DAU Webinar Pre-Event Questions (as of 10 April 23)

Central Question Themes:

1. Integration of Digital Engineering and Model-Based Systems Engineering (DE/MBSE) in T&E:
   - Questions: 8, 15, 16, 18
     - Main points: Challenges in integrating DE/MBSE, examples and lessons learned, leading groups, and applications in logistics T&E.
2. Security, Ownership, and Control:
   - Questions: 6, 7
     - Main points: Multi-level security, data sharing, ownership and control issues, roles of DoD and defense contractors.
3. Continuum and Role of Operational Test & Evaluation (OT&E):
   - Questions: 9, 10, 11, 14
     - Main points: Finalizing requirements, risks, early engagement of OT, guidance for funding and scheduling, and fitting OT&E in the continuum.
4. Confidence and Validation in DE Results:
   - Questions: 12, 13
     - Main points: Gaining confidence before validation data is available, handling software and cyber weapon systems.
5. Changing T&E Approaches and Funding:
   - Questions: 14, 17
     - Main points: Determining testing needs, guidance for funding and schedule allocation, OSD's role in facilitating the T&E continuum.
6. Lessons Learned and Examples from T&E Community:
   - Questions: 15, 18

TEaaC Webinar Questions

Question 1. How can data-centric approaches help catalyze Testing as a continuum? (GOVT)?

Answer 1.

By being data-centric we can treat all data as a pool of information for use at any time. To be data-centric we must focus not just on the data itself but its version history and its metadata. By doing so we make data useful for future users, not just those creating the data initially. This includes the ability to analyze existing data to determine which new data will support decision making. The T&E community can then consider data to be its major product rather than documents; a data-centric approach. LWE comment: model-centric approaches (a type of data-centric approach) in particular promote TEAAC because of the incredibly powerful ability to "test" a well-designed model through simulation and execution, then iterate on the model and retest several times ahead of a build cycle.

Sharing data is a key enabler, approaches that assure Government rights in data coming from Contractor Test and Development Test are necessary to have a smooth continuum.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

Question 2. How does T&E of AI affect the T&E continuum (perhaps drive more need for it due to iterative nature of AI)? (Industry).

Answer 2.

For ML based AI it becomes even more critical to have a continuum that spans development/DT (where the model is trained and tested) to the operational test (where AI enabled system is validated). This cycle will need to iterate, even after the AI enabled system is deployed, so we can adjust the model and redeploy it. Having infrastructure to enable this process (pipeline) is critical to the speed of execution of this process while maintaining evidence of the AIES training and testing. This infrastructure must also serve to protect the model, the data, the model parameters – to prevent the AIES from being compromised (e.g., data poisoning). The T&E continuum even goes back to the beginning of the process when prototyping and requirements definition occurs. T&E can engage at these early stages on a smaller scale to advise of the T&E of AI practices so the DOD can be informed of the complexities of a proposed system before the ink is dry on the requirements. All of this assumes that the DOD assures access and rights to the AIES sufficiently enough to validate it through the contracts and contract deliverables. The use cases for black boxed AI enabled systems must be critically unique enough to warrant the risk and higher costs of testing and evaluating it for its intended operational environment.

AI behavior is discontinuous, which complicates T&E significantly. Sharing test results across the continuum makes more results and up-to-date results available. While this improves the prospects for T&E of AI it's far from a complete solution to the issue.

Question 3. How to effectively use DE/MBSE tools for T&E when accommodating multiple customers with different "models" (e.g., aircraft SPO's, weapon SPO's, Multiple Orgs involved with Airworthiness, etc.)? (Air Force)

Answer 3.

There are techniques for model federation allowing disparate models to communicate without dictating explicit tools or structures on each independent creator. For example, SAIC has a method using "micromodels" that act as adapters between a set of models (NDIA 2022, Swickline). LWE comment: ultimately, we may need a forcing function to promote the use of standards. For instance, if our department only has a single software tool to evaluate MBSE models, programs must be required to deliver models that can be read and evaluated using that tool.

Sharing models won't be an automatic process, as shown in the long history of sharing simulations for testing or training. In selecting tools and templates it will be important to consider how widely applicable they are and what features they support to reduce the cost of model reuse.

Since we are in the nascent stages of digital transformation, we have an opportunity to take advantage of a unique inflection point in acquisition history. With everyone learning to use digital tools, the entire enterprise is looking to leverage and use models that are already

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

developed as opposed to building their own. If we can take advantage of this opportunity and establish collaboration across all of the pathfinder modelers (govt & industry), we can bring a unity across the DoD enterprise going forward that previously was not possible.

Question 4. How will programs determine how much and when testing is needed? What guidance will they have to determine the appropriate funding and schedule to allocate to continuous testing? (Air Force)?

Answer 4.

Statistical methods will become more ubiquitous. As digital engineering makes more data and quantitative models available, statistical analyses can be run to quantify uncertainties. This is an extension of the work of STAT COE into the digital engineering future. Our task will then be to determine our risk tolerance and statistical analysis can tell us how much more testing and of which types are most likely to reduce the risk to the desired level. LWE comment: the beauty of test-as-a-continuum is that "continuous test" funding as a separate line item is not necessarily required. With TEAAC, time dedicated to up-front design development reduces the need for "test" execution later down the line. If we leverage MBSE to create a simulatable, executable model, for instance, we learn a lot and improve the design multiple times before any hardware is built or software code is written. The genius of MBSE is that, leveraged appropriately (and especially once the new SysML v2 standard is in place) is that we can even "test a negative" and build "unhackable" systems using MBSE models (reference HACMS, HARDENS).

The same way they do today. The T&E continuum should be reflected in their test strategy and their TEMPs.

Question 5. Given the current MIL STD structured process and requirement, how do you propose we incorporate the early Vendor testing of commercial or DoD standards that currently occurs prior to DoD acq consideration?

Answer 5.

Mathematically we can use statistical and information based methods to treat all data collected in the past as part of a serial testing campaign. If the conditions under which the data was collected are well understood we can quantify our uncertainty when including it in an ongoing test continuum. Contractually we need to request access to the data and define our usage/purpose. In general, the Government should consider a test program or experimentation/VV&A plans that uses a combination of commercial and Govt testing to assess performance. LWE comment: once models become standard, it makes sense for DoD to request model artifacts together with certification evidence to complement the model. Full adaptation of standard MBSE practices can mean that vendors accomplish significant useful T&E ahead of DoD acquisition initiation as a way of demonstrating the capability of the product that they want DoD to acquire.

Sharing test results from prior work provides a mechanism to incorporate a vendor's own testing into the continuum. Acquisition processes should value these disclosures of results by recognizing the reduction in duplicative T&E that they enable.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

Question 6. How multi-level security is used for digital T&E information?

Answer 6.

There are many methods being developed so that we do not need to perform all evaluation at system high. Meta-data/data centric techniques will almost certainly play a large role in these solutions. Role based access techniques can be applied to data if it is tagged appropriately. Developing these tagging schemes is ongoing.

Sharing data always involves security considerations. T&E throughout the continuum must be sensitive to these security requirements and implement appropriate security controls. There multi-level solutions are available, they can reduce data sharing costs.

Question 7. Is the DoD developing these models in-house, or are they just being developed by Defense Contractors? (ownership and control issues)?

Answer 7.

Both. As an example, the DOD might create, own, maintain, and control mission and architecture reference models. These models can then be supplied to contractors who extend the models as part of a proposal and then further as a system design and test campaign. By segregating the role between DOD and contractor along abstraction and fidelity lines each set of models can inform the other while avoiding (some) ownership and control issues. This will not be the answer in all cases but can be effective in many if not most.

The development of T&E strategy, concepts, and TEMPs in a program is a collaboration between government and contractors. Both sides benefit from a shared model used for analysis, design, and testing. Contractors that are unwilling to share models that enable this benefit will cost their programs due to duplication of effort, and these higher costs might lead to such contractors not being selected for future programs.

Question 8. What do you envision as the biggest challenge integrating T&E/DE& MBSE?

Answer 8.

Reuse of models and data. This problem comes in three parts:

1. They are hard to find. Knowing which models and data exist and where they can be used (models and data are only valid when used under specific conditions) is extremely hard. We need a system for searching (this may or may not involve centralization) that includes definition of the validity criteria.
2. Once you find a model or data set how do you know it is the latest? We need a system for version control.
3. They are rarely plug and play. Manual effort is almost always needed to make them useful in a new environment. We need a system for defining interfaces.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

4. However, without reuse of models and data we will not achieve anywhere near the productivity gains we hope from digital engineering.

Some of the principal challenges are:

1. Teaching engineers to use MBSE as intended to leverage its full benefits (as opposed to using MBSE as an expensive tool to create PowerPoint slides). MBSE tools can (and should) be used to simulate and test systems ahead of system development, which is only possible with appropriate use of the MBSE tools.
2. Formalizing everything from MBSE language to requirements to MBSE system design. The tools will be incredibly powerful if used the right way; however, if wholescale adoption begins without understanding how best to use them, the real value wont be clear and the momentum needed to create real change wont appear.

The opportunity for a fresh start, a program that's "born digital", will occur infrequently. For many years we will be testing with a hybrid of programs with sound DE&MBSE and programs built with legacy processes. An overnight switch in approach isn't reasonable, so handling the shifting balance during this hybrid period is the major challenge.

Changing the current culture with everyone reinventing the wheel as opposed to working together in a collaborative environment is needed. On overarching change management strategy needs to be put in place.

Question 9. Since Testing is a means by which to verify requirements, in this environment when are requirements to be finalized and what is the risk if they are not?

Answer 9.

By Department regulation and law, operational requirements for major capability as well as middle tier of acquisition programs are set before substantial funding and programmatic activity can commence:

- The Materiel Development Decision (MDD) is the mandatory entry point into the major capability acquisition (MCA) process and is informed by a validated requirements document (e.g., an initial capabilities document (ICD) or equivalent).[1]
- Middle Tier of Acquisition (MTA) programs will not be subject to the guidance in Chairman of the Joint Chiefs of Staff Instruction 5123.01H and DoD Directive 5000.01. Each DoD Component will develop streamlined process that results in a succinct requirement document no later than 6 months from the time the operational needs process is initiated. Approval authorities for each capability requirement will be delegated to a level that promotes rapid action.[2]

Detailed performance specifications derived in part from operational requirements are defined in contracts for development and production at various times throughout a program as it proceeds.

[1] DoDI 5000.85 Major Capability Acquisition Procedures, November 2021.
[2] DoDI 5000.80, Operation of the Middle Tier of Acquisition (MTA), December 30, 2019.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

Under testing as a continuum, information will accrue continually throughout a program indicating whether operational requirements and contract specifications will be achieved. The risk is not in setting those requirements, but in failing to understand as early as possible what shortfalls in achieving them might occur, what the implications of the shortfalls could be for military capability, and whether and how to modify the program to correct the shortfalls if their implications for delivered capability could be significant. In contrast with the current practice of episodic testing conducted during a limited number of discrete events, conducting testing as a continuum should provide earlier and more comprehensive understanding of risks to satisfying requirements and their implications, as well as to the actions, if any, that should be taken.

When the objective is to have agile programs that respond to a changing battlespace, requirements can't be finalized. However, contractual baselines will refer to a baseline set of requirements. As the program evolves, and changes in the requirements evolve, T&E must flow continuously with those evolutions. It's important to capture and use data which defines which requirements were being tested in any event, so that the testing can be reconsidered in the face of shifted requirements.

DTE&A acknowledges that the MBE environment will better support introduction of requirements updates post-JROC requirements validation. I believe we want this flexibility to ensure the capability being delivered is designed and, at a minimum, assessed against the evolving threat/environment. This does lead us to the discussion regarding characterizing threat "space" (critical attributes/dynamics) – vs. specific instantiations of threats – so programs aren't chasing very specific instantiations that currently take very long to develop.

DTE&A also acknowledges the idea of moving to a requirements "model" – that starts to think of requirements in a behavior context vs. discrete KPPs/KSAs.

To be clear, all testing does not validate requirements. Test as a Continuum (TEaaC) provides better insight/structure to characterizing risk associated with requirements validation. And yes, TEaaC provides flexibility to accommodate requirements updates mid-acquisition – at least the ability to better assess risk of "doing nothing".

Question 10. Actionable ideas on how OT can engage early and help shape while remaining in our statutory independent role.

Answer 10.

By law, DOT&E can only provide advice on matters involving developmental testing (DT) (Title 10 USC Section 139). However, there is nothing prohibiting DOT&E from engaging with DTE&A and program offices early, for example beginning at Milestone A, to offer advice on how developmental testing, including modeling and simulation, can be conducted to provide information relevant to determining operational suitability and effectiveness, including information needed to verify, validate, and accredit modeling and simulation (M&S). This engagement could help reduce the time and resources needed to conduct dedicated OT. The use of model-based systems engineering (MBSE) can provide a means for continually reviewing plans for and the results of testing and M&S as it is conducted, and for documenting the results

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

relevant to evaluating operational suitability and effectiveness, as well as the effects of those results on subsequent testing, including both DT and OT.

OT independence prevents DOT&E from providing more than advice for DT earlier testing. However, that advice can be more strategic in early phases of a program. Early engagement can produce test results in CT and DT which are more useful in OT thanks to data sharing throughout the continuum. For example, scenarios for testing throughout the continuum could be common. Where common scenarios are exercised in DT, the program has more confidence that surprising results in OT will be avoided. OT assessments can be informed by more relevant DT results, increasing the set of relevant data points available to the OTA.

If we think about data curation early (in the context of the IDSK), then OT can fully leverage relevant data for OT requirements. Earlier capability-focused testing only helps OT.

Question 11. How does operational test & evaluation (OT&E) fit in the continuum, or does the concept mostly apply to developmental testing?

Answer 11.

See the answer to the question immediately above. Testing as a continuum applies both to DT and OT. In conjunction with the use of MBSE, testing as a continuum should provide a means for gaining and documenting information throughout a program on its likely operational effectiveness and suitability, providing effective mutual engagement is conducted by both the program offices and DOT&E, assisted by DTE&A.

See Q10, the continuum should be a smooth connection between tests. As processes become more cyclical, the distinctions between DT and OT will focus on test objectives rather than dates on a program schedule.

Question 12. How is confidence gained in DE results before validation data becomes available?

Answer 12.

It is rare that new DE articles are based on completely new concepts. Preliminary validations can, and should, be run on axiomatic cases and any previous data that already exists. Much of the time we are extending the capability of previous DE results. These extensions should at least in part be able replicate previous results.

This does point to the importance of reuse and interoperability of both data and models. We need data standards that enable this reuse as well as methods for identifying the context in which models are validated. An extension of a model can then be verified and validated first with existing data in the previous context. This gives us confidence that the new model works as expected. There remains the risk that the new model does not work in the new extended context. It will be up to the author and leadership how much T&E needs to be done to validate the extended context.

The engineering process has always had an incremental approach to confidence. As designs are refined, high risk components are prototyped to improve confidence. The TRL system has been

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

developed to understand confidence on a technology basis. DE results, like all early engineering designs, should be subject to skepticism until they are supported by validation data.

Confidence will be gained over time as MBSE models and M&S tools mature and learn based on actual validation & accreditation data.

Question 13. How does this address weapon systems that are entirely software (cyber weapons)?

Answer 13.

Pure software development has already, for the most part, achieved T&E as a Continuum. Modern Agile methods are a continuous process of identifying new requirements.

Software is more complex than hardware, which results in correspondingly more skepticism. The offsetting benefit of software is that is can be modified and refactored much more quickly that hardware. Systems that are entirely software are more amenable to test automation and routine regression testing than systems with expendable hardware components. The result is more test data throughout the continuum, and the need for analytics to assess the test data.

TEaaC fully supports current agile SW/DSO development processes – in fact we are trying to extend the agile SW framework to entire weapons systems. And by introducing cyber continually, it helps focus on cyber resilience vs. controls (RMF).

Question 14. How will programs determine how much and when testing is needed? What guidance will they have to determine the appropriate funding and schedule to allocate to continuous testing?

Answer 14.

The same way they do today. The T&E continuum should be reflected in their test strategy, program master schedules, and their TEMPs.

Capabilities Based Test and Evaluation (CBTE) evaluates the effects chain in a mission context. Benefits of CBTE includes reduced development time and cost through early discovery of mission related technical issues; enhanced Systems of System testing, specifically addressed via test design and planning; and enables closing the loop with warfare analysis – answers the big "so what" question on whether the delivered system meets fleet needs.

Question 15. What are some lessons learned previously from anyone in the T&E community that has been through a program that has been tasked to fully utilize digital engineering?

Answer 15.

Reuse is a key factor in achieving gains in productivity. Digital engineering can have higher cost up front with the expectation of a long-term return on investment. If the digital engineering products cannot be reused, then the long-term return on investment plummets. TRMC is working to ease this burden on programs and the T&E community through tools like TENA that allows

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

software and hardware on test ranges to interoperate through middleware rather than requiring rework.

Fully digital programs are few, but model-based T&E has been used in parts of many programs. Among the lesson learned are:

a) Access to models and model data must be clear in contracts before work begins so that contractors are aware of the government's need to share this information.
b) Models can be used for pre-flight prediction of test events, both to assure that the desired data can be collected and to facilitate the use of test events in model validation.
c) Large MBSE models take significant time to understand, and government personnel will need time to get up to speed on them.

Initial pathfinders are in development.

Question 16. What group is leading how T&E should use DE in each service? With the system of components being measured in the digital environment, how do we rapidly grade/evaluate in operational test to show true capability collectively and independently?

Answer 16.

The Navy is leading efforts out of the DON CHENG, Army has both ATEC and AFC, the Air Force has AFTC, and Space Force has STARCOM.

Question 17. The locked in Acquisition funding for T&E does not include iterative cycles of Develop/Test/Fix/Verify much less the "continuum"; how can OSD help with this?

Answer 17.

DTE&A has committed to engaging and working with program offices to help them implement testing as a continuum and tangibly demonstrate its benefits. As examples of those benefits accrue, DTE&A is similarly committed to sharing them throughout the acquisition community to build widespread acceptance of the need to conduct testing as a continuum and commit the associated resources. DTE&A has developed detailed explanations of the principles of testing as a continuum and its benefits that it is sharing with the acquisition community via briefings to OSD and Service leadership, participation in meetings, symposia, and webinars, and media publications. (See, for example, the upcoming ITEA Journal article "Test and Evaluation as a Continuum," by Mr. Christopher Collins and Mr. Kenneth Senechal.)

Building in validation via the IDSK will help better predict funding requirements for programs (and hopefully introduce stability in funding profiles). The continuum becomes embedded as how we cost out an overall engineering (ME, SE, T&E) effort.

Question 18. Please share example(s) of how Digital Engineering can support Logistics T&E.

Answer 18.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.

This depends on the intent of "Logistics." If you mean supply chain management, Digital Engineering can massively increase the ease of traceability of components. This applies both during the design phase when components may need to be tested for batch quality and in supply support when tracking spares or modeling alternatives for replacement parts. If you mean product support, Digital Engineering using model-based engineering can streamline –ility testing to ensure that the final product is supportable in the field. Predictive analysis either managed through digital techniques or modeled in digital simulations can optimize the supportability to program needs.

Question 19. What is the role in of the user or human system integration (HSI) labs in this process?

Answer 19.

The role of user or human systems integration (HSI) labs in the DTE&A's Test as a Continuum process is to ensure that human factors are effectively considered throughout the system design, development, testing, and evaluation stages. HSI labs play a crucial role in understanding and optimizing the interaction between humans and the systems they use or operate throughout the system's life cycle, ultimately leading to more effective, efficient, and user-friendly systems.

DTE&A question responses are opinions based on understanding of current best practices, policy, and guidance, and do not imply or constitute changes to official published policy.