The ALSP Joint Training Confederation: A Case Study of Federation Testing

Ernest H. Page
William E. Babineau

The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102
{epage,babineau}@mitre.org

Keywords:
Aggregate Level Simulation Protocol, testing , verification, validation and accreditation.

ABSTRACT: *The Aggregate Level Simulation Protocol (ALSP) resulted from a U.S. Defense Advanced Research Projects Agency (DARPA) effort to identify mechanisms suitable to facilitate the integration and interoperation of existing (so-called "legacy") constructive training simulations. The Joint Training Confederation (JTC) – the primary application of ALSP – has evolved from two models in 1992 to twelve in 1997 and supports several large-scale command post exercises (CPXs) each year, including the annual Ulchi Focus Lens, Prairie Warrior and Unified Endeavor exercises. Along with the Distributed Interactive Simulation (DIS) protocol efforts, the ALSP JTC serves as a key progenitor of the High Level Architecture (HLA). This paper highlights the testing approach used within the JTC and places it within the broader verification, validation and accreditation (VV&A) framework. The challenges presented by federation testing – some fundamental, others practical and programmatic – are discussed.*

1. Introduction[1]

The Aggregate Level Simulation Protocol (ALSP) began as a Defense Advanced Research Projects Agency (DARPA) funded effort in the spring of 1992 to investigate the feasibility of interconnecting existing constructive simulations to form a cohesive exercise environment [1]. Subsequent to a successful demonstration of these capabilities in the fall of 1992, ALSP was fielded as the infrastructure for a premier joint exercise environment. Since 1992 this environment, the Joint Training Confederation (JTC), has grown from two primary simulations to twelve and has supported several major exercises each year. For 1997 these include: Yama Sakura and Ulchi Focus Lens, hosted by the Korean Battle Simulation Center, Seoul, Korea; Prairie Warrior, hosted by the National Simulation Center, Ft. Leavenworth, Kansas; and Unified Endeavor, hosted by the Joint Training and Analysis Simulation Center, Suffolk, Virginia. For a more complete list of JTC-supported exercises refer to Page et al. [2].

Many aspects of ALSP and the JTC have been necessarily experimental in their nature. Although the technologies leveraged are for the most part mature, the JTC arena provides a novel context for their application. As a result, the processes and methodologies guiding the JTC have followed an evolutionary path. Many problems encountered have admitted solution, but many challenges remain.

The techniques and methodologies for both systems testing, and more broadly, verification, validation and accreditation (VV&A) are examples of this phenomenon. The techniques for VV&A of discrete event simulations are well studied and widely utilized (e.g., see [3]). Likewise, software testing methodologies are well known (e.g., see [4]) and are the basis of a variety of standards (e.g., see [5]). However, the application of these techniques and methodologies within the JTC is not straightforward — a variety of technical and programmatic barriers to their application exist. The size and scope of the JTC is increasing to keep pace with the ever-expanding training requirements, yet the time and dollars allocated to testing promises to remain fixed. In order to maintain JTC acceptability, which is a function of sufficient validity, the process of engendering and quantifying JTC credibility should strive to be both flexible and optimal.

This paper describes the latest efforts in refining the testing and VV&A processes within the JTC. Section 2 provides a review of the JTC development process, focusing on its integration and test aspects. Consideration of VV&A in this environment is given in Section 3. Section 4 identifies some current initiatives to improve cost-effectiveness of testing and VV&A in the JTC. Some conclusions are offered in Section 5.

2. The JTC Development Process

Page et al. [2] describe the VV&A practices for the ALSP JTC focusing on the testing process and the technical and operational challenges introduced through the Advanced Distributed Simulation (ADS) arena. A more general description of the JTC development process is given in [6].

The JTC development cycle comprises the transformation of user requirements and existing training simulations into a user-accepted system (a "confederation" of simulations). A portrayal of the overall JTC system project cycle is presented in Figure 1, as viewed using a system engineering "vee" model adapted from [7].

A single iteration of the system life cycle begins with user needs (requirements) at the upper left and concludes with a user-accepted system at the upper right. Descending the left side of the "vee" reflects decomposition, definition and development of the system. Ascending the right side of the "vee" represents system integration and testing. Each level of the model provides a correspondence between: (1) defined requirements and specifications, and (2) integration and test events. In order to promote rational system development, requirements and specifications must not be defined in a way that cannot be tested or verified. Similarly, integration and test events must have corresponding requirements and/or specifications to verify and/or validate.

[1] This paper is a republication (with minor updates) of: Tufarolo, J.A. and Page, E.H (1996). "Evolving the VV&A Process for the ALSP Joint Training Confederation," In: *Proceedings of the 1996 Winter Simulation Conference*, pp. 952-958, Coronado, CA, 8-11 December.

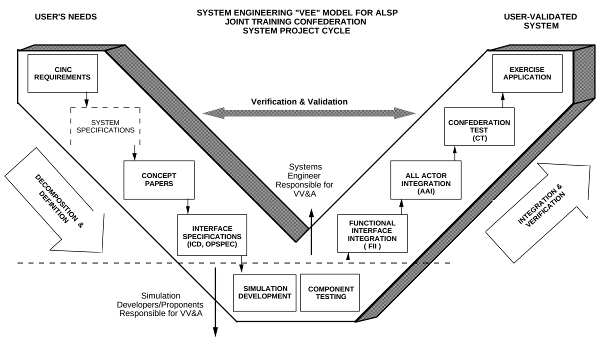

Figure 1: Vee Model

2.1 Overview

For the JTC, user needs are captured in the form of CINC/Service requirements. These high-level requirements are defined, agreed upon and prioritized for implementation. Based upon available funding (and the perceived availability of future funding) a collection of high-level designs, or concepts, are proposed. These concepts are evaluated against a variety of criteria and, if accepted, are refined into a set of model specifications and implemented within the various simulations comprising the JTC. Subsequently, the system of simulations is integrated, tested and evaluated. The "final" system consists of an accepted confederation (new versions of individual simulation software along with the ALSP Infrastructure Software) and associated documentation. It is this final system that is intended for deployment and subsequent accreditation and use at exercise locations.

2.2 Design and specification

From CINC/Service requirements, sets of *concept papers* are written. A concept paper describes a new or modified confederation functionality, identifies the relevant requirements, and suggests methods for achieving the desired capability. Concept papers outline the applicability of the concept to the JTC, impact on existing simulations, and estimated level of effort for affected simulations. These papers also provide a forum for bringing new simulations into the JTC.

Concepts are formulated and evaluated through the Interface Working Group (IWG). The interested reader should refer to Fischer [8] for a complete description of the organization and management of the Joint Training Confederation.

Subsequent to concept approval (by the ALSP Review Panel, typically upon recommendation of the JTC Systems Engineer) detailed designs are formulated. These designs span several documents: *Interface Control Documents* describe message syntax, content, and sequencing requirements. *Actor Implementation Documents* detail the interface implementation within a specific simulation ("actor" in the parlance of ALSP). The *Operational Specification* provides the model definition in terms of objects, attributes, interactions and parameters as well as simulation-object ownership and behavioral "scenario descriptions." The Operational Specification serves as the primary communicative model representation. Finally, the *Technical Specification* provides a technical reference manual for the ALSP protocols.

Currently the dashed item in Figure 1 labeled "System Specifications" does not exist as a singular entity for the JTC. Individual simulations maintain system specifications describing private objects and behaviors within their respective simulations. Ideally, these specifications might be consolidated to form a single system description describing objects and behaviors — both private and public — for the JTC *as one system*.

2.3 Development and implementation

The simulation developers use the design documents to generate modified versions of their simulation software. In accordance with the management structure originally installed for the JTC (see [8]) and pursuant DoDI 5000.61 [9], each Service has the responsibility of defining its own V&V process and naming its own V&V agent(s) for any simulation that participates within a "federation of models and simulations." The Systems Engineer assumes responsibility of the JTC as a whole, but has no dominion over the processes or agents appointed by the individual Services. The JTC is a democratic, cooperative organization. Participation within the joint training arena is ostensibly voluntary. Therefore, with respect to the JTC development process and the V&V of the system as a whole, individual simulation modification must sometimes be treated as a black-box activity. Methods and guidance for code-level testing and regression testing can be suggested, but no mechanism for enforcement exists. Similarly, the products resulting from V&V of the constituent simulations can be requested, but not demanded. Where these products are made available, they are utilized within the process described here.

2.4 Integration and test

Integration and testing occurs in four stages: (1) component, (2) Functional Interface Integration (FII), (3) All-Actor Integration (AAI), and (4) Confederation Test (CT). Test plans are developed in five areas: (1) component test, (2) technical test, (3) functional test, (4) operational test, and (5) load test. Currently, component tests for individual simulations are not developed nor managed by the JTC development process. However, the JTC Systems Engineer may assist simulation developers when conducting in-house component tests. The remaining test plans are formally constructed, evaluated, and executed within the JTC development process.

- *Technical testing* deals with matters of conformance and compliance with the ALSP protocols. The ability of an actor to connect to the ALSP Infrastructure Software, indicate its operating parameters, advance simulation time, and perform checkpointing are examples of items covered under technical testing.
- *Functional testing* is designed to verify that the JTC specification has been correctly implemented.
- *Operational testing* is performed to evaluate JTC behavior against the objectives of the training community.
- *Load testing* is performed to evaluate the performance of the JTC. To satisfy training objectives, the JTC must be capable of keeping pace with real-time over the period of a training exercise. Specifically, if an exercise begins at 1200 on a given date and ends at 1100 three weeks later (according to the wallclock and calendar), the "JTC time" at end-Ex should be very near 1100 (on start-Ex plus three weeks). Note however that this does not imply that the JTC needs to meet "hard real-time" performance measures. The training audience does not interact *directly* with the JTC simulations during an exercise; interaction is through actual "go-to-war" systems.[2] Deviations ranging from several minutes to several hours from wallclock time can be tolerated (in specific situations and for limited duration) as long as output to these systems can be otherwise mediated, e.g. exercise *scripting*.

Technical and functional test plans support both the All-Actor Integration and Confederation Test. In addition to functional and technical test plans, operational test and load test plans are developed to guide testing at the Confederation Test.

Developing *enumerations* is a critical item for the JTC. Enumerations are lists of valid attribute values agreed upon by the ALSP Interface Working Group for use within the JTC. The Operational Specification document captures these approved values, and is used as a reference for developers and exercise managers.

As indicated in Section 2.3, each Service defines the V&V processes to be used by, and designates the V&V agent(s) for, simulations that participate in federations such as the JTC. Therefore, the structure and execution of (standalone) component tests is outside the dominion of the JTC V&V agent. Interface-level component tests *are* subject to oversight within the JTC development process and the Systems Engineer assists simulation developers when conducting these in-house interface-level component tests.

Functional Interface Integration tests are conducted *as needed* between subsets of the JTC simulations. These integration tests are analogous to those used in modular program development and facilitate federation composition. Typically these subsets are formed according to functional interface – hence the moniker for the test.[3] For example, if substantive changes occur to the air-to-ground combat interface between JTC development cycles, the air model(s) and the ground model(s) typically will participate in a Functional Interface Integration prior to the all-actor test events (AAI and CT).

The All-Actor Integration and Confederation Test are the primary test events for the JTC. These events are typically two weeks in duration and require that all simulations be convened in a single location for testing.[4] The All-Actor Integration serves as the

[2] Where interoperability between the JTC simulations and extant go-to-war systems is lacking, a layer of intervening personnel is utilized.

[3] The principal organizing framework for the JTC specification, e.g. air-to-ship combat, ground-to-ground combat, sustainment, and so forth..

[4] To some, the application of centralized testing seems contrary to the notion of *distributed* simulation. If the operating environment is distributed, why shouldn't the testing environment be likewise distributed? In fact, distributed testing *does* occur within the JTC (often FIIs are distributed). However, centralized testing has several useful consequences. Obviously, the coordination of testing activities is simpler when all interested parties are co-located. A less obvious benefit, but a highly valuable one, is the camaraderie and sense of community that only results from close quarters interaction – a similar argument may be made for technical conferences and workshops. It is also worth noting that the primary benefit of distribution in the JTC is the distribution of the training audience. In the majority of JTC exercises "technical control" of the participating simulations is generally centralized within the simulation center hosting the exercise.

"developers test" and focuses on evaluation of JTC functionality with respect to the specification (functional testing). The Confederation Test serves as the "user's acceptance test" and focuses on operational realism (operational testing and load testing). During the Confederation Test the behavior of the JTC is evaluated with respect to the relevant training objectives.

2.4.1 Structure and execution of the Confederation Test

The structure of the CT merits some discussion. The Executive Agent acts as CT Test Director. The Systems Engineer assumes the overall responsibilities for test coordination and serves as director of both the technical test and the load test. The position of Functional/Operational Test Director is held by a Subject Matter Expert (SME) that represents the user community. Functional/operational testing is separated by functional area. A separate *test cell* is allocated to each functional area and the test plan for that area is executed under the direction of a Test Cell Coordinator.[5]

Two confederations are used to support testing, a *production* confederation and a *test* confederation. Each confederation is assigned a coordinator – usually a representative of the user community – who supervises and controls the technical aspects of the confederation activities, e.g. joining, resigning, scheduling and effecting confederation-wide saves, and so forth. Execution of the CT test plans occurs on the production confederation. When a test fails, a Problem Report (PR) is filed by the Test Cell Coordinator and submitted to a problem tracking system which is generally under the purview of the Systems Engineer. Problem tracking personnel determine the source of the failure and adjudicate the disposition of the PR. If software changes must be made to permit test passage, the software is modified and the test evaluated on the test confederation. Upon successful passage of the test (and any necessary regression testing as determined by the Test Confederation Coordinator) the modified software is "rolled" into production and the test is re-executed there. A test is only considered passed when it passes in the production confederation.[6]

The AAI and the CT provide the opportunity to rehearse this deployment.

[5] For the 1997 CT, ten test cells were utilized: air-air, air-ground, surface-air, ground-ground, ship/ground, ship/air, air/sea lift, electronic warfare, intelligence, and sustainment.

[6] The twin confederation paradigm used during the CT is also often used within JTC exercises

2.4.2 Validation techniques applied within the Confederation Test

Of the 13 subjective validation techniques identified by Balci [10] *event validation*, *face validation*, *sensitivity analysis*, and *submodel testing* are the primary techniques used to evaluate the JTC. *Schellenberger's criteria* also apply to the JTC approach although the formal delineation of model assumptions is missing in the current practice.[7] In some sense, the previous year's exercises serve as *field tests* for the Confederation under test. Year-to-year changes in model functionality are rarely "drastic" and are usually localized within the overall structure of the simulation code. As a result, some level of general confidence in a model accrues over time. But demonstrated successes in actual exercises not only provide a level of confidence in the model itself, but perhaps just as importantly, they engender a level of confidence in the model developers and model operators. Generally speaking, the probability of a successful exercise is *higher* given a bad model that has competent, experienced model operators, and skilled programmers who are intimately familiar with the code, as opposed to a good model with inexperienced operators and "newbie" programmers. The reasoning here is simple: no model is completely impervious to "attack" from bad or unexpected input. And at exercises, *bad input happens*. Simulations crash. In these situations the success of the exercise hinges on the ability of the operator(s) to recover from the crash, or the ability of the programming team to quickly develop and install the necessary patch. Some other issues affecting exercise success are discussed in Section 3.

2.5 Deployment

Confederation deployment includes delivery and employment of new simulation and ALSP Infrastructure Software at user sites. The JTC Systems Engineer provides support during *first-use* — the first major exercise to operate a newly accepted JTC. Associated with the JTC deployment, exercise centers typically conduct additional on-site testing of the JTC using their specific environment and data. This supplementary testing, to include Systems Engineer support if requested, is funded by the exercise sites. It is the purview of each exercise site to accredit the JTC for use within its exercise(s).

[7] Refer to [11] for a comprehensive survey of the verification and validation techniques for discrete event simulation.

3. Evaluation

An evaluation of the VV&A process within the Joint Training Confederation must begin with a recognition that no singular VV&A Agent role is defined for the JTC [8]. Such an agent, if singularly defined, would possess the omnipotence to establish and oversee the VV&A within the individual simulations as well as the VV&A of the confederation as a whole. In the current JTC management structure each Service (or model proponent) has the responsibility of VV&A for its simulation. The ALSP Systems Engineer assumes responsibility for VV&A of the system as a whole (see the demarcation of VV&A responsibilities in Figure 1). As Balci [11] indicates, a collection of valid sub-components does not necessarily result in a valid system when these components are integrated. Therefore, validation of the whole requires validation beyond that of the parts. However, the limited authority of the Systems Engineer in this role implies that some of the testing undertaken to support the system-level validation activity is necessarily functional or "black-box."

3.1 The vanilla approach to systems testing

Adopting the term used by Harel [12] we describe the *vanilla approach* to systems testing (testing "in the large" as delineated in [4]) as being comprised of four stages: (1) component test, (2) integration test, (3) system test, and (4) acceptance test (assuming the existence of a customer for the system). Figure 2 depicts the JTC test process superposed on the vanilla approach, although the mapping is not as clean as the figure connotes. The strength of this framework is that, for the majority of software systems, it provides the recipe for cost-effective development. Of course, there
are a variety of ways to tailor aspects of this approach, pair-wise integration versus big-bang integration for example (see [13]). But, generally, the most cost-effective approach to system testing will reflect this framework, and recognition of this fact influenced the formulation of the JTC development process.

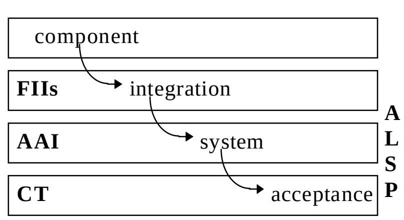

Figure 2: JTC Test Process Superposed on the Vanilla Approach

3.2 Problems associated with a traditional approach

The additional requirement for accreditation complicates the direct application of the vanilla approach within the JTC environment. Ideally, we would like to (annually) develop, evaluate and accredit the JTC and then field it to support the exercises that occur during the year. But several factors stand in the way of this. A few of these are reviewed below.

- *The test system and fielded system differ in scale.*

The systems and personnel assembled to support and participate in a JTC exercise are extensive, typically involving thousands of people and hundreds of workstations for a month or more. In contrast, the 1997 ALSP JTC Confederation Test (user acceptance test) involved roughly 300 people for two weeks, and approximately 40 simulation workstations.

Both practicality and cost limit the size and scale of JTC testing. A test cannot cost as much as an exercise. In fact, testing costs arguably must be many orders of magnitude less than an exercise. As a result, the system that is used to support training is much larger in scale than the system tested.

- *The test system and fielded system differ by platforms.*

Exercises are usually held at military simulation centers (for examples, see Section 1). Each center has a significantly different hardware, software, and network infrastructure. Therefore, test results are only applicable (for certain measures, e.g. performance) with respect to a particular infrastructure (the one used for the test) and, of course, subject to the caveat of the scale problem noted above.

- *The test system and fielded system differ by composition.*

A JTC-supported exercise may be viewed as a layered architecture. At the hub are the JTC simulations (e.g. CBS, AWSIM, RESA, JECEWSI, JCAS, TACSIM, CSSTSS, MTWS, AMP, LAD, AFSAF and PSM for 1997 — see [2] for details) that exchange information and coordinate the advance of simulation time over ALSP. Many of these simulations are interactive, i.e. they have user input terminals and battlefield situation displays. These devices are generally referred to as *controller workstations* and provide a layer around the simulation hub.

An additional layer of systems (referred to in the JTC community as peripheral software systems (PSS)) are routinely attached to the JTC when employed at an exercise. These systems are influenced by actions and changes made within the JTC, and facilitate information exchange with the training audience. The training audience generally does not interact directly with workstations or PSS, but is (ideally) immersed in a wartime-equivalent environment. In some instances, the PSS provide information in a facsimile of real-world formats. A very small number of the PSS support linking real-world command and control (C2) systems and command, control, communications, computers and intelligence (C4I) systems to the JTC. More often the linkage requires a manual interface (so called "human-in-the-loop") to convey information between the JTC and the training audience. The mélange of workstations, PSS, C2/C4I systems, and manual interface mechanisms provides the outer "external systems" layer of the exercise architecture.

During a CT, a system consisting of the JTC and a collection of external systems anticipated for the coming year's training exercises is tested. Unfortunately, the tested system is never identical to the one used at any exercise. Subsets of the simulations and external systems are selected, often based on political or cost rationale that supersede technical considerations. New external systems are occasionally used in exercises (to satisfy exercise requirements) that were not tested during a CT.

- *The test system and fielded system differ by code.*

The JTC development can be viewed as multi-level development life cycle. Each simulation component comprising the JTC has a separate life cycle (necessary to satisfy primary Service requirements) in addition to that of the JTC. These individual development life cycles must be coordinated to coincide with the JTC life cycle. Currently no formal mechanism exists to facilitate such coordination. Modifications to software can occur throughout the year (as dictated by the primary Service development cycle). By organizational agreement, configuration management (CM) authority falls to the Service. Further, imposing a singular CM process that encompasses the peripheral systems mentioned above (to include the "real world" battlefield systems) is a practical impossibility.

- *The test system and fielded system differ by data.*

Many military simulations, including those in the JTC, are essentially data driven, i.e. the outcome of any particular event in the model is sensitive (at varying levels of sensitivity) to values in the model database (see Davis [14] for design rationale). For the JTC simulations these databases include information regarding the types of battlefield equipment, order-of-battle (hierarchical) information, and details of the geographical region being played.

These databases differ greatly between exercises. For example, exercise Prairie Warrior includes next-generation (2010) weapons systems in a fictionalized European scenario, whereas Ulchi Focus Lens uses current weapon systems in a Korean defense scenario.

The results of any test are tightly coupled to the collection of databases underlying the simulations. As a result, database tests must be conducted prior to each exercise. However these tests also fail to be definitive. Database parameters may be changed during an exercise, typically to correct an inappropriate action observed by exercise controllers, or to ensure the exercise scenario evolves as required to meet training objectives.

- *The test system and fielded system differ by personnel.*

Several interfaces between the training audience and the JTC simulations have been automated (i.e. a direct link between C2/C4I systems and simulations). Many of them have not. This is the primary role of an exercise controller: to communicate with the training audience via real world mechanisms, and interact with the simulations using controller workstations.

There is also a level of critical personnel generally referred to as *technical control*. These personnel operate the simulations, the ALSP infrastructure software, and the attendant systems and networks. Both the exercise controllers and technical personnel vary from exercise to exercise. Each can introduce mistakes that unintentionally disrupt the exercise. Novice workstation operators often input unexpected values, producing equally unexpected events in the simulations. System and network personnel will adjust systems and networks differently, resulting in varying behavioral characteristics.

- *System behaviors are subject to evaluation/modification.*

A mechanism for fault tolerance is typically used in training exercises under the nomenclature of *white-cells*. White-cells contain a group of personnel that behave as referees and are used to adjudicate questions regarding the state of the simulation. When an undesired outcome (in terms of training objectives) is detected in the state of the simulation, an exercise white-cell can intercede and realign the simulated conditions to better meet the objectives.

For example, during a recent exercise, an initial engagement involved a massive cruise missile attack by opposing forces (OPFOR) against friendly forces. As a result of operator errors during exercise set-up, many friendly forces did not possess fire control for their air defenses. These units were virtually wiped out during the opening attack. Such an imbalance of forces threatened to seriously impact the training objectives for the exercise, so the referees interceded and the friendly force structure was replenished to an expected post-attack level.

3.3 Causes of system failure

In "traditional" settings (i.e. simulation used for analysis) failure of the simulation is strictly a function of simulation invalidity. This isn't customarily the case in the interactive training simulation world. Numerous problems can be traced to causes outside the simulations themselves, including computer failure, network failure, and operator error. To a lesser extent, general software errors occur. Consequently, the items that a substantive portion of our testing efforts focus on, and the corresponding results that lead us to consider the simulations as "sufficiently valid" are frequently *not* contributing to exercise problems. Although the testing methods are proper and successful test results desirable, the downside is we end up with a confederation of valid simulations which can fail to support an exercise!

For example, in the first days of one exercise the network performance became significantly degraded such that the simulations were running so slow (at a ratio of 0.4:1 with real time) that the training audience was becoming affected. After several hours of investigation the problem was tracked to the fact that a workstation had been added to the network that morning and given the same IP address as one of the mainframes hosting a primary simulation. Of course, this shouldn't happen. A fairly stringent process *was* in place regarding machine connection to exercise network. But the process was inadvertently subverted, and the exercise was jeopardized. How VV&A can (or should) address phenomena like this is an open issue. The important point is that in "traditional" settings, establishing the accuracy of the representation with respect to the objectives is most (if not all) of the battle.[8] In an interactive simulation environment, this is not the case

3.4 Is a traditional approach infeasible?

The accumulation of factors noted above begs the question, "to what extent can model validity be determined during acceptance testing?" The answer seems to be, only partially. The differences in scale, composition, data, and personnel demand that much of the validation activity may only be undertaken within the immediate context of an exercise — using the existent training system, data and personnel as the basis for testing. If this is true, then is there any value in conducting the traditional (component, integration, system, acceptance) tests? Should these tests be abandoned, and all test expenditures redistributed to pre-exercise efforts? Such a course of action would seem unwise. The benefits gleaned from a disciplined, bottom-up testing approach extend beyond validation. They contribute to early error detection as well as verification, and provide feedback regarding the overall reliability of the system.

A cost-effective solution would seem to lie somewhere in the middle, as it often does; its precise location perhaps only identifiable through experimentation.

4. New Directions

Although the challenges to the conduct of VV&A are numerous — some technical and others programmatic — affecting change within the process is necessarily slow. In addition to the natural resistance to change, the role of the JTC as a proven and reliable environment supporting training exercises demands that modifications to a process that works sufficiently well should be made with caution. There is wisdom to the philosophy "if it ain't broke don't fix it".

Thus, the changes envisioned for the future are intended to automate portions of the testing and validation activities, to reduce the time (and therefore cost) of the AAI and CT efforts, and to enhance the structure of activities that normally precede an

[8] Type I error, or "model builder's risk," exists in any modeling situation. This is the risk that results from a valid simulation are dismissed by the decision maker [3].

exercise. Several initiatives are underway for the 1997 development cycle.

- A new test harness has been designed to support component testing at the system level. The harness will be "programmed" to stimulate an actor with respect to the interfaces it participates in (as described in the Interface Control and Actor Implementation documents).
- To address the sensitivity of model validation on the underlying model databases, a mechanism is being provided to allow each actor to *register* the values of key database elements with the ALSP infrastructure during the confederation initialization process. The database elements are enumerated attribute values defined at the confederation level — so-called *enumerations mapping*. A tool has been developed to collect this information and provide cross-referencing to facilitate the detection of incomplete entries or errant mappings. This facility could significantly improve what has been an entirely manual, labor-intensive and error-prone process. However, not all of the JTC actors can easily register their enumerations mapping. In some cases this mapping is resident in code rather than a separate database, thus making the creation of a routine that extracts this information a difficult (if not impossible) task.
- Another tool has been developed to monitor the message traffic and examine messages for invalid enumerations. Note that should each actor have the capability to register its enumerations mapping (as noted above) *and* if no database changes occur during an exercise, this tool would provide no additional benefit. However, in the absence of such capabilities and guarantees, this tool, while providing less information than an enumerations mapping cross-reference, can provide a valuable mechanism for quickly identifying certain common errors that occur during confederation operations.
- Over the next two testing cycles, the length of the AAI and then the CT will be reduced. This reduction is expected to be achieved by reducing the time allocated for functional testing, and the time allocated for load testing. Such a reduction reflects the reduced pace of changes being introduced into the JTC during this period.
- To address the problems noted in Section 3.2, pre-exercise testing efforts will be enhanced. Specifically, the formalization of pre-exercise load tests is being investigated.

5. Conclusions

Within the JTC development process, cost-effectiveness serves as an overriding objective. However, while costs must be constrained, system viability cannot be sacrificed – training must continue. Although a variety of factors complicate their application, the testing and VV&A activities within the JTC process seem to provide cost-effective evaluation leading to the annual production of a viable system. The track record of the JTC as a training vehicle has few blemishes. While the size and scope of the JTC continues to expand to keep pace with ever-expanding training requirements, the time and money allocated to testing promises to remain fixed (or even diminish). Given that model validity is a function of cost versus benefit, the challenge to the Systems Engineer for the JTC is to maximize the benefits (i.e. the validity) in the face of fixed costs.

In a recent memorandum signed by the Undersecretary of Defense for Acquisition and Technology, ALSP and its cousin, the Distributed Interactive Simulation (DIS) protocol, are targeted for replacement within the U.S. Department of Defense (DoD) by the emerging High Level Architecture (HLA) for models and simulations. The JTC itself is also scheduled for replacement by the Joint Simulation System (JSIMS). As JTC Systems Engineer, MITRE actively supports the transition to these future systems and architectures, and it is our hope that the lessons learned from the ALSP JTC will enhance their realization and success.

3. References

[1] Defense Advanced Research Project Agency (DARPA): "Advanced Distributed Simulation Program." Project Report. Advanced Research Project Agency, March 1990.

[2] E.H. Page, J. A. Tufarolo and B. S. Canova: "A Case Study of Verfication, Validation and Accreditation for Advanced Distributed Simulation." *ACM Transactions on Modeling and Computer Simulation*, **7**(3), July, 1997 to appear.

[3] O. Balci: "Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study." *Annals of Operations Research*, **53**, pp. 121-173, November 1994.

[4] B. Hetzel: *The Complete Guide to Software Testing*. QED Information Sciences, Inc. Wellesley, MA 1988.

[5] Department of Defense: “MIL-STD-498: Software Development and Documentation.” United States Department of Defense, 1988.

[6] G. Miller and A. Zabek: “The Joint Training Confederation and the Aggregate Level Simulation Protocol. *Phalanx*. **29**, pp. 24-27, 1996.

[7] K. Forsberg and H. Mooz: “The Relationship of Systems Engineering to the Project Cycle.” *Engineering Management Journal* **4**(3), pp. 36-38 1992.

[8] M.C. Fischer: “Aggregate Level Simulation Protocol (ALSP) Managing Confederation Development.” In: *Proceedings of the 1994 Winter Simulation Conference*, pp. 775-780, Lake Buena Vista, FL, 11-14 December 1994.

[9] U.S. Department of Defense: “DoD Modeling and Simulation (M&S) Verification, Validation and Accreditation (VV&A).” Directive 5000.61, April 1996.

[10] O. Balci: “Guidelines for Successful Simulation Studies.” In: *Proceedings of the 1990 Winter Simulation Conference*, pp. 25-32, New Orleans, LA, 9-12 December 1990.

[11] O. Balci: “Principles of Simulation Model Validation, Verification, and Testing.” Technical Report TR 94-24. Department of Computer Science. Blacksburg, VA, June 1994.

[12] D. Harel: “Biting the Silver Bullet:Toward a Brighter Future for System Development.” *IEEE Computer* **25**(1), pp. 8-20, January 1992.

[13] B. Beizer: *Software System Testing and Quality Assurance*. Van Nostrand Reinhold, New York, NY, 1984.

[14] P.K. Davis: “Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations.” R-4249-ACQ. The RAND Corporation. Santa Monica, CA, 1992.

Author’s Biographies

ERNEST H. PAGE is a member of the technical staff of the MITRE Corporation where he is currently working on the Aggregate Level Simulation Protocol (ALSP) program and is the Principal Investigator for a MITRE Sponsored Research project on web-based simulation. He received the Ph.D., M.S. and B.S. degrees in Computer Science from the Virginia Polytechnic Institute and State University in 1994, 1990 and 1988 respectively. He serves as a member of the Defense Modeling and Simulation Office (DMSO) Verification, Validation and Accreditation Technology Working Group and is a member of the Test Forum Planning and Review Panel for the 1997 Simulation Interoperability Workshops. He has been active within the High Level Architecture (HLA) initiative, working with both the Test and Interface Specification Working Groups. Dr. Page is the Vice Chairman for the Association for Computing Machinery (ACM) Special Interest Group on Simulation (SIGSIM) and is a regular contributor to *ACM Computing Reviews*. His research interests include discrete event simulation, parallel and distributed systems, and software engineering. He is a member of ACM, SIGSIM, IEEE CS and Upsilon Pi Epsilon.

WILLIAM E. BABINEAU is a member of the technical staff of the MITRE Corporation where he is currently working as the lead engineer for VV&A and testing on the ALSP program. He received the M.S. degree in Computer Science from the Stevens Institute of Technology in 1986, and the B.S. degree in Electrical Engineering from the Rensselaer Polytechnic Institute (RPI) in 1981. He has over 15 years of software systems development and test experience with several aerospace and defense corporations, having worked on applications in automatic test equipment, data storage devices, and sonar, infrared and navigation systems. He is a member of ACM and IEEE CS.