1/31/2011

Validation Referent

Introduction

Referent has a number of connotations, each of which may be appropriate in the context in which the term is used. The following definition is most often related to modeling and simulation (M&S):

Referent: A codified body of knowledge about a thing being simulated

This is the definition for referent in the Department of Defense (DoD) glossary of M&S terms.[1] The definition derives from work on M&S fidelity in the Simulation Interoperability Standards Organization's (SISO) Simulation Interoperability Workshop (SIW) during the late 1990s.[2] In M&S, the referent has two functions. It supports simulation development (and modification), and it supports simulation assessment.

In supporting simulation development, the referent is the information upon which development of the simulation is based. That information determines the entities, processes, and interactions that should be included in the simulation to represent the **simuland** (the system being simulated). That information determines the algorithms and data that should be used in the simulation. For clarity, it is helpful to call the referent used to support simulation development or modification the development referent.

In supporting simulation assessment, the referent is concerned with fidelity, validation, and accreditation. The referent provides the information with which simulation results are compared to determine its fidelity and validity. That comparison provides factual information to support an accreditation decision. For clarity, when a referent is used in this way, it is helpful to call it an assessment referent or a validation referent.

This Special Topic is focused on use of a referent as a validation referent. This Special Topic is primarily for those who will be involved in DoD M&S validation and accreditation assessments. It provides specific guidance so that those who select, describe, approve, and use validation referents in M&S validation and accreditation assessments can do so effectively and efficiently. A checklist for validation referent identification and specification is provided at the end of the Special Topic.

An Expanded Definition

Validation of models and simulations is defined as "[T]he process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model."[3] This connotation for validation is generally accepted, both within DoD and elsewhere, but sometimes people modify the connotation in various ways. For example, some restrict the term "validation" to situations in which simulation results are compared to high-quality test data. Every such modification either adds to or subtracts from the connotation of validation in the DoD definition and should be avoided. Even those outside DoD would do well to use the term "validation" with the DoD connotation to avoid confusion because of the extensive use of the term with DoD models and simulations.

The DoD definition for validation states what validation is, but does not describe how the accuracy of the real-world representation is determined. Determination of representational accuracy requires comparison of the simulation, its data, and its results with something. That something is the validation referent. The referent definition presented in the Introduction is not the best definition for the context of a referent used to support validation and accreditation assessment because the definition is too general. An expanded definition of validation referent will be more useful.

An expanded definition for *validation referent* shown below is based upon a 2004 study sponsored by the Defense Modeling and Simulation Office that involved participants from four nations.[4] U.S. participants came from Air Force, Army, and Navy organizations, the Missile Defense Agency, the National Academy of Sciences, academia, and industry.

> **Validation Referent:** The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for the simulation's intended use.

Various words and phrases from this definition are discussed below to ensure that the expanded definition for validation referent is fully understood. The words and phrases discussed are shown in **bold type**.

The **information** used as the validation referent may consist of the following types of information, alone or in various combinations:

- "Data" (observations of the simuland, either under controlled circumstances as in tests and experiments, or under natural or operational circumstances)

- Theories as expressed in algorithms that describe characteristics, behaviors and relationships (preferably theories validated against observations of the simuland)
- Simulation results (preferably from simulations that have been objectively and quantitatively validated by comparison with data about the simuland)
- Expert human knowledge which, in M&S assessment, is described as subject matter expert (SME) estimations

Consequently, the quality (accuracy, reliability, comprehensiveness, credibility, etc.) of the information in a validation referent may vary from one validation referent to another, and within a validation referent. In cases where explicit observations and theories do not provide a comprehensive and sufficiently reliable description of the reality represented in the simulation (i.e., the simuland), information from theory, other simulations, and SMEs may have to serve as all or part of the validation referent. Such SME information may not be explicitly articulated and systematically organized, since it is a form of qualitative assessment, and the bases for qualitative assessments usually are not as explicitly described as those for quantitative assessments.

M&S communities that deal mostly with physics-based models that may in principle be compared against quantitative data tend to be less willing to accept SME information and other qualitative assessments as part of the validation referent than are other M&S communities. The computational science and engineering community, which has major verification and validation (V&V) concerns in applications of computational fluid dynamics and computational solid mechanics simulations, is an example of that kind of M&S community. It should also be noted that the computational science and engineering community has been more prolific in V&V publications than most M&S communities. Publications of this community include V&V guides from professional communities as well as the only college/graduate-level textbook on V&V.[5,6,7]

However, subjective assessment is present even in the most quantitative situations. Test results are checked to ensure that the results have not been contaminated by bad data. Bad data can be introduced by faulty instrumentation, errors in installation of the instrumentation, etc. The erroneous data is removed from data considered as valid test results. Sometimes the rationale for removal of such results may not be clear. At times, data has been removed merely because the results were not in the range expected by those conducting the test. The impact of this qualitative judgment on quantitative test results is seldom addressed explicitly.

The information used as a validation referent is a **codified body of information**. The word “codified” has several connotations. It implies a system and organization, both useful aspects of information to be used as a validation assessment referent. Codified often implies authority and is frequently associated

with policy. Wherever appropriate and possible, the validation referent should be information drawn from credible, authoritative data sources.

Unfortunately, a codified body of information many not cover the full spectrum of the intended use domain, even when that codified body of information draws upon theory and SME estimations as well as observations and test data. In addition, the codified body of information may have contradictions (or apparent contradictions) among different information items. Such aspects have to be addressed when encountered in validation referent identification, selection, and specification.

The definition above deals with the "**reality represented in the simulation**." This refers to the simuland. Typically that reality includes actors, systems, or entities interacting with other actors, systems, or entities by various processes through or in one or more environments. The validation referent pertains to all of these: actors, systems, and entities; processes; interactions; and environments. Connotations associated with these may vary by M&S type and by the kind of application. For example, the simulation operator would not be part of the validation referent for a batch-run constructive simulation, but the simulation operator might need to be part of the validation referent in an interactive simulation (such as a game or war game).

Three challenges arise from the validation referent as the **best or most appropriate** information available that describes **characteristics and behavior** of the reality represented in the simulation **from the perspective of validation assessment for intended use of the simulation.**

1) Developing and organizing the **best** information possible may be too costly and/or take too long for schedule and resource constraints of a particular M&S project. This leads to use of an *adequate validation referent*; one that is not the best information possible but that is adequate as the basis for validation assessment within the context of the simulation's intended use. Such an adequate validation referent can be the "most appropriate information available."

   The idea of an adequate validation referent increases the importance of exact and precise specification of the simulation's intended use. A poorly specified intended use increases the risk of decision error, i.e., the risk either that a simulation will be judged acceptable (valid) for the intended use when it is in fact not acceptable, or that an acceptable simulation will be judged unacceptable. Consequences of such mistakes depend upon the impact of the simulation on the overall project. There may be little consequence from such a decision error for a simulation that produces only background information about a subject, but catastrophe could result from such a decision error about a simulation used as part of a real-time decision aid for a safety-critical system. The ultimate authority for the acceptability of a validation referent is the accrediting activity (or Accreditation Authority), the person or organization responsible for the

decision that the simulation is appropriate for the intended use. This Special Topic develops rationale for identification, selection, and description of a validation referent, which also serves as a basis for its acceptance by the accrediting activity.

2) The **most appropriate** information has adequate fidelity to serve as the basis for comparison in a validation assessment of the simulation for its intended use. It is a truism that one cannot demonstrate greater fidelity for simulation results than the fidelity of the validation referent to which those results are compared. Consequently, the validation referent must be able to address the capabilities required by the intended use at a sufficient level of fidelity; otherwise the body of information selected is inadequate to support an assessment of the simulation results; there would be insufficient evidence to support a judgment. Vague specification of the simulation's intended use can make it difficult to determine required fidelity for an adequate validation referent. This leads to increased decision risk and makes it more likely that an inadequate simulation will be considered acceptable for the intended use. For more information see *Advanced Topics>Special Topics>Fidelity*.

3) As a general rule, the most appropriate collection of information to serve as the validation referent is the least expensive set of information that has adequate fidelity to support the intended use of the simulation. Otherwise, the validation referent becomes "gold plated" and its development wastes resources. For example, if simulation results are to provide ballpark estimates of performance as supplemental information to the User, SME judgment may be an acceptable validation referent for such "back of the envelope" fidelity. Expending time, effort, and resources collecting data and formatting referent data beyond the scope of what is needed to supports the intended use can result in a waste of resources.

Because the validation referent **describes characteristics and behavior of the reality represented in the simulation,** specification of the validation referent is particularly difficult for simulations that generate new knowledge or additional knowledge about the reality represented in the simulation (as opposed to simply being a reliable representation of that reality). This could be the case with a simulation demonstrating emergent behavior. This also could be the case if the simulation employs some form of judgment in uncertain conditions (such as representation of human decision-making) or if it portrays a future reality that is difficult to predict (such as social structure, military posture, etc.). There are two implications here for validation referent specification. One is that iteration of the validation referent specification may be necessary as new knowledge is generated. More iterations can translate into a draw on resources. The other is that the intended use of a simulation with such uncertainties in the simuland implies that the validation referent specification will have low fidelity. It is not the length of time projected into the future that is the issue; astrophysical simulations project conditions billions of years in the future with high fidelity. It is the

uncertainty in the representation that determines the fidelity achievable in the validation referent specification.

Additional Comments Related to the Expanded Definition

The expanded definition for validation referent presented above can apply to all varieties of simulation and their full range of potential applications, but how one identifies the information appropriate for a particular validation referent and selects an appropriate set of information as, well as how one describes the validation referent, may vary with simulation type and application. Subsequent sections will discuss these topics.

The focus on validation referent use in validation and accreditation assessment makes it necessary to consider issues related to statistical independence (or dependence) of the information used for simulation development relative to information used as the referent in validation and accreditation assessments when simulation results are to be used to predict performance or behavior. If the same referent is used in both development and assessment, the model is essentially "tuned" or "calibrated" to the referent. Many M&S communities, especially those with models or simulations that have heuristic factors that are adjusted to make simulation results fit experimental data better stress the importance of separating data used in calibration from the data used in simulation validation assessments.[6] The problem is similar to that encountered with clinical trials in the health field, in which randomization of patient involvement in a clinical trial (with its associated unpredictability) is used to "protect against the unpredictability of the extent of bias in the results of non-randomized clinical trials".[8] Many do not appreciate the issues associated with using the same information for simulation validation assessment as was used for simulation development or the limitations that such use places on what can be determined about simulation accuracy or predictive capability under those conditions.

For editorial simplicity, the term "validation assessment" may be used to encompass all referent aspects of both validation and accreditation assessment. It is important to appreciate the difference between validation assessments and accreditation assessments, however. The validation assessment determines whether the simulation has adequate fidelity to support intended uses, and perhaps can even quantify the likelihood that the simulation has the required fidelity with a statement like "we have 90 percent confidence that simulation results will differ from the validation referent by less than 5 percent for any conditions within the specified application region." Such a quantitative statement is the most precise kind of validation assessment, and depends upon a very well defined intended use, an explicitly defined validation referent, and a precise measurement capability for the simulation. Few validation assessments have had such quantitative precision.

Accreditation assessment uses the validation assessment, but supplements it with risk, programmatic, and other considerations to determine if the simulation is

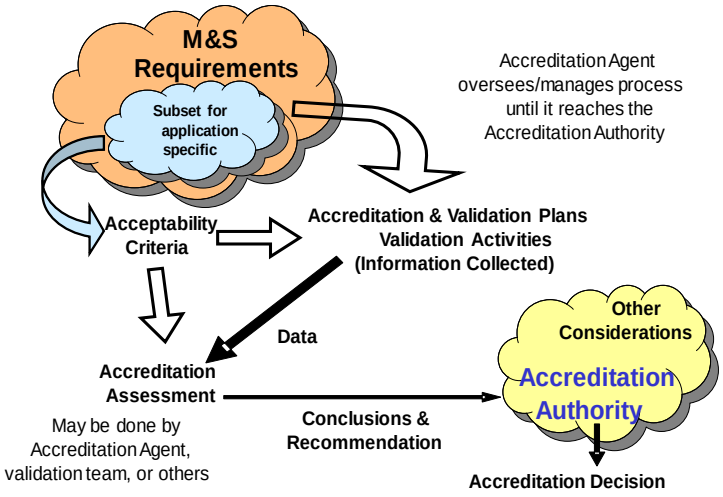
appropriate for use in a particular situation. Whenever accreditation assessment is mentioned in this Special Topic, the only concern will be its validation assessment component – other aspects of the accreditation assessment are outside the scope of this Special Topic. This makes it convenient to use the term validation assessment to refer to either or both validation and accreditation assessments. It is important to remember that validation assessment is always in the context of the simulation's intended use, and that the accreditation activity (or Accreditation Authority) is the one who determines the acceptability criteria for the intended use and, thus, in essence is also the one with final approval authority for the validation referent.

Validation Referent Identification and Selection

This section is concerned with three topics. First, how to identify possible validation referents for simulation validation assessment; second, how to select an appropriate validation referent when there is more than one possible referent; and third, what to do when an appropriate validation referent does not exist or cannot be obtained within available time and resources.

Validation Referent Identification

Two factors make identification of a validation referent non-trivial: a) simulation requirements and b) the level at which the reality is to be represented in the simulation (i.e., how close does it need to represent the simuland). Requirements identify what is to be represented in the simulation (actors, systems, entities, processes, interactions, environments, etc.) and the simulation purpose determines the level of fidelity that the representations require. The figure below illustrates how requirements drive the accreditation assessment process.



The Role of Requirements in the Accreditation Process

To illustrate how these factors make validation referent identification nontrivial, consider the validation referent of the natural environment for an M&S application whose purpose is to predict sensor performance capability. If a radar sensor is involved, natural environmental factors that could be pertinent include terrain (to determine where radar signals would be masked by terrain features); precipitation intensity, extent, and duration (since radar signals can be attenuated by precipitation); temperature and pressure gradients with altitude (these affect radar ducting); etc. These factors can depend upon geographic location and vary with time of day and time of year. They may change from one day or year to the next. For some purposes, a simple free-space radar range computation will be adequate (a computation that is not affected by any of the factors mentioned above), and in other cases, all of these factors must be considered. In some cases, parameter values from standard handbooks or catalogues will provide the information needed. In other cases, detailed observations must be made at the specific site where the radar will be located. In some real-time applications, the information may have to be current, not historical (neither recent history from the previous day nor history from years before will be adequate).

The simulation's purpose and the nature of the reality to be represented, as specified in the requirements, determine what characteristics the validation referent should have (what parameters should be represented and how). Information sources that contain all pertinent factors for the characteristics specified provide an initial set of possible validation referents. Simulation fidelity requirements then determine accuracy and uncertainty characteristics of the information needed for it to qualify as a possible validation referent.

When some information items are not available in a possible validation referent (or information items do not have the needed fidelity, or information fidelity is uncertain), it will be necessary either to institute a test program to develop the needed information, accept the uncertainty in the simulation results resulting from an insufficient referent, or re-scope the simulation requirements to address the referent limitations. This may result in only part of the intended use domain space being addressed by the assessment or the use of a less rigorous referent such as a theoretical value or SME estimation for selected parameters instead of requiring test data of a specified fidelity. For a complex simulation, the validation referent is often drawn from a collection of information sources, not from a single information source.

Correlating the validation referent with simulation requirements is essential; otherwise, the simulation may not be able to support its intended use. Whether theory or SME estimation has adequate fidelity for the validation referent, or extensive use of precision data from tests is required for the validation referent to have adequate fidelity, is determined by the requirements and the intended use.

Validation Referent Selection

The validation referent is usually selected on the basis of direction, convenience, economics, a decision to use proxy information for the validation referent, extent of coverage of the intended use domain, community acceptance of the information (i.e., its credibility), or some combination of these. Each of these is discussed below. Any possible appropriate validation referent must be acceptable in terms of scope, reliability, credibility, and fidelity.

> Example:
>
> Multiple acceptable validation referents are possible; consider the possible validation referents for the length of a meter. From 1889 to 1960, the official international validation referent for the length of a meter was an international prototype of a meter with an estimated accuracy of less than 10^{-6} meter error. From 1960 to 1983, the international referent for the length of a meter was based upon the wavelength of 86 Kr with an estimated accuracy of less than 10^{-8} meter error. Since 1983, the international referent for the length of a meter has been based upon the speed of light in a vacuum with an estimated accuracy of less than 10^{-9} meter error. People often use a less accurate referent for measurements than the international standard. Such less accurate referents range from the ancient informal standards of the distance from finger tip to nose or a long stride to contemporary approaches such as a standard measuring tape, a yard stick, or a more precise measuring device. Depending upon the application, measuring devices (referents) with accuracies measured in terms of a few percent (or more) could be acceptable. One need not go to the expense of measurement accuracies associated with international standards for length if the measurements are only going to be used to estimate how much paint will be needed to paint a wall. Measurement accuracies of 80 to 90 percent will be more than adequate for that application.

This simple example illustrates how a variety of referents may be adequate for a particular application. Selection of the validation referent that is most appropriate depends upon the circumstances, particularly upon the simulation's intended use.

Validation Referent Selection by Direction

When the validation referent selection is directed, the validation referent is specified by the M&S sponsor or User and approved by the accreditation activity or authority. Often a particular validation referent will be selected due to its compatibility with other aspects of a large program or for compatibility with related programs. It should be noted in that if the specified validation referent does not have sufficient fidelity, it will be necessary to communicate the referent limitations to the User and/or Accreditation Authority and determine options for enhancing the defined referent, re-scoping the intended use, or accepting the resulting risk.

If a SME is serving as the validation referent or must supplement facts and theories as part of the validation referent, it is essential the identified SME has the proper domain knowledge. A SME's credibility with the User and/or Accreditation Authority may be the result of past interactions, the SME's curricula vita, the SME's organizational association, as well as establshed technical expertise.[9] For

more information see *Advanced Topics>Special Topics>Subject Matter Experts and VV&A*.

Validation Referent Selection by Convenience

When the validation referent is selected for convenience, the validation referent may be the one easiest to access, one already available rather than one expected in the future, or one that the M&S team knows and is comfortable working with. As long as the validation referent has acceptable fidelity, no issues arise from selecting the validation referent on the basis of convenience.

Validation Referent Selection by Economics

Since cost of information is usually related to quality (better costs more), it makes economic sense to balance between quality and cost. Selecting or developing the best validation referent requires a balance between the level of representation needed to satisfy the intended use and the resources required to develop a sufficient enough referent. If economic considerations preclude the selection or development of an adequate validation referent, then options for enhancing the referent need to be defined, the intended use may need to be re-scoped, or the risk resulting from the limited referent may need to be accepted.

Validation Referent Selection by Proxy

It may not be possible to select or develop a validation referent for the specific domain being simulated. This is nearly always the case when the reality represented by the simulation is a future system. In that case, performance of similar tasks on existing systems may be adapted to provide a referent. For example, when assessing reaction and decision times for operators of a new combat system or machinery control system, there may be no data on operator performance; however, past experience with operator performance in the accomplishment of similar tasks on existing systems may be used as the validation referent for simulations of the new system.

Validation Referent Selection by Coverage Extent

For situations in which the simulation intended use application domain is very large, it may be desirable to limit the number of referent sources to minimize the problems associated with ensuring the cohesiveness and consistency of the validation referent. Determination of information adequacy when the information covers a broad spectrum of parameter values always needs careful attention. Often the fidelity across information sources varies with parameter value. Thus, coverage extent has to be within the context of acceptable fidelity for the intended use.

Validation Referent Selection by Community Acceptance

In some situations, community attitude toward simulation results is more important than in other situations. In situations where community attitude is important, selecting the acceptable validation referent with the greatest community acceptance will facilitate community acceptance of simulation results, since the community will trust the standard for validation assessment. Confidence in the assessment standard facilitates confidence in simulation results.

Validation Referent Selection by a Combination of Considerations

More often than not, validation referent selection will involve more than one of the factors mentioned above. There is no general rule for how to weight the different factors. Weighting of the various factors depend upon the particular situation. The overriding, fundamental principle, however, regardless of what factor or factors are involved in referent selection, is that any validation referent selected must be sufficiently detailed to support the intended use.

Consequences Arising from Lack of an Adequate Referent

Sometimes the information that is available or likely to become available within the expected time and resources before the intended use of the simulation will not be adequate to support validation assessment for the intended use. In such a situation, one of the following three options must be taken.

- Accept that validation assessment will not be possible because of lack of an adequate validation referent
- Accept the risk associated with applying an incomplete referent. Define and document areas of uncertainty
- Provide additional time and resources so that an adequate validation referent can be developed
- Modify the simulation's intended use (and associated requirements) so that available information can provide an adequate validation referent

The Accreditation Agent, V&V Agent, or others involved in identification, selection, and preparation of the validation referent must clearly explain the situation so that the appropriate authority can decide which option to select.

Validation Referent Specification and Description

This section addresses of the following topics:

- Where in the M&S life cycle the validation referent should be specified

- What should be included in description of the validation referent
- How the description of the validation referent depends upon M&S type, the particular M&S application, and the availability of referent data

Sometimes the description of the validation referent consists of pointers to detailed information elsewhere (such as test documentation and results), with the referent description in simulation documentation merely providing the pointers and explaining how the indicated information fits in the validation referent. Other times, simulation documentation of the validation referent will include both the information and its description. Discussion in this Special Topic applies to both situations.

Validation referent descriptions have five fundamental aspects: 1) context, 2) scope, 3) resolution, 4) parameter uncertainty quantification, and 5) information coherence. Specification or description of the validation referent should include all five aspects. Each of these aspects is discussed below.

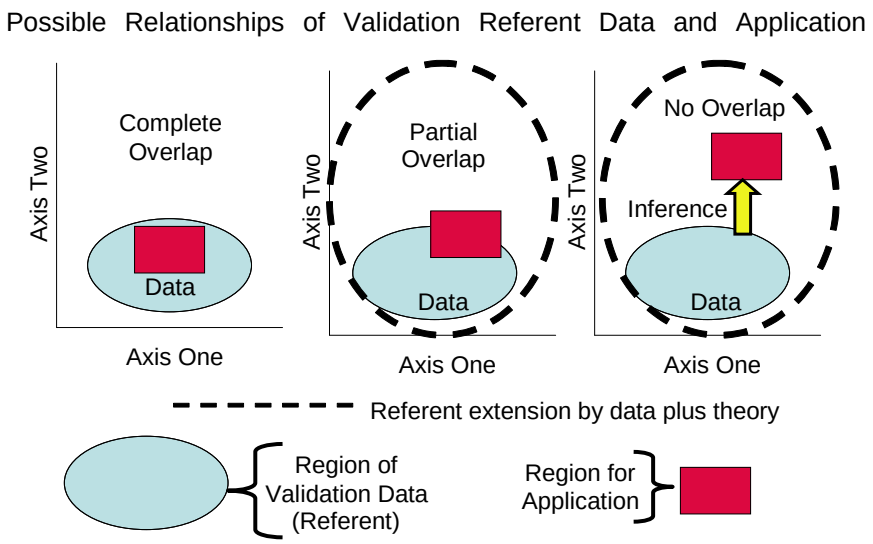
Some general characteristics should apply to every validation referent specification. Identification and specification of a validation referent should be definite and unambiguous. Just Identifying and specifying a validation referent consistently is an improvement over current M&S practice. If information in the validation referent uses only part of information from a particular source (such as results from one test in a set of tests), what is not to be used should be clearly and specifically delineated. It usually helps to add an explanation of why this is being done.

> Example:
>
> If the flight profile of a missile is to come from a particular test, but the radar signatures of the missile in that test are not appropriate as part of the validation referent for the missile represented in the simulation (perhaps the signatures of the missile in the test were augmented to facilitate range tracking of the missile during the test), that should be noted explicitly and clearly in the validation referent identification and description. This eliminates any possible confusion about what is to be used as the validation referent.

1) **Context** – Context addresses the environment within which the validation referent information is applicable. It is a critical component of a referent specification and is particularly important for situations in which there may be debate or uncertainty about the scope of the validation referent. Rather than describing the actors, systems, entities, processes, and interactions represented in the simulation or their fidelity, context describes the parameters and factors that might influence those descriptions. Context information can range from conditions under which information (e.g., human performance information) is collected to physical conditions (temperature, pressure, radiation, etc.) that might influence measured parameters in the validation referent which are not addressed specifically by algorithms.

2) **Scope** – Scope or domain coverage is concerned with the range of parameters or application domain(s) of concern for the referent. The intended use defines the appropriate application domain(s). The description of the validation referent should indicate what portion of that application domain is addressed by the referent. For example, test data used as a validation referent may reflect steady-state, smooth, undisturbed flow conditions for a parameter (such as fluid volume passing through a pipe, traffic on a road, time delay waiting for a technician in on-line support, etc.) but such data are not appropriate for use as a referent in transition, turbulence, or disturbed flow conditions (such as might be experienced if an obstacle were in the pipe or on the road, or the on-line support shift is short-handed).

   Limitations in validation referent domain coverage can force consideration of inference. The figure below illustrates three possible relationships between data for a referent (the blue ovals within the solid lines) and intended use (red boxes) in the application domain. If the validation referent is restricted to data, only when the data completely overlap the application area can there be high confidence in a quantitative assessment of the relationship between simulation results and the validation referent. In the other two situations (partial overlap and no overlap), indeterminate uncertainty is present for applications outside the data region. If reliable theory exists, then either of these cases might be capable of being reduced to the equivalent of the complete overlap case by use of theory for any point outside the data region.



Data Application Domain Overlap Possibilities

General parameter regions are fairly easy to understand as shown in the following example.

Example:

If a simulation is to address what happens when bodies collide (as with a missile defense kill vehicle hitting its target), the range of interest for collision velocities may vary from nearly zero (a scenario in which the kill vehicle approaches from behind the target) to very fast (10 km/s or more for a fast interceptor against a ballistic missile in a head-on encounter). Data from full-scale tests may be very limited, not only in the number of tests, but also in the portion of the speed regime for which there are tests and in the availability of precise information from the tests. Data from surrogate tests (sled-tests or light gas gun experiments) may supplement the full-scale test data, but they introduce uncertainties because of test artifacts (such as need to scale results, differences between the surrogate and the real object, etc.), and such tests may not fully cover the parameter regime of interest. Then theory or perhaps very high fidelity simulation results may serve as "data" for parts of the parameter regime that testing (either full-scale or surrogate) cannot address. Finally, expert opinion may be used to fill in any remaining information gaps in the domain (and to reconcile discrepancies among the various types of information).

The defined referent will indicate how information about the M&S actors/systems/entities, processes, and environments throughout domains of interest should be addressed. This description also identifies the information sources to be used for each portion of the domain, gaps that may exist, methods to reconcile information from different sources, and methods to combine information in the validation referent.

Whether a partial or no overlap situation is adequate for the validation referent (possibly supplemented by theory and SME estimations) depends upon the simulation intended use and the level of fidelity required for validation assessment.

3) **Resolution** – Resolution is focused on the level at which distinctions can be made in the representations characterized in the validation referent. The validation referent describes all aspects of the simuland to include the actors/systems/entities, their interactions, and the environment(s) they operate in. As illustrated earlier, not every possible attribute may be significant for the intended use. The level of resolution needs to satisfy simulation requirements and fully describe the domain coverage required with all pertinent variations indicated. For example, if the object size or color changes with temperature, and the object's size or color is important for the intended use of the simulation, that attribute trait should be specifically noted.

4) **Parameter Uncertainty Quantification** – There are two uncertainty dimensions in fidelity and validation assessments: simulation uncertainty, and uncertainty in the referent. Simulation uncertainties arise from imperfect algorithms, computation characteristics (such as table look-up errors), input errors, etc. Often when simulation results are compared with a standard (theoretical curve, test data, etc.), it is assumed that all error or uncertainty is a result of simulation uncertainties. This usually is not the

case. The other uncertainty dimension in fidelity and validation assessment is uncertainty in the referent. Only when referent uncertainties are characterized can simulation uncertainties be characterized. When both uncertainty dimensions are fully characterized, then fidelity and validation assessments can be rigorously performed.

A great deal of attention has been given to quantifying uncertainties, both simulation uncertainties and referent uncertainties. Various validation metrics have been suggested, guidance in regard to validation experiments has been developed so that referent information with characterized uncertainty may be articulated. A proper validation referent description always includes specification of parameter uncertainties in validation referent information, even when description of parameter uncertainties is limited (such as simply bounding the uncertainties, e.g., the uncertainty is expected to be less than *x* percent). Sometimes the available information does not permit a better characterization of the uncertainty.

Validation referent uncertainty can also be characterized by the kind of information used in the referent. Three basic kinds of information are a) quantitative test and experiment data, b) fully formulated theory that has been correlated with substantial evidence and data, and c) and loosely formulated theory including SME estimations.

Standard statistical processes can be applied to quantitative test experiment data to estimate the uncertainties (and accuracies) of a validation referent based upon such information. The more data one has, the better the estimation of such uncertainty and accuracy.

Standard statistical processes can also be applied to fully formulated theories to estimate the uncertainties (and accuracies) of a validation referent based upon such information. A fully formulated theory has demonstrated its reliability by being successfully applied in numerous circumstances by a wide variety of people.

Loosely formulated theories and SME **estimation**s have limited repeatability, i.e., different people get different results when these assessment methods are applied and the uncertainties of these assessment methods may be difficult to estimate. One method of dealing with the uncertainties associated with SMEs use is to use multiple SMEs as the basis of the referent. This permits statistical analysis of SME estimations, providing some indication of variability in the SME responses as well as revealing trends in SME responses. For more information on SME usage see *Advanced Topics>Special Topics>Subject Matter Experts and VV&A*.

5) **Information Coherence –** As indicated previously, information for the validation referent may come from multiple sources. Some will be

redundant (same parameters for the same part of the same domain), some will be supplementary (different parameters for the same part of the same domain, or same parameters for different parts of the same domain), and some will be disjoint (different parameters and different domains). Some information will have parameter uncertainties quantified, and other information will not. Information coherence is concerned with how information is combined so that information about a particular aspect of the validation referent (parameter, actor, system, entity, process, interaction, environment, etc.) makes sense and is compatible with other information in the validation referent. The validation referent description should explain how information combination is performed so that information coherence is achieved.

The two examples below illustrate some of the ways this might be done. Neither example tries to show a best or preferred way for combining information. Decisions about what is best or most appropriate in a particular situation will depend upon details of the situation.

Example 1:

Information about parameter *x* will be taken from three tests, all of which address how the parameter varies with respect to *y* over the same set of values for *y*. In Test 1, uncertainties in the measurement of *x* are fully characterized. Tests 2 & 3, which were done at the same test facilities using the same equipment and test personnel, do not characterize *x* uncertainties nor describe environmental conditions fully. Options for referent information about *x* include:

- Use all data from Tests 1–3 and ignore uncertainties about *x*
- Use only data from Test 1 and fully characterize *x* uncertainties
- Use a plot of *x* from Test 1 with uncertainties indicated, overlaid with data from Tests 2 and 3, as a basis for an equation to characterize *x* vs. *y*

Example 2:

Information about parameter *x* is available from Test 1 for a portion of the domain of interest. Information about a parameter *y*, to which it is believed that *x* is proportional, is available for a portion of the domain of interest not covered by Test 1 (but without any information for the part of the domain that is addressed in Test 1). Two SMEs gave widely divergent opinions about what multiplication factor should be assumed in the proportional relationship between *x* and *y*. Options for referent information about *x* include:

- Use only data from Test 1
- Use both data from Test 1 and a band for *x* determined by proportionality to *y* (with the two SME estimations establishing the edges of the band)
- Declare that *x* is unknown

When and Where to Specify the Validation Referent

There are many ways to describe the M&S life cycle. One way to describe it is to break it into eight phases, which may be passed through serially or repetitively, with more than one of the phases occurring concurrently. These phases occur regardless of M&S development paradigm employed, whether a serial paradigm or an iterative one. The phases are:

1) Requirements definition (expression of the needed or desired simulation capabilities)

2) Planning

3) Conceptual model

4) Design

5) Implementation

6) Testing

7) Use

8) Modification

Early specification of the validation referent in a new simulation development (or in a major modification of a legacy simulation) is needed to support conceptual validation assessment of the simulation. The same validation referent should be used for conceptual validation assessment as is used for results validation and for the accreditation assessment. Otherwise the standard by which the simulation is evaluated changes from one aspect of assessment to another, and such shifts can have negative impacts, both on the on efficiency of the development and on the resulting simulation's credibility.

When a legacy simulation is to be used, "early specification of the validation referent" means specifying the validation referent early in accreditation planning. Since, by definition, a legacy simulation already exists, specification of the validation referent occurs after simulation development.

Delay in specifying the validation referent until late in a new simulation development or modification (or until late in accreditation planning or execution for a legacy simulation) can introduce a variety of problems. For a new development or modification of a legacy simulation, such delay means that conceptual validation may have been performed using a different referent from validation referent to be used in assessing the implemented simulation. Opportunities for unnecessary problems and extensive rework abound in such conditions. Additionally, validation data sets need to be developed and a delay in specifying the referent limits an

early understanding of data requirements. This may result in insufficient time to build the required data set to support validation activities.

For both legacy and new/modified simulations, late specification of the validation referent can delay simulation accreditation, because needed information which could have been obtained by appropriate early planning is not available. Such delay in validation referent specification also can increase VV&A costs by not facilitating efficient V&V planning and execution.

Early specification of the validation referent for new simulation developments (and for simulation modification) normally occurs during the requirements or planning phases. Typically specification of the validation referent will occur in conjunction with development of acceptability criteria if the validation referent has not been specified earlier. It is also helpful to specify the validation referent in documentation of the conceptual model to ensure an appropriate focus and context for conceptual model validation, since the conceptual model should be validated against the same standard that will later be used for results validation.

Where the validation referent is specified may vary. In some cases, the validation referent will be documented by itself, with pointers to it from the Accreditation Plan, V&V Plan, and conceptual model. The DoD standard for documentation of VV&A does not use the term “referent,” but its templates for the Accreditation Plan, V&V Plan, V&V Report, and accreditation Report all include “Basis of Comparison” as Appendix C.[10] This appendix is where the validation referent would be described in those documents. In other cases, the validation referent may be documented in one of three places (with pointers to it in the other two): in the Accreditation Plan, the V&V Plan, and the conceptual model. In yet other cases, the validation referent may be documented all three places.

For accreditation of a legacy simulation, the validation referent should be specified in the Accreditation Plan. Depending upon the circumstances, the validation referent may also be described elsewhere, or there may be other products that point to the validation referent described in the Accreditation Plan.

Validation Referent Specification Dependence on M&S Type and Application

The spectrum of models and simulations in use within DoD is vast. Every M&S technology is used somewhere within DoD M&S: from desktop and hand-held devices to super computers, from single-computer simulations to large Live-Virtual-Constructive exercises involving many simulations and live forces, etc. Applications of DoD M&S include warfighting simulations (wargames, M&S-supported decision aids, simulations to explore tactics and strategy, etc.); all aspects of science, engineering, manufacturing, and logistics; education and training (from flight simulators to remote learning simulations); business simulations (planning, finance, personnel, etc.), and more. The guidance provided

in this Special Topic about validation referents is applicable to the whole spectrum of DoD models and simulations.

It has been suggested that validation referent guidance depends upon the M&S type and upon the M&S application. Five M&S type and five M&S application types were selected to demonstrate appropriateness of the guidance presented here. Insights gained from considering validation referent guidance for these ten situations give confidence that the guidance about validation referents in this Special Topic can be applied for any M&S type or M&S application within DoD.

The five M&S types considered were:

1) Models and simulations that make extensive use of adaptive programming
2) Models and simulations that involve human behavior representation
3) Distributed simulations
4) Models and simulations that make extensive use of aggregation
5) Simulations with system/hardware/software/people in the loop

The five areas of M&S application considered were:

1) Computational science and engineering applications
2) Engineering-level applications
3) Gaming and training applications
4) Military theater-level and campaign-level applications
5) Non-physical applications (such as economic simulations)

Each of these ten areas is discussed briefly with special emphasis on points related to their validation referents.

Validation Referent Dependence on M&S Type

Information needed for the validation referent can vary with the model or simulation type. This is illustrated by discussion of validation referents for five types of M&S. Guidance about validation referent identification, selection, and specification/description presented earlier applies to all types.

1. Models and Simulations That Make Extensive Use of Adaptive Programming

Many adaptive programming techniques are employed in modern simulation: artificial intelligence, knowledge-based systems, genetic algorithms, fuzzy logic,

agent-based simulation, neural nets, etc. These simulations are also referred to as complex adaptive simulations. Some of these adaptive simulations "learn". Some evolve and change the simulation structure and algorithms. Simulations that are based on adaptive programming can be designed to demonstrate how a person/system might behave given a set of circumstances where their actual behavior is unknown.

At present, there is no standard information that is considered the validation referent for such a simulation.

2. Models and Simulations That Involve Human Behavior Representation

Human variability and limited awareness of context factors that can impact human behavior create much uncertainty in human behavior representation, impacting the ability to define a validation referent. As indicated by the conclusion of a recent report[11]: "There is currently no common way to implement, measure and validate human behavior in models." The conclusions about human behavior representation validity in this report also note that, "Current constructive entities are often brittle, showing unrealistic behavior for even slight departures from the design space due to an over-reliance on simple, rigid rule sets and strict behavioral templates that capture standard operating procedures and idealized, purely doctrinal behavior. Validity on the constructive and predictive levels is often weak because of the prohibitive cost of extensive validation and because of limitations in experimentation with threatening conditions."

The difficulties in deciding what information to use for comparison with a simulation's performance show the importance of the validation referent specification explicitly addressing both the context of the information and its scope (which parts of the application domain of simulation's intended use are covered by the information). This illustrates the value of the five aspects of validation referent description (context, domain coverage, attribute distinctiveness, parameter uncertainty quantification, and information coherence) discussed earlier. Careful description of validation referent context and scope helps to ensure that an appropriate perspective is used when simulation results are assessed relative to the referent.

Validation referents described in accordance with guidance presented earlier will help the User and Accreditation Authority think through what the simulation is to be accredited for. Is human behavior representation to be validated at the level of individual behavior (such as the action of an individual soldier or crew member) or at the group level (such as the crew of a combat vehicle)? Often there is more reliable information about the behavior of a group than of an individual. Clarity in such matters can help to determine what information is pertinent and appropriate as a possible validation referent.

3. Distributed Simulations

The validation referent serves as the basis of comparison for the simulation results. This subsection discusses general principles involved in identifying and selecting an appropriate validation referent for a distributed simulation and makes clear that guidance presented earlier about validation referent identification and specification also applies to distributed simulations. Specific guidance may be available for a distributed simulation employing a particular paradigm. Such guidance for High Level Architecture (HLA) distributed simulations may be found in the VV&A Overlay for the Federation Development and Execution Process.[12] The Federation Referent is used in conjunction with development of acceptability criteria, in validation of the Federation Conceptual Model, and in validation of federation results. The material provides general guidance similar to that presented in this Special Topic, but does not provide detailed guidance about information that should be contained in the Federation Referent or its format. Ideas presented in this Special Topic and those about the Federation Referent are fully compatible. The Federation Referent has a special challenge in regard to combining information from different sources for the validation referent, i.e., how to combine information from validation referents for the federates that comprise the federation. The importance of credibility of the Federation Referent with the User/Sponsor of the federation is emphasized.

Validation Fundamentals for Distributed Simulations

Distributed simulation applications pose a special validation assessment problem in that they are intrinsically more complex than many unitary simulations because they involve both the interaction of multiple processes and the infrastructure that makes the interaction possible. A coarse assessment of simulation performance may be obtained if the results can be compared to a validation referent consisting of a previous “real-life” event, either from an operational event or in a test/training situation. Because of limited controls in real-life events, there is significant variability in data from live events. The intrinsic variability of real-life events places limitations on the ability of one or a few real-life events to serve as a referent. The variability must be addressed.

Problem decomposition is a basic procedure used in scientific methods. It can also be applied to the validation referent for a distributed simulation. Problem decomposition is a particular application of what some call “piecemeal validation” or, more specifically, the impact the validation of the distributed simulation’s components (e.g., federates) has on the validation on the overall distributed simulation (e.g., the federation). Simpson’s Paradox states that the potential for the direction of a conclusion from examining the pieces may be at odds with the direction of a conclusion from examining the whole.[13] Thus the validation of the components is necessary but not sufficient for validation of the overall distributed simulation. Simpson’s Paradox should be considered when defining the validation referent for a distributed simulation.

Because of the variability of real-life events, some prefer to decompose the federation into its functional components and to identify validation referents for individual components. Then each component can be validated individually. This permits sensitivity analyses and comparisons against theory and historical data to be performed more easily. At the functional level, results may be available from other simulations that have comparable or greater precision and that have a prior validation history. Also at the functional level, more options are typically available for selection as possible validation referents. Once the individual functions have been validated, their interactions among each of the simulation components must also be evaluated. The V&V Agent must follow the logical threads across the interfaces to ensure that the data being received is, in fact, what was expected. While many algorithms use the same terms and produce results with the same units, they may have underlying assumptions or approximations that unacceptably skew the results when they are utilized in subsequent processing.

The V&V Agent should seek to identify as many potential validation referents as possible for the distributed simulation. Once the list is assembled, each potential validation referent should be analyzed to ensure that it is adequate with regard to context, scope, uncertainties, etc. It helps to have multiple validation referents in order to ensure that the set of potential validation referents addresses all relevant capabilities and test objectives and incorporates the lowest uncertainty.

Insights gained from experiences with Live-Virtual-Constructive exercises have helped to refine methods for assessing their validity.

Validation Referents for Humans, Software, and Systems in Distributed Simulations

Validation referents for humans, software, and systems in distributed simulations are discussed below. Some aspects of human behavior representation were discussed earlier.

1) **Human Factors –** Probably the greatest uncertainty in simulation performance is encountered in the representation of human performance. The following should be noted:
    - The “average person” is not the sum of the “average” body parts. There are significant variations between the sexes; in addition to size, one finds differences in reach and mobility. These have great significance when modeling the physical properties of the human body; and when simulating its motions.
    - Differences in decision and reaction times have been observed between members of the Army, Navy, and Air Force performing the same task in operational situations in which all had completed the same training for the task and had comparable experience. This has been ascribed to differences in the underlying command structures and operational doctrines of the services. This is a specific example of context that should be described in the referent.

- Age, experience, outside distractions, and fatigue can cause significant variations in human performance. These factors must be taken into account in the simulation of Command and Control systems and in the representation of equipment operation.

When selecting a validation referent for human performance, the first criterion is that the data sets should be as representative of the simulated population as possible. The second selection criterion is that the task represented by the data set should be representative of the action to be taken – motion or decision. Note that there can be significant differences in timing between training and exercise conditions and actual operating conditions; however, as proficiency increases, the two values will converge.

In situations in which data sets are not available from tests, textbooks and industry consensus standards can be consulted for “typical” or “standard” values for similar tasks. These data may be used to calculate a referent value for the simulation. Care should be taken to ensure that the sources selected for the referent are independent from those selected by the M&S Developer for use in developing the simulation.

2) **Software** – Two forms of software can be used as a validation referent. First, a software package that has previously been accredited as representative of the system being simulated may be used as a validation referent or comparison standard for the simulation being validated. Second, a specialty or general-purpose software package may be used to construct a reference model of the system being simulated. In each case, accuracy and uncertainties associated with the software should be noted (as a standard part of referent description).

The use of accredited software packages for validation assessment is particularly applicable when complex systems are being simulated. An example of such a situation is the incorporation of a representation of the Navy’s Aegis Combat System or the Army’s Patriot Missile System into a federation. In such a case, a simulation provided by the program office could be used as the referent for validation assessment of representation of the system used in the federation.

For notional or developmental systems, reference simulations can be developed using commercial software packages. Examples of such packages are COMNET® for communications systems and Extend® for process modeling. Care should be taken when constructing reference simulations for use as validation referents. First, the precision and accuracy of the simulation must be sufficient to satisfactorily predict performance. Second, the simulation can only reflect the performances specified in the requirements and the characteristics identified to the point that the simulation was constructed. In this situation, the simulation, used

as a validation referent, really represents a second solution to the system being simulated. Accordingly, many uncertainties exist. The V&V Agent should attempt to identify them and, where possible, quantify them, for inclusion in description of the referent.

3) **Real Systems** – Often, an operational system is represented in a federation. When this occurs, data about the system from representative tests may be used in validation assessment of the system's representation. Care must be taken to ensure that the simulation setup representing the operational system is the same as that of the operational system at the time the data were taken. Test data about the operational system itself generally are the most authoritative source of material for the referent.

4. Models and Simulations That Make Extensive Use of Aggregation

Because there is seldom appropriate test data for items represented by extreme aggregation in a simulation, the standard validation referent for aggregation is based upon more detailed representations of the entities, processes, and phenomena represented in the aggregated simulation. These more detailed representations typically come from more detailed simulations. For example a campaign-level simulation may use results from an engagement-level simulation to determine the impact of countermeasure in an aircraft-on-aircraft engagement rather than dynamically representing the detailed engagement. The results of the engagement simulation would be "rolled-up" or aggregated at the campaign level, so results would be reported for a squadron instead of an aircraft. As noted by Bigelow and Davis,[14] "rolling-up" system performance does not always give better results in terms of aggregate-level behaviors.

In the context of aggregate-level representation, specification of the referent context is particularly important, since there are often differences between the context (assumptions, conditions, etc.) of the aggregated simulation and the context of the "feeder" simulations that can impact the appropriateness of the referent for the intended application.

5. Simulations with System/Hardware/Software/People in the Loop

Some validation referent considerations for this M&S type were addressed earlier in the subsection on Distributed Simulations.

Simulation can be used for understanding, describing or predicting the behavior of a single item or a population of items.

Example:

A flight simulator can be used to predict the likely performance of a particular pilot (or crew) in flying a particular aircraft. This is an example of a "single" item. If the same flight simulator is used to estimate performance of pilots (or crews) for a particular kind of aircraft, then it is an example of "items from a population" in the first sentence. In the second case, the personnel involved may not be a good validation referent. They be aces or duds or otherwise not representative of behavior and performance variations of the population of interest.

Thorough understanding of the population of interest, its characteristics and variations, is essential in determining what should be the validation referent for this M&S type. Typically that issue is ignored, and often a single item is blithely assumed to be representative of the population of the item. Consideration of characteristics and variations of the item is essential in determining what should be the validation referent for this M&S type.

Validation referent specification for this M&S type should indicate whether the intended use is for a single item or for items representative of a population. If it is a single item, then the relationship of that item to the validation referent should be clear. When the validation referent is the item itself, then variations in the context should be clearly delineated and the impact of the variations estimated (such as the "white knuckle" reality of real flight versus experience in a simulator, or the lack of vibration and acceleration forces in a missile seeker in a hardware-in-the-loop simulation vice one in actual missile flight). If items representative of a population are intended, then the validation referent should specify the characteristics and variations of the population and the relation of the items used in the referent to that population.

Validation Referent Dependence on M&S Application Category

Information needed for the validation referent can vary with the M&S application. However, guidance about validation referent identification, selection, and specification/description presented earlier applies in all cases.

1. Computational Science and Engineering Applications

The computational science and engineering M&S community is focused on experimental data (from particular tests) or in some situations from standard benchmark cases as referents for validation assessments. In that community, "benchmarks" are usually part of verification activities. "Benchmark solutions refer either to analytical solutions, i.e., exact solutions to the partial differential equations with the specified initial conditions and boundary conditions, or to highly accurate numerical solutions. However, we believe that in the solution of nonlinear partial differential equations or solutions with discontinuities or singularities the most reliable benchmark solutions are analytical solutions. In validation activities, accuracy is measured in relation to experimental data, i.e., our best indication of reality."[15] Some in this community reject use of the term validation for comparisons

with results from other simulations (no matter how accurate those simulation results are perceived to be) or for comparison with theoretical predictions. They prefer to restrict the term validation to acceptable comparison with experimental data.

The computational science and engineering community has produced the only college/graduate-level textbook on V&V.[7] General V&V materials and that for other communities typically are a chapter in a textbook, or just conference papers and journal articles. Two professional societies related to this community have produced V&V guides for M&S.[5, 6] Both guides stress use of data from special tests (validation experiments) as the most appropriate source for the validation referent.

2. Engineering Level Applications

Engineering level simulations tend to have resolution at the component or sub-component level and use algorithms that describe item characteristics and behavior at that level or higher. Typically the information used as the referent in validation assessments is drawn from specifications and requirements (for future systems or items), from tests and other data (for existing systems and items), or from higher resolution simulations (such as computational science simulations). Key aspects of validation referent specification are: identification of information sources, description of how information from different sources will be combined, quantification of variations and uncertainties in the information, and delineation of the information context.

3. Gaming/Training Applications

The validation referent for a simulation used in a gaming or training application depends upon the function of the application, i.e., its intended use. If the function is education, the validation referent probably can be very different from the validation referent when the function to train tank drivers in the physical tasks of driving a tank. For example, in a simulation supporting an educational endeavor, physical appearance of an object may be distorted cartoon-like to make lesson points clearer. Whereas in a simulation supporting tank driver training, realistic appearance and behavior of objects is important. Generally information sources suitable as possible validation referents for description of characteristics and behavior of physical systems represented in the simulation are the same ones which might be used as referents for other kinds of simulations except when the intended use dictates unrealistic appearance or behavior. Simulations used in games and training may mix human players/participants with computerized human behavior representation in the simulated environment, with human behavior representation generated by varieties of adaptive programming. Validation referent issues for simulations with adaptive programming, human behavior representation, and people in the loop are also relevant to gaming/training simulations.

Concern about validation of computer/computer-supported games has a long history. Hermann's conclusions about game validity published in 1967 still hold today: 1) validation is always a matter of degree, which means a simulation may be more valid by some criteria than by other criteria; (2) validation cannot be separated from the simulation's purpose; (3) given multiple validity strategies, several of the broadly applicable criteria may be applied in a particular phase of the simulation life cycle; and (4) human participation in the simulation significantly alters the required validation procedures. Hermann also noted that at that time only face-validation had been explored much.[16] Consequently, the simulations they were using were of unknown validity. As in other areas, the information that might be appropriate as a validation referent depends upon the situation, and often will involve use of SMEs.

It still appears that validation referent guidance presented earlier applies in this application domain as well as elsewhere: identification of possible validation referents as a function of the situation; selection of validation referents early in the process of M&S development/assessment; and specifying the validation referent in the Accreditation Plan (and perhaps elsewhere also) such that the validation referent is clearly and unambiguously designated using the five-fold aspect description presented earlier.

4. Military Theater-Level and Campaign-Level Applications

Two aspects of the problem make validation assessment of theater-level and campaign-level simulation one of the most vexing problems facing V&V personnel within DoD. The first is the inherent complexity of the system being modeled. The second is the number of uncertainties that must be dealt with. The consequence is that there are many problems in identification and selection of validation referents. Four basic methods are used to address these problems. Usually these methods are used in a combination of methods.

- One method is to compare results of the theater-level simulation with more detailed engagement-level simulations. Great care must be taken to ensure that the simulations employ the same context. Often the theater-level simulation and the engagement-level simulations are focused on different issues, which can preclude valid comparisons of their results.

- A second method is to disaggregate the theater-level simulation into operational or functional areas – logistics, weapon effects, maneuver, and so forth. Then the results of detailed simulations of each area can be compared with the performances found in the theater-level simulation. The validation referent (results from more detailed simulation) in this method is the same as in the previous method.

- The third method is to examine theater-level simulation performance in terms of the comparability of its results with other simulations of greater or lesser detail and with live-test and operational data.

- The fourth method is reliance upon SME judgment that the simulation's performance complies with the expectations of knowledgeable experts.

Validation issues associated with comparing or combining information from simulations with different levels of resolution have been studied extensively.[14] It is important to specify how the possible sources of information that will be used for validation of the simulation are going to be combined and applied.

Although simulation in this application domain presents serious V&V challenges, validation referents can be selected and specified in accordance with guidance presented earlier. The thorough description of the validation referent prescribed by the guidance in this Special Topic will help to ensure that the uncertainties and context of the validation referent are identified and that their potential impact on the validation assessment is clear.

5. Non-physical Applications

Non-physical applications, such as simulations related to economic theory, are different from the warfighting simulations widely used in DoD. The primary sources of validation referent information for non-physical applications are (1) theory related to the topic, or (2) historical experience or evidence related to the reality represented in the simulation. Regression techniques are frequently employed to show how simulation results correlate with the historical information. It is important that such historical information be decomposed in all relevant domains and that correlations be performed to ensure that the simulation can support the intended use.

Validation referents for simulations in this application domain can be identified and specified in accordance with the guidance provided earlier.

Validation Referent Approval

Because of the important role that validation referents play in accreditation assessment, the validation referent should be approved by the Accreditation Authority. This may be done as part of approval for the Accreditation Plan (which should specify the validation referent) or as a standalone item for approval by the Accreditation Authority.

Validation Referent Checklist

This section provides a checklist that can be used to ensure that a validation referent is properly identified and specified. It addresses the suggestions and guidance presented earlier in this Special Topic. The checklist is designed for those who identify and specify the validation referent.

1) Has guidance been given relative to validation referent selection and specification?

   a) From whom (User, Accreditation Authority, Accreditation Agent, V&V Agent, other authority)?
   b) What does the guidance address?
      - Validation referent selection
      - Validation referent description
      - Validation referent use
   c) Has guidance been followed? If not, why?

2) Are needed M&S information items available in acceptable form (intended use statement, requirements, acceptability criteria, Accreditation Plan, V&V Plan)? If not, how is needed M&S information to be developed to support validation referent identification, selection, and specification?

3) Is an adequate validation referent available? If not, has an appropriate decision been made about what should be done in the absence of an adequate validation referent?

4) Are multiple adequate validation referents available?
   a) What basis is used for selecting among them? Direction, convenience, economics, scope (portion of intended use domain covered by the possible validation referent information, best information available, broadest community acceptance, other factors), etc.?
   b) If more than one information source is used in the validation referent, how will information coherence be achieved?

5) Has the validation referent been specified appropriately with descriptions of context, domain coverage, attribute distinctiveness, quantification of parameter uncertainty, and information coherence?

6) Has validation referent dependence upon M&S type and kind of application been addressed explicitly?

7) Has expected validation referent use in validation and accreditation assessments been defined adequately?

8) Has the validation referent been developed in accordance with specific guidance for referents for that kind of M&S?

9) Are there any validation referent issues that need to be brought to the attention of the User, Accreditation Authority, Accreditation Agent, V&V Agent, or some other authority? How will the issue be brought to the attention of those who need to be aware of the issues?

10) Has the validation referent been approved by an appropriate authority (User, Accreditation Authority, Accreditation Agent, V&V Agent, other authority)?

11) Is the referent documented appropriately (e.g., in Accreditation Plan/Report and V&V Plan/Report as stated in MIL-STD-3022, as well as in the conceptual model and elsewhere)?

Validation Referent Use in Validation and Accreditation Assessment

This section differs from the preceding section in that it is concerned with using, rather than identifying and describing, the validation referent. It is included here because it emphasizes the importance of proper identification and description of the validation referent. Three basic situations are discussed:

1) Validation referents are data rich (the information about the simuland for the validation referent is both abundant and accepted as pertinent and reliable).

2) Validation referents are data poor (there is little information about the simuland for the validation referent, but data that do exist are recognized as pertinent and reliable).

3) The validation referent is poorly defined, poorly identified, or otherwise inadequate – the data that exist are not recognized as pertinent and reliable by all concerned.

Frequently the data quality of the validation referent varies across the simulation use domain. For part of the intended use domain, there may be data-rich information suitable to serve as part of the validation referent, but there will be other parts of the simulation use domain which are data poor or for which no reliable data exist. In such situations, the validation referent will have multiple parts, with appropriate characterization of the uncertainties of information in each part in the validation referent description and specification.

This section also discusses the important issue of the relationship of the validation referent used for validation and accreditation assessment to information used for M&S development and to information used as inputs in simulation runs.

1. Data-rich Validation Referents

A data-rich referent is always the most desirable referent situation, especially if uncertainties about parameters of the referent are well specified throughout the entire domain of the simulation's intended use. Such referents provide the basis for the most reliable and most credible validation assessments of simulation

results, since they provide an objective and factual basis for statistical comparison of results. However, even in situations with data-rich referents, care must be taken to ensure that the information used for the referent is truly pertinent to the simulation's intended use. For example, if human size is part of the referent (as might be pertinent for simulations concerned with passenger movement in a new vehicle design), one must be sure that the information about human size is recent, because humans are bigger now than they were a couple of generations ago (at least in America).

In a data-rich environment, it should be possible to separate the information used for the referent from information used to design and develop the simulation, so that the simulation's validity and capability to make reliable predictions can be assessed more robustly. The need for statistical independence between information used for M&S development and information used for validation assessment was mentioned earlier and is discussed below as well.

2. Data-poor Validation Referents

In a data-poor situation, the referent is by necessity mainly theoretical. The theory may be well articulated and explicitly formulated, as in an astrophysical simulation of processes insides stars, for which we have only inferred data, no *in situ* observations.[17] On the other hand, the theory may be unarticulated, as is often the case when SMEs are used as part of the referent. The SME may not make explicit how a judgment was reached or upon what evidence and logic the assessment is based. The credibility of fidelity and validation assessments based upon SME judgments in data-poor referents is always severely limited, but a thoughtful probing of SMEs (and good documentation of their insights) makes the best of such a situation.

In data-poor environments, it is seldom possible to separate information used for simulation development from information used for validation assessment. Some, especially the computational science and engineering community, would claim that this limits simulation assessment to calibration and prohibits validation of simulation predictive capabilities. It is recommended that the validation referent description note that validation referent information was also used for simulation development so that appropriate caveats may be associated with the validation assessments. This Special Topic does not attempt to prescribe how such caveats are determined, only that the lack of statistical independence between information used for M&S development and for assessment be noted in the validation referent specification so that appropriate caveats can be indicated with the assessment.

3. Inadequate Validation Referents

In the third case, the validation referent either cannot be clearly identified or has recognized inadequacies (such as contradictory information, so that if simulation results agree with some parts of the referent data, they will not agree with other parts). If there is no clear way to sort the validation referent information so that it is

coherent and non-contradictory, then it is necessary to clearly document the inadequacies.

However, as Hodges and Dewar[17] noted long ago, even a model that cannot be validated can have utility and value. In the data-poor situation discussed above, the use of theory or SMEs as the validation referent allows one to make fidelity and validation assessments, even though the credibility of such assessments may be low. In contrast, when the validation referent is contradictory, it is difficult to even make a low-credibility assessment. Inadequate validation referents challenge the professionalism of V&V personnel, since inadequate validation referents force them to declare to sponsors, Users, Accreditation Authority, and others that validation assessment is not possible because of validation referent inadequacies and the simulation could only be used with a significant level or risk.

Validation Referent Relation to M&S Development and Run Information

Information about the reality represented in the simulation is used to develop it, to run it, and as the referent in validation and accreditation assessments. It is reasonable to inquire about the relationships among these three sets of information since they all deal with the same reality.

In a data-rich environment, it is most desirable that the set of information used for M&S development, the set of information used as inputs to running the simulation, and the set of information used as the validation referent be statistically independent. This is particularly important when simulation results are used to predict how things will be in regions for which data are sparse or absent. The rationale is simple: this approach provides the highest likelihood that simulation results will be most representative of the reality represented in the simulation.

The medical community has found that conclusions from clinical trials can vary significantly from what is believed to be more correct if appropriate care is not taken to control such statistical issues.[8] Similar concern is needed in all simulation assessments. Many do not appreciate the issues and risks associated with using the same information for the validation assessment that was used for M&S development.

Concern about this general problem, and its specific related problem of not letting a developer have access to experimental outcomes before the simulation describing the same situation is built and run, is abundant in the computational science and engineering community. The need for the developer to know the exact conditions of the experiment before running the simulation is understood. Numerous guidelines are presented for the ways that “empirical adjustable parameters” – *knobs, dials, fudge factors* in more colloquial terminology – must be treated for simulation results to be acceptable in peer-reviewed circumstances (Roache[7]).

In data-poor environments, statistical independence among the three sets of data may be impossible. There simply may not be enough data to permit use of separate data for the validation assessment referent, given what is needed to develop and run the simulation. The three data sets (development data, run data, and assessment referent data) may even have to be identical. This condition limits what can be claimed about the validity of simulation results. One can describe with quantitative precision how well the simulation reproduces its input data, but one cannot make meaningful comments about simulation predictive capabilities. In that regard, one may be unable to make a stronger assessment than one can with an inadequate validation referent. Candor about limitations of validation assessments in such circumstances is an important aspect of V&V professionalism. The words of statistician George Box,[19] who at the time was the Past President of the American Statistical Association and President of the Institute of Mathematical Statistics, should not be forgotten: “all models are wrong, but some are useful.”

References

1. Department of Defense, *DoD Modeling and Simulation (M&S) Glossary* (Draft), March 2010.

2. Gross, David C., “Report of the Fidelity Implementation Study Group,” *Proceedings of the Spring 1999 Simulation Interoperability Workshop (SIW)*, paper 99S-SIW-167A, 1999.

3. Department of Defense, *DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A),* DoD Instruction 5000.61*,* December 2009.

4. Pace, Dale K., "The Referent in Simulation Development and Assessment," *Proceedings of the European Simulation Interoperability Workshop*, June 28-July 1, 2004, Edinburgh (Scotland), paper E04-SIW-024, available at http://www.sisostds.org.

5. American Institute of Aeronautics and Astronautics (AIAA), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA G-077-1998.

6. American Society of Mechanical Engineers (ASME), *Guide for Verification and Validation in Computational Solid Mechanics*, ASME V&V 10-2006, December 29, 2006.

7. Roache, Patrick J., *Verification and Validation in Computational Science and Engineering,* Hermosa Publishers (Albuquerque, NM), 1998. ISBN 0-913478-08-3.

8. Kunz, Regina, and Andrew D. Oxman, “The Unpredictability Paradox: Review of Empirical Comparisons of Randomized and Non-randomized Clinical

Trials," *British Medical Journal*, October 31, 1998, pp. 1185-1190, available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9794851&dopt=Abstract (accessed December 2010).

9. Pace, Dale K., and Jack Sheehan, "Subject Matter Expert (SME)/Peer Use in M&S V&V," Paper A4 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

10. Department of Defense (DoD) Standard Practice: *Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations*, MIL-STD-3022, 28 January 2008.

11. North Atlantic Treaty Organization (NATO) Research and Technology Organisation (RTO), *Human Behavior Representation in Constructive Simulation*, Technical Report TR-HFN-128, September 2009.

12. Institute of Electrical and Electronics Engineers (IEEE) Standard 1516.4™ (IEEE Std 1516.4™-2007), *IEEE Recommended Practice for Verification, Validation, and Accreditation of a Federation – An Overlay to the High Level Architecture (HLA) Federation Development and Execution Process*, 20 December 2007.

13. Pearl, Judea, *Simpson's Paradox: An Anatomy*, UCLA Cognitive Systems Laboratory, Technical Report (R-264), November 1999. Available at http://singapore.cs.ucla.edu/R264.pdf (accessed December 2010).

14. Bigelow, James H., and Paul K. Davis, *Implications for Model Validation of Multi-Resolution Modeling (MRM) and Exploratory Analysis*, RAND Corporation Report DRR-3045-AF, April 2003.

15. Oberkampf, William L., Timothy G. Trucano, and Charles Hersch, "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," Paper B1 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

16. Hermann, Charles F., "Validation Problems in Games and Simulations with Special Reference to Models of International Politics," *Behavioral Science*, Vol. 12, pp. 216-231, 1967.

17. Calder, Alan, Jonathan Dursi, Bruce Fryxell, and eight others, "Validating Astrophysical Simulation Codes," *Computer Science Engineering*, Vol. 6, Issue 5, pp. 10-20, September 2004.

18. Hodges, J. S., and J. A. Dewar, *Is It You or Your Model Talking? A Framework for Model Validation*, RAND Report, R-4114-AF/A/OSD, 1992.

19. Box, G. E. P., "Robustness is the Strategy of Scientific Model Building" in *Robustness in Statistics,* eds., R. L. Launer and G. N. Wilkinson, Academic Press, 1979, p. 202.

Acronyms

AIAA	American Institute of Aeronautics and Astronautics
ASME	American Society of Mechanical Engineers
CD	Compact Disk
DoD	Department of Defense
HLA	High Level Architecture
IEEE	Institute of Electrical and Electronics Engineers
M&S	Modeling and Simulation
M&S CO	Modeling and Simulation Coordination Office
MRM	Multi-Resolution Modeling
NATO	North Atlantic Treaty Organization
RPG	Recommended Practices Guide
RTO	Research and Technology Organisation
SISO	Simulation Interoperability Standards Organization
SIW	Simulation Interoperability Workshop
SME	Subject Matter Expert
UCLA	University of California at Los Angeles
V&V	Verification and Validation
VV&A	Verification, Validation and Accreditation

The appearance of hyperlinks does not constitute endorsement by the DoD, M&S CO, the administrators of this website, or the information, products, or services contained therein. The DoD does not exercise any editorial control over the information found at these locations. Such links are provided consistent with the stated purpose of this website.