.
Follow the RCRM policies from earlier in this checklist to clarify comments, gather inputs, close the RCRM to further inputs, adjudicate any remain comment, and elevate the issues as needed.

☐ 23. Update stakeholders on action item completion.

Completed within 1 week of meeting.

☐ 24. Close action items; send post-review MBTD products for stakeholder awareness.

This is especially important to informing stakeholder who could not attend the meeting of MBTD changes.

☐ 25. Request 01B inputs to internal CRM within 2 weeks of completing IPR-2.

☐ 26. Update iBOSS PM (IPR-2 complete).

☐ 27. Offer lessons learned.

The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.

IEF Routing Annex

Purpose: Rapid routing allows the MBTD snap-shot to be accurate at signature. Do not route until minimal leadership comment is expected.

1. Finalize the draft IEF.

> ***Finalize both content and formatting. The document will route much faster if reviewers are not finding admin issues to fix. Consult the template, along with the OPTEVFOR document writing guide for correct formatting.***

2. Run IEF database checksum. Fix all issues.
3. Route IEF for signature with CRM and RCRMs attached.

> ***Any changes made to the MBTD contents of the IEF must be reflected in the tool. Keep the tool up to date.***
> ***The division owns the documents. Route the CRMs with the document detailing what the division has done with each comment. 01B is not required to be included in the final IEF routing process.***

4. Update iBOSS PM (IEF signature date).
5. Notify stakeholders that the IEF is signed.
6. Ensure signed IEF is posted to iBOSS Documents.

> ***Be certain the workbooks are properly included in the Word file. Failure to properly attach them could result in loss of the signed versions of the workbooks. At minimum, it means anyone looking back at the signed IEF has to waste time searching for the files.***

7. Ensure all three RCRMs are posted to iBOSS Documents as supporting documents.
8. Offer lessons learned.

> ***The MBTD process has improved (including changes to this checklist) through the lessons learned by core teams. Provide any significant observations you have that may help future teams.***

9. Update the IEF as required.

> ***Consult OT&E manual chapter 4. Any time MBTD material is changed, the associated changes should be signed out, perhaps via IEF update. Draft the necessary products and complete the necessary meetings IAW prior steps of this checklist based on the material being changed.***

SUT / SoS Annex

Purpose: Clarify DODAF architecture use in detailing SoS, and provide additional detail on SUT/SoS development.

> ***The Operational Viewpoint (OV)-1 (and corresponding All Viewpoint (AV)-1), OV-2, 0V-3 OV-4, OV-5B, OV-6C and Systems Viewpoint (SV)-1, SV-2, SV-6, and SV-10C DODAF views used in conjunction with the ICF Integrated Capability Technical Baseline (ICTB) Level 3 (ICTB-3), where applicable, may assist with SUT/SoS determination. The list of views is not exhaustive and not all views may be available for use by the OTD/OTC (DoDAF architecture development varies between programs).***

1. The use of the DoDAF views is summarized here:
   a. OV-1 (High Level Operational Concept Graphic): Describes the mission, class of mission or scenario in graphic form. This view gives the user a high-level overview of how the system interacts in the battlespace (players, operations, etc.). This simple graphic may be useful to include with the SUT/SoS discussion in the IEF.
   b. AV-1 (Overview and Summary Information): The AV-1 provides executive-level summary information that describes the concepts contained in the pictorial representation of the OV-1.
   c. OV-2 (Operational Resource Flow Description): The primary purpose of the OV-2 is to define capability requirements within an operational context. The OV-2 depicts Operational Needlines that indicate a need to exchange resources (information, personnel, material, etc.).
   d. OV-3 (Operational Resource Flow Matrix): This view presents the resource flow information from the OV-2 into matrix form and also describes the attributes of those resource exchanges.
   e. OV-4 (Organizational Relationships Chart): The primary purpose of the OV-4 is to show the organizational structures and interactions for the system. This can be used to determine the roles of the humans operating and supporting the system (role-based OV-4) and the

stakeholders in terms of command structure at a point in time (actual OV-4). The OV-4 may be used to identify users within the SoS that will be critical to test execution.

f. OV-5B (Operational Activity Model): The OV-5B describes the operational activities that will be performed by the system within a mission or scenario. This view may be useful in determination of what tasks will be appropriate for the SUT to perform within a specific mission context.

g. OV-6C (Event-Trace Description): The OV-6C provides a time-ordered examination of the resource flows as a result of a particular scenario. They are also referred to as sequence diagrams, event scenarios or timing diagrams and allow the tracing of actions in a scenario or critical sequence of events. The OV-6C can be very complex but can be used in conjunction with the other OV views to determine the flow of mission execution.

h. SV-1 (Systems Interface Description): The SV-1 addresses the composition and interaction of systems, showing the performers/organizations and personnel types involved. It also shows capability integration and system integration information. This view is helpful in the determination of the external systems interacting with the SUT and may be helpful in determination of the SUT fielding configuration as well as which SoS elements should be represented in the test.

i. SV-2 (Systems Resource Flow Description): The SV-2 is complementary to the SV-1, providing a precise depiction of the system connections between the SUT and the SoS. This view shows all of the data flows within and external to the SUT. It is useful for further understanding SUT-SoS connections and determining which interfaces should be tested and associated SoS elements included in the test.

j. SV-6 (System Resource Flow Matrix): The SV-6 specifies the system data exchanges (resources) that flow across the system boundary. This view is typically very complex and should be used to supplement other views when determining SUT/SoS, methodology for testing interfaces and the SoS elements to be included or represented in the test.

k. SV-10C (Systems Event-Trace Description): The SV-10C provides, for unique scenarios or situations, the time-ordered interactions between functional resources. It has a corresponding textual description. This view shows the flow of mission tasks and interactions over time, which can be useful in the understanding of SUT and SoS as well as the tasks appropriate to the SUT for the current phase of test with the corresponding SoS elements to be represented in test. This view is complex and may not always be appropriate for use by the OTD/OTC.

2. Supporting Documents: Additional documents may be useful in defining SUT interaction with and integration into the SoS. OPNAVINST 3501 series Required Operational Capabilities/ Projected Operational Environment instructions define the mission areas applicable to the SUT. These missions form the basis for determination of the SoS aspects to be included in the test. Once these missions are understood, the architecture products discussed above provide the mission context and detailed interactions of the SUT within the SoS. In addition to the SUT DoDAF products, MTB and ICTB documents may provide additional context while relating the SUT and SoS relationship to TACSIT and Kill/Effects chains.

PMT View Annex

Purpose: Describe the PV-0, and provide guidance for initial creation of this PMT View.

PV-0 Tabs

The MBTES PV-0 excel file export has an "Overview" tab, containing a 'roll-up' graphical display of each COI, along with the 1st-level subtasks of each COI. This tab provides a snapshot of the entire MBTD in a single view. Next, it includes a tab for each Effectiveness COI, a tab containing all Suitability COIs, and separate Suitability COI tabs. Each of these contains all of the subtasks, measures (except orphaned measures), and conditions developed in the MBTD process for the associate COI(s). The tabs for the separate suitability COIs are not required to be kept (likely only useful for suitability COIs with complex task breakdowns). The first PV-0 tab, which must be inserted after MBTES export, is always "Program Information" in order to provide a quick reference for everyone using it. The IEF template contains a sample PV-0 from which this first tab can be copied, and inserted to new PV-0s.

PV-0 Export

There are two places to export a PV-0 from MBTES. Within the 'Other' menu, select the 'PMT View Shell' for the PV-0 containing the entire MBTD (for IEF purposes). The 'Test Plan PMT View Shell' selected under a specific document within the 'Test Plans' menu will give only the COIs, tasks, measures, and conditions applicable to the test (tasks excluded from the test plan are marked gray). This truncated PMT View Shell supports reporting under a single test document

PV-0 Production Guidelines

For ALL spreadsheets beyond the Program Information tab, the following guidelines apply:

1. The first row of the spreadsheet, and every second row thereafter will be a "Divider" row, and the row height will be established at 3.0 using Excel functions.
2. The first column of the spreadsheet, and every second column thereafter will be a "Divider" column, and column width will be set at 0.3 using Excel functions.

The combination of 1 and 2 above will establish a matrix where each data cell can be given a specific cell border color/weight, and it will remain visible. Without these "Divider" rows and columns, adjacent cells could not be provided an individual cell border due to overwrite.

3. Each tab contains a heading section. The heading is a standard format that will be output from MBTES. It will include the SUT name, Phase of Test, PMT View Variant (PV-0, PV-1B, PV-3, etc.), PMT View Variant Title (Complete Performance PMT View, Test Program Progress PMT View, etc.), as well as revision identifier for configuration control, and date of revision as further described below.
4. The revision number for a PV-0 will be a whole number, beginning with 0 for the original PV-0. The date will be the date of approval of the revision by the approval authority. The revision number for all other PMT View variants will be in the format "Revision 2.4", meaning that it is the fourth update using the format of PV-0 Revision 2.
5. All text within all PMT View variants will use Calibri font.
6. Font sizes will be as follows (if applicable, adjust outside MBTES):
    a. Titles and top Heading line: Font Size 14
    b. Tasks and Subtasks: Font Size 8-12, as necessary for readability
    c. Measures: Font Size 12
    d. Conditions: Font Size 8-12, as necessary for readability
7. Measure fonts will be further defined as follows:
    a. OTA Critical Measure (***BOLD ITALIC***)
    b. Program KPP, KSA, or CTP Measure (<u>UNDERLINED</u>)
    c. OTA Critical AND Program KPP, KSA or CTP (***<u>BOLD ITALIC UNDERLINED</u>***)
    d. Not an OTA Critical Measure, KPP, KSA, or CTP (Normal)
8. Task fonts will be further defined as follows:

a. OTA Critical Task (***BOLD ITALIC***). Note: On the Overview tab, 1st-level subtasks that have critical lower-level subtasks will also be marked bold italic.

b. Not an OTA Critical Task (Normal)

9. All Measures will be identified by the Measure ID, and will include a comment box containing the following information:
   a. Measure Type (Specified/Derived/Other)
   b. Measure ID (Example: M12)
   c. Measure (Example: Engagement Range)
   d. Criterion (Example: Median >= 20 nm)
10. All conditions will be identified by the Condition ID, and will include a comment box containing the following information:
   a. Condition ID (Example: C 1.3.2)
   b. Condition (Example: Visibility)
   c. Condition Type per DOE, if applicable (Continuous, Categorical, Discrete)
   d. Descriptors for the condition
11. There will be no colors used in cells within the PV-0 EXCEPT:
   a. Measures and Tasks in the test design but not in scope (i.e. not being evaluated) in the current OT phase will be filled with the color Gray. Any task levels below gray tasks are also gray.
   b. OTA RV measure cells and the Conditions that are RV Factors will be highlighted in Light Blue for identification.
12. Subtask cells that contain no task state "None". They are the same fill color as their immediate parent task.
13. Cell border and cell fill color legends will be included as part of the PV-0 to establish the program-specific standards. These legends will be used as the PV-0 becomes other views.
   a. The PV-1 legend must be included, and cannot be changed from standard.
   b. When the PV-0 is converted into another view, the appropriate legend will be kept, and the others deleted.
14. Truncated PV-0s must be marked as such. These include all tasks for any included COI (if the COI is not included in the test plan, the tab will not appear). Any task that is not in the test plan (not in scope, for whatever reason), there will be no measures or conditions displayed under the task, and the task will appear in gray.

Date	Change
Jun 12	Updated Fishbone Diagram on Page 1
Dec 12	Removed and updated all references to PINs — guidance has been incorporated into OTD manual
	Clarified A/B codes should be invited to TP2
	Added guidance for review and creation of Table 2.1 at IPR1
	Additional administrative corrections (typos, clarified wording)
Feb 13	Added guidance regarding OTD review of JCIDS documents to the very first step
Nov13	Updated guidance for WCB. Added clarity to measures, traceability and DRs. Regression note added. Note added on derived/other clarification with sponsor.
	DOE section adjusted for latest best practices and agreements with DOT&E
	Emphasis on Test methods as notes and other updates to IPR 2 checklist
Mar 16	Back cover POCs updated.
	MBTD Diagram updated. Checklist summary (after figure) improved.
	Process admin section added.
	Standard steps leading up to every MBTD review improved in all sections.
	Descriptions on how each MBTD review should run added in all sections.
	TP-1 checklist improved. In-scope SUT concept included. SoS discussion improved. Employment concept added.
	TP-2 checklist improved. Discussion of tasks and conditions upgraded.
	Document routing checklist removed.
	SUT/SoS Annex added (BP 24 incorporation).
Jul 16	IPR-1 checklist improved. Discussion of all steps (especially DRs) upgraded.
	Minor edits to TP-1, TP-2, E-IPR, and IPR-2 Checklists
	Migration of checklist annexes to electronic-only format on the Y-drive.
Jan 17	IPR-2 checklist improved. Discussion of all steps (especially vignettes) upgraded.
	Minor edits to TP-1, IPR-1, DWG (stakeholder contact expectations), and E
Jan 18	Expanded process admin section: database management processes, RCRM procedures, CRM guidance, MBTD vs. phase.
	Various review meeting directions throughout: meeting minutes, action items, lessons learned.
	Added guidance on Employment Concept section and COI wording.
	Cybersecurity guidance updated for new Cyber Survivability process: removal of most MBTD for cyber, added MCSM direction, Cyber CTF meeting attendance, and cyber overview inputs

Date	Change
Apr 18	CBTE awareness added to process admin section.
	WCB link to capabilities write-up in section 1. 01X invite to TP-1/2, IPR-1
	Removed “Unthresholded”. Replaced with “No Threshold”.
	RCRM posting to the Y-drive following IEF signature.
Apr 21	Front (title) and back (contacts) covers updated. “Cyber CTF” removed.
	01X removed from checklist
	RCRM rules clarified (all programs, draft, close-out)
	Cyber involvement clarified within steps leading to TP-1 / TP-2. Includes working engagements, and meeting attendance, early admin.
	N942, ONI, and/or independent DT rep invitation/attendance for various meetings clarified.
	PMT View & Section 2 approval at IPR-1 added for TIEFs stopping there.
	M&S resourcing for replication of live runs and RFR review at DWG added.
	Analysts review of DRs for RVs prior to DWG added.
	PMT View presentation at DWG and approval at E-IPR added
	ONI receipt of DWG read-aheads added.
	EIPR stakeholder attendance expanded. Important slides emphasized.
	CS test approach steps added to IPR-2 checklist and meeting.
	CS vignettes, M&S, limitations clarified. 01D added to IPR-2 attendance.
	Logistic Supportability annex added (electronically). Other mentions of electronic annexes clarified
	Analyst handbook mentions replaced with suitability handbook.
	PMT View annex added.
Jun 23	Working file moved to Word.
	Signature blocks and date blocks deleted. No longer tracked on paper.
	Meeting attendees blank spaces removed. No longer tracked on paper.
	MBTES guidance extensively reworked in process admin. MBTES function implications on other MBTD steps clarified throughout.
	TEPS removed. iBOSS added. IEF routing direction updated to match.
	12-step diagram updated to new version.
	SWP-related step identified with blue font. SWP direction added.
	Added discussion of MBTD for DT observation.
	LTD discussion added. CBTE discussion expanded.
	IEF input versus eventual TEMP negotiations.
	MBTD review meetings lead-up, execution, and follow-up improved.
	“In-scope” use in MTA/SWP/UON clarified.
	CS direction for COIs, tasks, conditions, measures, DRs, vignettes.
	Clarity on use of No Threshold, sponsor criterion approval, derived req’ts.
	1st-level subtask flexibility restriction due to issues search.

Date	Change
Jun 23	Custom descriptors for run matrix vs. all descriptors in directory.
	DR categorization optional. Expanded direction on DR drafting
	WCB removed throughout.
	M&S information needed for E-IPR brief.
	Vignette test methods clarified as an optional MBTD input.
	Vignette summary input to MBTES.
	Suitability vignette resourcing.
	DRTMs deleted from IPR-2 required products.

01B - Test Design and Analysis

01B - Mr. Dave Beyrodt ..... (457-6029)
01BB - CDR Rob Beauchamp ..... (457-6173)

CORE TEAM FACILITORS

Paul Coslick (lead CTF) ..... (457-6319)
Bill Lamping ..... (457-6075)
Josh Tribble ..... (457-6278)
x ..... (457-6139)

MBTES SUPPORT

Aaron Hommell ..... (457-6154)
Devon McGuire ..... (457-6154)

ANALYSTS

Rachel Robinson (lead analyst) ..... (457-6460)
Maya Copeland ..... (457-6345)
Josh Tribble ..... (457-6278)
Lee Stenson ..... (457-6393)

M&S and Targets

Travis Rauch ..... (457-6556)

CALL YOUR
CTF FIRST

"Happy Campers!"

Version 4.0
July 2023