Generic Methodology for Verification and Validation for Training Simulations

Manfred Roza
TNO Defence, Security and Safety
Soesterberg, The Netherlands
Manfred.Roza@tno.nl

Jeroen Voogd
TNO Defence, Security and Safety
The Hague, The Netherlands
Jeroen.Voogd@tno.nl

Martijn van Emmerik
TNO Defence, Security and Safety
Soesterberg, The Netherlands
Martijn.Vanemmerik@tno.nl

Ad van Lier
Simulation Expertise Centre, Netherlands MoD
The Hague, The Netherlands
Ac.vanlier@mindef.nl

ABSTRACT

Modelling and simulation (M&S) is increasingly being exploited as an enabling technology to support tactical, operational and strategic training objectives within the military domain. Verification and Validation of M&S assets are intended to ensure that only correct and suitable training results are obtained thereby facilitating risk management with regards to M&S use for military training purposes.

The Generic Methodology (GM) for Verification and Validation (VV&A) is a generic and comprehensive VV&A methodology for acceptance of M&S assets. The GM-VV is currently in the processes of standardisation within the Simulation Interoperability Standards Organization (SISO), and at the same time under consideration by various national defence directorates (DoD, MoD, etc) to be incorporated as part of their M&S policies. The GM-VV provides a framework to efficiently develop arguments to justify why M&S systems are acceptable for a specific intended use or not. This argumentation is intended to support stakeholders in their risk-based decision-making process on the development, application and reuse of such M&S systems.

This paper presents the technical substance of GM-VV in detail, gives a status update on the standardization process along with its relationship to other VV&A standards and practices. GM-VV is a generic methodology, based on sound system engineering principles and likewise needs to be tailored for specific M&S application domains. The paper illustrates how GM-VV can be tailored for and applied to training simulations. This illustration is based on some results and lessons learned from several currently running (joint) technology demonstration programs of the Dutch, Swedish, German and French MoD's inside NATO NMSG-073 context.

ABOUT THE AUTHORS

Manfred Roza is a member of the scientific staff at the Training Innovations Department of TNO Defense, Security and Safety. He holds a M.Sc. and Ph.D. in Aerospace M&S Technology from the Delft University of Technology. Besides VV&A his research interests include modeling and (distributed) simulation, human-factors and training systems. Currently Manfred is the chief editor of the GM-VV product development group inside SISO and technical lead for NATO NMSG-073 task group on GM-VV case-studies and associated recommended practices.

Jeroen Voogd is a member of the scientific staff at the M&S and Gaming Department of TNO Defence, Security and Safety Division. He holds a Ph.D. in Computational Physics from the University of Amsterdam in the field of modelling and simulating of biophysical systems on parallel and distributed computing platforms. A recurring theme in his work of the last years is the quality of simulations. This includes VV&A of M&S assets, as well as quality assurance within TNO.

Martijn van Emmerik is a member of the scientific staff at the Training Innovations Department of TNO Defense, Security and Safety. He holds a M.Sc. in Human Factors Psychology (Utrecht University) and a Ph.D. in Social and Educational Sciences from the University of Twente. His research interests include transfer of training, vehicle simulator training, cybersickness, serious gaming, and validity.

Ad van Lier is a member of the staff at the Simulation Expertise Centre of the Royal Netherlands Armed Forces. . He holds an M.Sc in Information Technology (University of Liverpool). He is tasked with standardization, interoperability, VV&A and re-use of M&S. Areas of personal interest focus on intelligent training within M&S systems.

Generic Methodology for Verification and Validation for Training Simulations

Manfred Roza
TNO Defence, Security and Safety
Soesterberg, The Netherlands
Manfred.Roza@tno.nl

Jeroen Voogd
TNO Defence, Security and Safety
The Hague, The Netherlands
Jeroen.Voogd@tno.nl

Martijn van Emmerik
TNO Defence, Security and Safety
Soesterberg, The Netherlands
Martijn.Vanemmerik@tno.nl

Ad van Lier
Simulation Expertise Centre, Netherlands MoD
The Hague, The Netherlands
Ac.vanlier@mindef.nl

INTRODUCTION

The Generic Methodology for Verification and Validation (GM-VV) was developed in an international project called Common Verification, Validation and Accreditation Framework for Simulation (REVVA) with government participation from France, The Netherlands, Sweden, Canada and Denmark, and industry partners from Great Britain. This cooperative effort aimed at the development of a uniform and generic framework for verification and validation (VV&A) of models, simulations and data, which are (to be) shared between these nation's defense organizations. The GM-VV is currently in the processes of standardization within the Simulation Interoperability Standards Organization (SISO), and at the same time under consideration by various national defense directorates (DoD, MoD, etc) to be incorporated as part of their M&S policies.

GM-VV is a generic methodology which means that it is defined independently from any specific M&S application, domain or technology. This makes the methodology generally applicable and compatible to any class of VV&A problems inside the M&S domain. However, this also makes GM-VV an abstractly defined methodology that has to be instantiated, specialized, extended and optimized for such a particular M&S application or technology. Metaphorically speaking, GM-VV can be compared to an alphabet, which is a symbolic representation to write words independent of language. Each language, however, has special rules for combining those symbols to form meaningful units (words) and many have extended the set of symbols, or optimized it for their specific purpose (think of characters like the German 'ß', or the use of accents to change pronunciation: à, ç, č, or ñ). Furthermore, the rules for combining these symbols are different in each language (and are not specified in the alphabet itself).

This paper starts with an introductory overview of the GM-VV basic principles, concepts, methodology components and their interrelationships. Next the paper presents the SISO product nomination, its current status and position with respect to other VV&A effort and standards. The paper continues with presentation how GM-VV may be tailored for VV&A of training simulations. This part provides initial recommended practices to help instantiate, extend and optimize GM-VV for training simulations. This effort is illustrated with some results and lessons learned from several running (joint) technology demonstration programs of the Dutch, Swedish, German, and French MoD's.

GM-VV BASIC PRINCIPLES & CONCEPTS

Systems Engineering Approach to M&S

GM-VV adopts a systems engineering approach to M&S and associated VV&A practices. This means that within GM-VV models and simulations are considered to be systems of systems that have a life-cycle and are subjected to system engineering practices. Moreover, in GM-VV models and simulations are considered to be part of a larger context, called the *frame system*, in which they are used. From this perspective M&S is just a system engineering specialization, and the same holds for VV&A practices. Therefore the methodology design is compliant, builds-upon and complements well-established systems engineering and other international standards (ISO/IEEE 15288, ISO 15026, ISO/IEEE 12207, etc.). GM-VV can thus easily be fit in or used as an extension to any existing systems engineering methodology or paradigm.

M&S Based Problem Solving Approach

The basic premise of GM-VV is that models and simulations are always developed and employed to help fulfill specific needs of their users and other stakeholders (trainers, analysis, decision makers, etc.). GM-VV assumes that VV&A always takes place within such a context and uses a four-world view to structure this larger context (Figure 1).

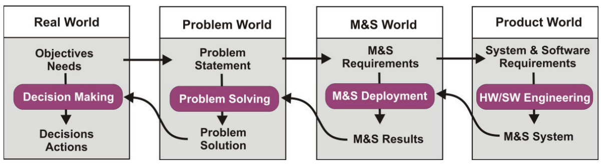

Figure 1: M&S Based Problem Solving Approach

In the first of these four worlds, the real world, certain needs may arise. Once the objective is set in the real world to address these needs, these needs translate into a problem statement, which has to be solved in the problem world. The M&S world is the way to help solve this problem by first setting M&S Requirements for an M&S System. The M&S system itself is the result of hard- and software engineering within the product world.

The M&S deployment consists of well-controlled executions of the M&S system, i.e. simulations, from which M&S results are obtained. These results are used in the problem solving process of the problem world. Finally the problem solution is transferred to the real world where it is deployed in the operational environment. When the work has been done properly the resulting effects will ultimately fulfill the needs.

Compared to standard system engineering approaches, the M&S world is specific to M&S system engineering.

VV&A Problem Solving Approach

GM-VV considers VV&A as a separate problem domain with its own specific needs, objectives and issues. This domain is referred to as the VV&A world (Figure 2) and is defined in parallel to the previously mentioned worlds. The primary objective of the VV&A world is to support stakeholders in making acceptance decisions on using M&S artifacts in the problem world. The major question in such decisions is whether the components of the M&S system have capabilities that collectively satisfy the real-world objective.

In general it is not possible to demonstrate with absolute certainty that the objectives will be satisfied. Consequently there is always a probability that the M&S based solution fails to satisfy the real world need and its use results in an undesirable impact, i.e. there is a risk.

The acceptance decision should only be positive if the stakeholder has sufficient confidence in the claim that an M&S artifact will satisfy the real world needs without posing unacceptable risks. The stakeholder decision making process therefore requires well-informed arguments to justify that claim. This stakeholder is the *VV&A problem owner*.

VV&A World objective is the delivery of a stakeholder oriented argumentation for an acceptability claim: the *acceptance recommendation.* Although the acceptance decision procedure is an input for developing the acceptance recommendation, the *VV&A intended use*, its actual execution is in the real world and thus outside the scope of the VV&A world.

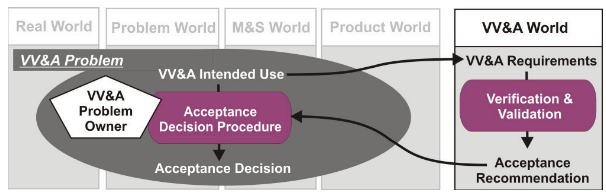

Figure 2: VV&A Problem Solving Approach

Acceptability Criteria Satisfaction

The *acceptance goal* is to convincingly show that an M&S system will satisfy its purpose in use. This abstract acceptance goal must be translated into a set of necessary and sufficient acceptability criteria that are concrete. For these criteria convincing evidence must be obtained. GM-VV defines three classes of acceptability criteria for M&S artifacts, called VV&A *properties* (Figure 3) that each addresses and provides a set of assessment metrics for a specific part of an M&S artifact:

- *Utility* properties are used to assess the effectiveness, efficiency, suitability, and availability of an M&S artifact in solving a problem statement in the problem world. Utility properties address aspects such as value, risk and cost.
- *Validity* properties are used to assess the level of agreement of the M&S system replication of the real world systems it tries to represent i.e. fidelity. Validity properties are also used to assess the consequences of fidelity discrepancies on the M&S system utility.
- C*orrectness* properties assess whether the M&S system implementation conforms to the imposed requirements, and is free of error and of sufficient precision. Correctness metrics are also used to assess the consequences of implementation discrepancies on both validity and utility.

Besides the aforementioned properties GM-VV defines a special class of properties, *Meta*-properties. These meta-properties are used to assess the level of confidence with which the utility, validity and correctness have been assessed, i.e. the convincing force of the evidence for these three properties. Meta-properties typically include aspects such as completeness, consistency, independency, uncertainty and relevance.

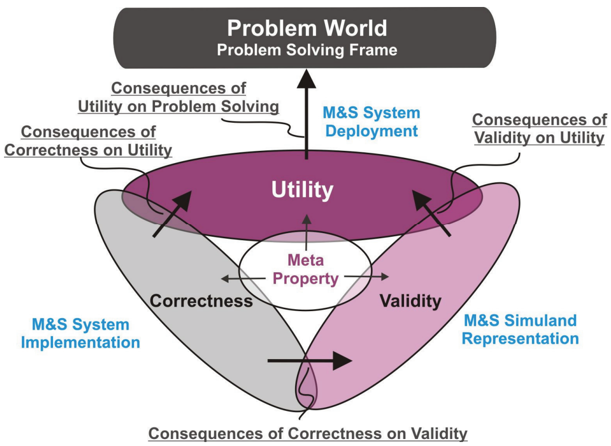

Figure 3: Utility, validity, correctness & meta-properties relationship diagram

Goal and Evidence Based Reasoning

VV&A problem owners should understand and have a clear picture of the whole VV&A problem before attempting to solve it or make an acceptance decision. Particular for large-scale and complex M&S system, or in case M&S based solutions are used in safety critical real-world environment, this may involve identification and definition of many interrelated acceptance criteria, sub-criteria, sub-sub-criteria, etc. Similarly, many items of evidence may be elicited to proof the satisfaction of these acceptance criteria, with varying convincing force and may even be contradicting (i.e. counter evidence). To effectively and efficiently address these issues GM-VV defines two major techniques which are based on argumentation network based reasoning: the VV&A goal network and the VV&A claim network.

The *VV&A goal network* is an argumentation network based on sound goal-oriented requirements engineering principles. The network is a directed graph consisting of goals, strategies and supporting contextual information for transforming the *acceptance goal* (AG) into a set of well reasoned, traceable and auditable *acceptability criteria* (AC) along with evidence solutions (Figure 4). *Evidence solutions* (ES) specify what and how evidence must be elicited in order to obtain convincing evidence for each of the specified acceptance criteria. An evidence solution comprises specifications for the M&S system test method, referent or expected results generator, results comparator and evaluator. When implemented they form the *VV&A experimental frame*.

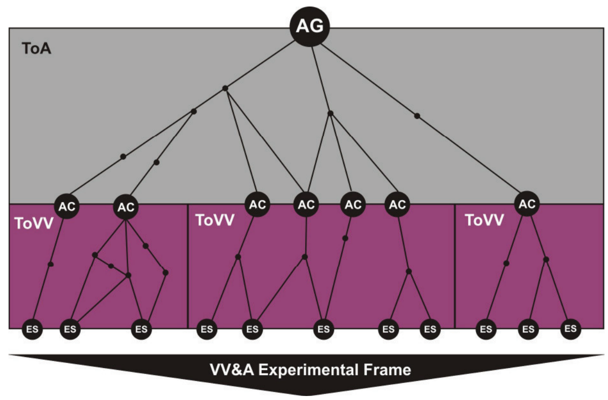

Figure 4: Schematic VV&A Goal Network

The *VV&A claim network* is an evidence-based argumentation network for transforming any collected *items of evidence* (IoE) into a well reasoned, traceable and auditable set of *acceptability claims* (AbC) on the satisfaction of each of the specified acceptability criteria (Figure 5). These acceptability claims are further aggregated by argumentation into a claim on whether the M&S system as a whole is acceptable, the *acceptance claim* (ACC). Furthermore, this network provides the means to reason on and express the level of confidence that can be placed on each of these claims. The VV&A claim network is also a directed graph but now consisting of items of evidence, claims and arguments. The VV&A goal network serves as a basis for developing the VV&A claim network.

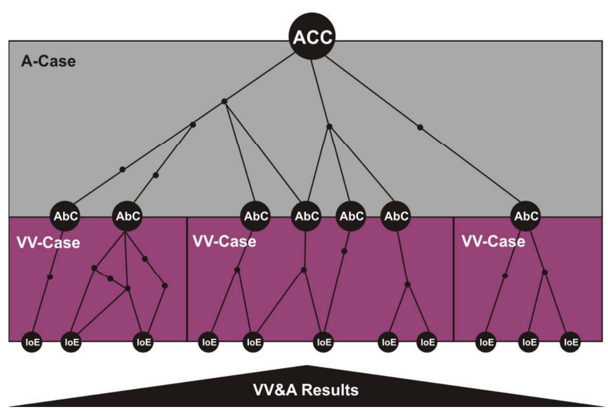

Figure 5: Schematic VV&A Claim Network

VV&A Business Approach

As indicated by the 4 world model, GM-VV considers VV&A in a larger context and in that setting provides products and services. The organization of the VV&A work uses existing organizational concepts to establish VV&A business enterprises [IEEE15288, etc.]. GM-VV distinguishes two agency types:

- *VV&A acquirer agency:* a customer organization that acquires VV&A products or services
- *VV&A supplier agency*: a production organization that develops VV&A products and services.

Military organizational units responsible for the acquisition of M&S assets are examples of VV&A acquirer agencies. A VV&A acquirer agency tasks the VV&A supplier agency to provide products and/or services that solve the VV&A problem (Figure 2). The contract between the agencies is the *VV&A agreement*. These agencies can be two independent organizations but also be different units within a single organization.

From the VV&A acquirer agency perspective GM-VV provides mechanisms on how to best acquire and embed VV&A products and service in the organization structure, policies, M&S related projects, etc.
From the VV&A supplier agency perspective corresponding mechanisms are provided. On project level additional concepts are defined such as *VV&A project*, *plan, report* and *project memory* to manage the *technical* VV&A work in project environments. On VV&A enterprise level concepts such as *VV&A cost model*, *maturity model* and *enterprise memory* are defined to facilitate VV&A optimization, quality improvement, knowledge reuse and management.

Tailoring
Although similarities may exist, each VV&A project is unique. The VV&A project organization and its technical execution should fit the particular acceptance problem at hand, within context and within the VV&A project's enterprise environment. To achieve this, GM-VV offers an intrinsic and systematic adjustment approach, called *tailoring*. Tailoring in GM-VV means instantiating the methodology on enterprise, project and technical level, by means of *expansion*, *specialization* and *optimization*.

Expansion means integrating supplemental components with GM-VV. Usually, expansion is needed to comply with existing enterprise standards, regulations and policies. Specialization means extension of generic GM-VV components by providing domain specific implementations or details such as methods, tools and techniques. Both expansion and specialization must be performed in accordance with the GM-VV conformance rules, which guarantee consistency, and coherence of the applied method.

Tailoring by optimization means seeking the optimal cost-benefit-ratio for how the VV&A project solves the VV&A problem (Figure 6). Optimization may require expansion and specialization.

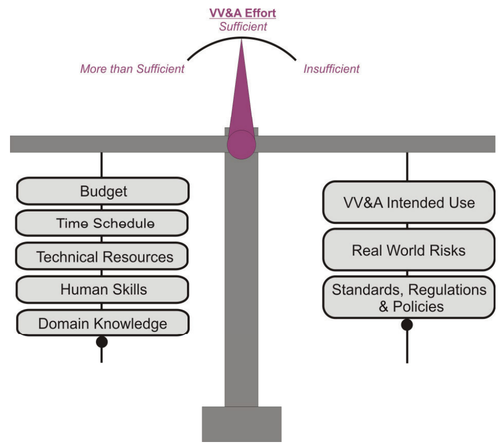

Figure 6: The VV&A Effort Optimization

In optimally solving the VV&A problem three major boundary conditions on the right side of the balance must be satisfied. First, the VV&A intended use must be achieved. Second, sufficient confidence must be established such that no unacceptable real world risk is incurred when the M&S based solution is used. Practically this means avoidance of *VV&A type 1 errors* (accept the M&S based solution while it should have been refused) and *VV&A type 2 errors* (refuse the M&S based solution while it is acceptable). Third, the VV&A solution may have to comply with standards, policies and regulations which are imposed on the project.
On the left side are the VV&A project resource variables that can be chosen to establish an optimum. These include budget, time schedule, technical resources (infrastructure, methods, tools and techniques, etc.), VV&A practitioner skills and available domain knowledge.

The right side of the optimization balance determines the level of rigor and detail with which a VV&A project must be executed to provide an acceptance recommendation that satisfies the VV&A problem. Optimization balancing is done prior to and during VV&A project execution. Optimization is also an important mechanism in assessing the feasibility of a VV&A project prior to its execution in the agreement phase between VV&A acquirer and supplier agencies.

GM-VV ARCHITECTURE AND COMPONENTS

The GM-VV methodology architecture builds around a three pillar design. GM-VV assumes that the overall

VV&A goal is satisfied by the right VV&A products and not directly by processes or organizations (Figure 7). The components of the GM-VV product pillar defines all the essential VV&A products that must be developed during the VV&A project, and those that must be kept up-to-date to sustain a viable VV&A enterprise environment. The GM-VV process pillar defines all VV&A life-cycle processes that collectively produce and modify the VV&A products as defined in the GM-VV product pillar. The GM-VV organization pillar defines all necessary organizational components to setup enterprise and project organizations.

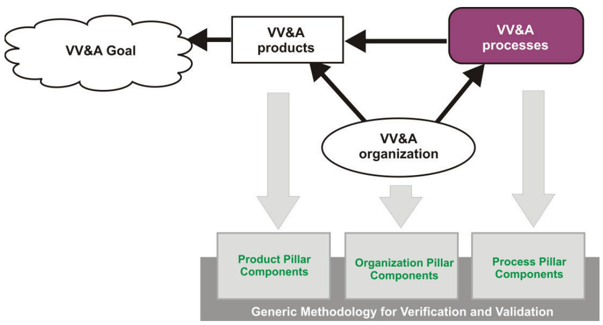

Figure 7: GM-VV three pillar design

Product Pillar Components

The products in this pillar are organized into enterprise, project and technical level. The GM-VV products are:

Agreement Products
1. VV&A Contract

Enterprise Products
1. VV&A Cost Model
2. VV&A Maturity Model

Project Products
1. VV&A Project Plan
2. VV&A Project Report

Technical Products
1. VV&A Requirements Specification
2. VV&A Preconditions Specification
3. VV&A System of Interest
4. VV&A Experimental Frame
5. VV&A Results
6. VV&A Goal Network
7. VV&A Claim Network
8. Acceptance Recommendation

All information underlying these products is managed in two *enabling products*:

1. VV&A corporate memory
2. VV&A project memory

The above listed products are the minimal set of products that must become available during a VV&A project or in the enterprise to conform to the GM-VV. As a result of the tailoring the content of some of the products can vary from very little to a huge amount of information.

Process Pillar Components

The process pillar describes all processes related to the life-cycle of the VV&A products. The life-cycle processes collectively deliver the VV&A products of the GM-VV product pillar. The GM-VV life-cycle processes can be considered as an instantiation of and extension to standard system life-cycle process models. The GM-VV defines the following sets of processes:

1. *Agreement Processes* needed to establish a (sub)contractual agreement between a VV&A acquirer and supplier agency.
2. *Enterprise Processes* needed to establish and maintain a VV&A supplier agency in which VV&A projects can be run.
3. *Project Processes* needed to establish and maintain a VV&A project environment in which the technical processes are conducted.
4. *Technical Processes* needed to do the actual technical VV&A work in order to produce the acceptance recommendation

Acceptance problem based tailoring determines the rigor and timing with which these processes are executed for a specific VV&A project.

Organization Pillar Component

This pillar describes all necessary organizational components to create a VV&A organization model. The organizational pillar provides two sets of components:

VV&A agencies specifications
1. VV&A supplier agency
2. VV&A acquirer agency

VV&A roles specifications
1. VV&A problem owner
2. VV&A sponsor
3. VV&A enterprise manager
4. VV&A project manager
5. Acceptance leader
6. VV&A leader
7. VV&A implementer

These organizational components are specified in terms of a name, responsibilities and obligations. The obligations are also found as activities and tasks in the process pillar. Individuals may play multiple roles, and multiple instantiations of roles and agencies may exist.

GM-VV STANDARDISATION EFFORT

SISO Product Nomination for GM-VV

In 2006 the Simulation Interoperability Standards Organization (SISO) formally approved the GM-VV product nomination by installing a GM-VV drafting group (DG) and product development group (PDG). The GM-VV product nomination consists of three interrelated documents (Figure 8):

- *GM-VV Handbook* provides an introduction to the methodology concepts and components. It also provides guidance on how to deploy and apply GM-VV within organizations to establish, execute and support VV&A projects.
- *GM-VV Recommended Practice Guide* (RPG) provides the VV&A life cycle processes, activities and tasks to develop the GM-VV products. This includes guidance on tailoring and strategies to develop goal and argumentation networks.
- *GM-VV Reference Manual* provides an extensive reference description for the GM-VV concepts and components.

These documents are organized such that one should start with the GM-VV handbook. From there the reader is pointed to the RPG for practical implementation hints, tips and other how to's. The reference manual is referred to whenever a deeper understanding of GM-VV is required.

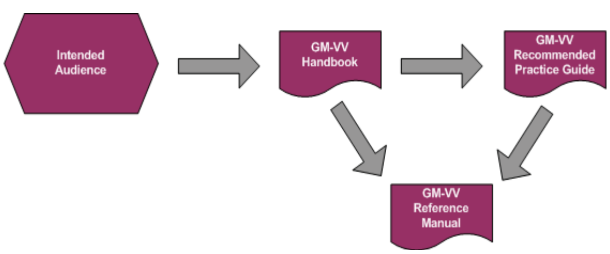

Figure 8: GM-VV Product Nomination Outline

Since the GM-VV reference manual is a companion reference document to the standard this document is not officially balloted like the other two documents. Currently, the GM-VV handbook will enter its fourth comment round within the SISO PDG community. The GM-VV RPG and reference manual will enter their third comment round. It is expected that the comment rounds will finish before end of 2010. After that the GM-VV handbook and RPG will be put forward for official balloting in 2011.

GM-VV and other VV&A Standards or Practices

The key objective of GM-VV is to offer the international M&S community a methodology for VV&A that encompasses a wide range of M&S technologies and application domains. Another objective is to provide this community with a common VV&A language, concepts and components to better facilitate the communication and co-operation within VV&A projects and between agencies. Therefore, the GM-VV attempts to leverage and harmonize existing well established international VV&A standards and practices into a single general applicable approach for VV&A. As result the GM-VV is rooted in and builds upon on contributions of the Euclid project REVVA [REVVA, 2009], IEEE 1516.4 Standard on the VV&A overlay on the FEDEP, ITOP 7.2 working group procedure for VV&A, and efforts within DoD DMSO (VV&A RPG, Templates, etc.). The major additions GM-VV makes to these existing VV&A methods are:

- Detailed VV&A product definitions with clear semantic specifications
- Introduction of a business model for VV&A and associated organizational components specifications
- Pragmatic goal and evidence based reasoning techniques for VV&A
- VV&A process model and process component specification that are applicable across M&S development paradigms and technology
- Project and corporate memory that support VV&A information and knowledge exchange, consolidation, reuse, and development within VV&A projects and enterprises
- Direct support for the development of CASE tools for VV&A to facilitate the cost-efficient implementation and execution of VV&A

The existing VV&A standards and practices are considered to be specializations of the GM-VV, and can be improved by integrating some of the new technical additions GM-VV offers. GM-VV also provides the referential basis for development of future M&S problem and application specific VV&A methods, tools and techniques. If built on the same basis and utilizing the GM-VV tailoring mechanisms (expansion and specialization), they will be interoperable with each other on (re)use of VV&A results, information and knowledge level.

GM-VV and NATO MSG-073

In 2009 NATO chartered a task group on GM-VV: NMSG-073. In this task group the MoD's of the Netherlands, France, Sweden, Germany, Canada and

Turkey work together to support the SISO PDG. Key objectives of NMSG073 are to refine and consolidate the methodology, to develop a body of recommended practices, to build support tools and techniques, and to provide a basis for training material. To achieve these objectives, a serie of practical case-studies is executed, either collectively or on a national level as part of MoD's own technology demonstration programs. These case-studies vary in nature but have a strong focus is on training simulation.

VV&A FOR SIMULATION BASED TRAINING

The Dutch MoD together with *Netherlands Organization for Applied Scientific Research* - TNO initiated at the end of 2009 a two year technology demonstration program that is focussed on the application of GM-VV to training simulation applications. This program contributes to a driving simulator case-study which is executed in NMSG-073. On a national level the program runs a VV&A case-study in support of the development of a moving-base bridge simulator for the Dutch navy. This simulator is intended for training and evaluation of heavy weather ship handling doctrines. Based on the first results and lessons learned from these ongoing case studies, initial recommended practices to help instantiate, extend and optimize GM-VV for training simulations are presented in the next sections.

Defining a Training Simulation Context for VV&A

Many simulator-based training development methods, such as MASTER, are strongly rooted in system engineering practices [Farmer et.al, 1999] and they all have key elements in common, see Figure 9. Since the GM-VV is also based on system engineering a relatively simple mapping is possible of simulation-based training on the GM-VV world views (Figure 1) [Roza et.al., 2009].

Figure 9: Simulation-Based Training System

In the real world training needs lead to the specification of the training problem, objectives, requirements and constraints for the design of the training program in which the training simulation will be employed. In the problem world the training program is the frame system that specifies how the training simulation is employed and how its results are utilized to solve the training problem. The outcome of a training program is the composite observable named training transfer.

The simulation-based training system is the aggregation of both the training program, which includes the trainees, and the training simulation. This training simulation maps to the GM-VV M&S world, while the actual development or acquisition of the training simulation maps to the GM-VV product world.

The training problem itself is derived from some operational need in the real world. The reason to use simulation based training usually is that real-life training is too costly, unfeasible or unsafe. In case the training transfer is insufficient there are real world consequences such as: damage to equipment, personal accidents or even loss of lives, collateral damage to environment, etc. This is the start-point for risk assessment required for tailoring of the VV&A project by optimization.

Concurrent and post-hoc VV&A

Simulation-based training systems should undergo VV&A during development to assure that the right training simulation is delivered and developed in the right way. This is called *Concurrent VV&A*. The Dutch navy bridge simulator VV&A case study is an example of such a concurrent VV&A project.

Post hoc VV&A is applicable or sometimes even mandatory when training needs changes due operational or doctrine changes, reuse of the training simulation in a totally different context, etc. The above mentioned NMSG-073 case study is an example of post-hoc VV&A. Here an existing driving simulator is assessed to determine if the simulator is acceptable for a road accident research study.

GM-VV Expansions for Training Simulations

Expansion (integrating supplemental components) is one of the three ways of tailoring GM-VV. An example of this kind of tailoring applied to training simulations is found in the Dutch navy bridge simulator case-study [Roza et. al., 2009]. Here the GM-VV life-cycle process for defining VV&A requirements is expanded by integrating and interfacing an existing method that allows for systematic specification of need statements (i.e. requirements) for simulation based training systems. This method is called SLIM (Specification of Learning means in an Iterative Manner) and has been developed for the Royal Netherlands Army who use it in their acquisition processes for training simulations. SLIM comprises an iterative process that is executed in a workshop setting, and ensures that all needs are developed in a well balanced manner [Verstegen et. al., 1999].

When SLIM is used for specifying training need statements it is an implementation of the GM-VV activity of providing the VV&A precondition specification. Since defining VV&A requirements is also a form of needs statement development, a tailored version of the SLIM method can facilitate the GM-VV VV&A requirements definition process as well. Both applications of SLIM are elaborated upon below.

SLIM Expansions to VV&A Precondition Provision
The VV&A precondition specification is a standard GM-VV product that specifies all prerequisite contextual information for a VV&A project. VV&A preconditions typically comprise information about the M&S system intended use, use risks, requirements, constraints and the M&S system (architecture, design, models, performance capabilities, etc.). The GM-VV activity *Provide VV&A Preconditions* can be expanded by integrating the next three additional tasks:

1. *Monitor the SLIM session.* The VV&A problem owner and acceptance leader join in the SLIM session on the backbench and collect information for the development of the VV&A precondition specification.
2. *Verify Existing Training Simulation Need Statements.* A shortened SLIM session is executed to check the existing need statement for correctness, completeness and consistency.
3. *Complete Non-Available Training Needs Information.* A SLIM session is executed to develop the lacking information such as training program, training objectives and risks, training simulator requirements, etc.

SLIM Based VV&A Requirements Definition
The activities and tasks as provided by the GM-VV for the *VV&A requirements definition* process do not impose a specific method of execution. This process can internally by expanded by organizing and executing SLIM moderated workshops. This means the integration of a transversal activity *Make group decision and assess feasibility of the VV&A requirements* with the existing GM-VV activities. In this activity training simulation requirements and use risk, time schedule, available resources, budget, cost and other aspects related to the VV&A project execution are assessed in relationship to the VV&A requirements specification for adequate balance and feasibility. Based on these outcomes collaborative decisions will be made on the VV&A requirements specifications and development process. The number of iterations and the duration of the process largely depends on complexity of the VV&A problem of the training simulation, associated use risks and available resources.

GM-VV Specializations for Training Simulations
Examples of tailoring GM-VV by specialization (extension of generic GM-VV components by providing domain specific implementations or details) for training simulations come from the work on the NATO driving simulator case-study. GM-VV specializations for training simulations focus on two parts of the argumentation network. First, it focuses on instantiations of the generic VV&A properties and associated acceptance criteria for *utility*, *validity* and *correctness* (Figure 3). And second, GM-VV specialization focuses on training simulation specific implementations of the generic *evidence solution* structure provide by GM-VV goal network (Figures 4 and 10).

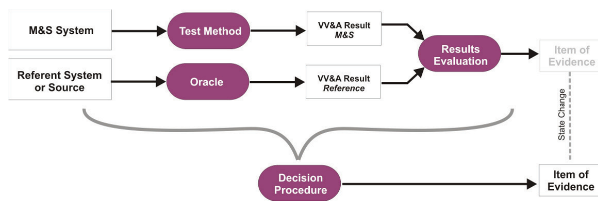

Figure 10 GM-VV Evidence Solution Structure

In the domain of training simulation, more than in any other domain, validation is not just driven by 'the degree to which the simulator accurately represents the real world'. At "trainee proficiency" is the desired output of the simulation based training system rather than the real word representation, see Figure 9. It is the Transfer of Training (ToT) that determines the utility. Positive ToT implies that trainees who practised with the simulator can reduce the amount of training on the real system to master the task. Unfortunately the opposite also occurs: negative ToT. Common utility properties for ToT are quantified by [Farmer, 1999]:

1. <u>Percent Transfer Measure (PTM)</u>: *PTM = 100[(Tc-Te)/Tc],* where *Tc* refers either to the time in training to a proficiency criterion outside the simulator. *Te* is the remaining time on the operational equipment after training in the simulator
2. <u>Transfer Effectiveness Ratio (TER)</u>: *TER = (TE_sim-TE+sim)/TE-insim*, where: *TE_sim* is the training effort needed to learn the task without the use of a simulator. *TE+sim* is the amount of time to learn the task with some training in the simulator. *TE-insim* is the amount of effort required to learn the task in the simulator.

3. Training Cost Ratio (TCR): *TCR = cost of operating training simulator / cost of operating actual equipment for training*
4. Cost Efficiency Ratio (CER): *CER = TER / TCR*.

Defining acceptability criteria based on such a utility property means defining compliance rules in terms of acceptable margins, for example: $0.8 < TER < 1$.

Within the training domain three major evidence solutions exist to collect evidence to prove satisfaction of these acceptability criteria for training simulation utility. These solutions are:

1. Simulator to Real Task: here an experimental group is trained on the simulator for a certain time (*GM-VV Test Method,* see Figure 11). Afterwards they receive additional training on the operational equipment to reach a specific level of competence. A control group is trained on the real task (only) to reach the same level of competence (*GM-VV Oracle*). Afterwards the performance of both groups is compared and is used to determine the simulator utility (*GM-VV Results Evaluation*). In practice there are many practical constraints that make it hard to make a fair comparison between both training groups. In those cases it must be decided whether the evaluation results are good enough to use as evidence (*GM-VV Decision Procedure*).
2. Within Simulators or Quasi Transfer: an experimental group is trained on a simulator with one or more simulator variables (colour, sound, mechanical motion, etc) degraded or omitted (*GM-VV Test Method*). A control group is trained for the same tasks on the fully operational simulator (*GM-VV Oracle*). The difference between the groups in performance reveals the relative contribution of the manipulated variable(s) on the utility properties of training simulation (*GM-VV Results Evaluation*). This solution does not provide absolute measures and is based on (unproven) assumption that the fully operational simulator is 'valid' with respect to the real world task. Hence a decision must be made on whether the evaluation results are good enough to use as evidence (*GM-VV Decision Procedure*).
3. Expert and Subject Opinion: experts (operators, instructors and training specialists) and trainees are asked to give their opinion on the training simulation utility. Experts and subjects collectively form in this case the *GM-VV Test Method*, *Oracle* and *Results Evaluator*. An obvious drawback is that most people have limited expertise with regard to the cues that facilitate learning; hence their opinions may be biased to overestimate the value of those cues that are (technologically) appealing. A decision must be made on whether the evaluation results are good enough to use as evidence (*GM-VV Decision Procedure*).

GM-VV Optimization for Training Simulations

Examples of tailoring GM-VV by optimization (seeking the optimal VV&A effort cost-benefit value) can also be found in the NATO test case. Relevant utility properties, adequate compliance rules and evidence solutions are identified by means of the GM-VV goal network. Developing an appropriate GM-VV goal network that balances the real world risks, costs, resources, etc. (Figure 6) is a run-time form tailoring by optimization of the VV&A project.

GM-VV is based on non-refutation. In practice it is impossible to prove for complex systems, which simulation-based training systems are, that they provide the solution to the real world need. No amount of VV&A, and thus resources, can help that. Therefore optimal choices need to be made with regard to limited available resources to make sure that the best possible VV&A activities are performed.

GM-VV Optimization focuses on the choices to be made during the construction of the GM-VV goal network and on the choices between alternatives of instantiations of training simulation specific implementations for the generic evidence solution.

During the construction of the goal network the top goal is disaggregated in a number of steps to the level of acceptability criteria (AC). A goal is split into sub-goals via a strategy that specifies why proving the sub-goals means proving the upper goal. Not all sub-goals, however, contribute evenly to the upper goal in terms of risk. Sub-goals for which it can be shown that their contribution to risk is negligible small may be left undeveloped. In the driving simulator case-study, at some point in the argumentation network, showing that the simulated car is good enough is split into two parts: parts of the simulator have been used for many similar studies and one part was build specifically for the current study. The VV&A problem owner indicated it had much experience with the already existing part and indicated that they posed no risk to the current study. Only the new sub-goal was further developed.

When the goal network has been developed up to the AC, an evidence solution for each AC must be specified. For a given AC often several different tests can be specified that have different levels of resource consumption and have different certainty of their outcomes.

A multi criterion optimization must be made that takes into account the contribution of the ACs to the overall risk, and for each AC the costs in terms of resources and the convincing power of all test alternatives.

In the driving simulator case-study one of the ACs stated that it must be shown that the update rate of the velocity and steering behaviour data must at least be 25 Hz. One possible test was to look at the time-stamps of the data points in the log-files and determine that the required update rate was indeed always reached. This is a relatively cheap and quick test. However, it leaves the possibility that the simulator internally had a lower update rate and produced the higher frequency data points by e.g. extrapolation. A more convincing test is to examine the source code and determine whether the produced data points really are individual data points; a much more time and budget consuming test that can only be performed by specialized personnel.

CONCLUSIONS

This paper presented the technical substance of the Generic Methodology for Verification and Validation (GM-VV) to support acceptance of models, simulations and data. GM-VV is submitted for standardization within the Simulation Interoperability Standards Organization (SISO) and targets for balloting in 2011. GM-VV provides a more general applicable and unifying approach to VV&A for M&S than currently existing VV&A standards and practices. GM-VV provides the referential basis for development of future M&S problem and application specific VV&A methods, tools and techniques. If build on the same basic components they will be interoperable with each.

Within the NATO tasks group MSG-073 GM-VV is currently applied in various case studies to improve the methodology, provide application and problem domain specific recommended practices for GM-VV (methods, tools and techniques), develop training material and VV&A CASE tool specifications.
Based on the first results from this NATO effort and the Dutch MoD technology demonstration program on GM-VV, the paper showed how GM-VV can be tailored for application in the training simulation domain. Future work will focus on developing a much more detailed recommended practice for GM-VV for the VV&A of training simulations, using new results from the current running case-studies and several new ones.

REFERENCES

SISO GM-VV PDG (2010a), Generic Methodology for Verification and Validation (GM-VV) to support Acceptance of Models, Simulations, and Data, Recommended Practice Guide, SISO-GUIDE-00X.2-201x-DRAFT-v1.5.1

SISO GM-VV PDG (2010b), Generic Methodology for Verification and Validation (GM-VV) to support Acceptance of Models, Simulations, and Data, Reference Manual, SISO-GUIDE-00X.3-201x-DRAFT-v1.3

SISO GM-VV PDG (2010c) Generic Methodology for Verification and Validation (GM-VV) to support Acceptance of Models, Simulations, and Data, Handbook, SISO-GUIDE-00X.2-2010 DRAFT-v1.5.1

REVVA (2009), REVVA Reference Manual, version 1.0.5., www.revva.eu

IEEE (2007), IEEE 1516.4, Recommended Practice for Verification, Validation, and Accreditation of a Federation — An Overlay to the High Level Architecture Federation Development and Execution Process

Roza et. al (2009), Specification of Needs Statements for Training Simulation Development, Verification and Validation, 09F-SIW-078, SISO SIW.

Emmerik van et. al (2010), GM-VV Illustrated: An Educational Example from the Human Driving Behavior Research Domain, 10S-SIW-035, SISO SIW.

Verstegen et. al. (1999). *The Use of Simulators for Training and Instruction*, (Report No. TM-99-B003). Soesterberg, The Netherlands: TNO Human Factors.

Gagné et. al. (1992). Principles of Instructional Design (4th ed.). Forth West(etc.): Harcourt Brace Jovanovich College.

Farmer et.al. (1999), Handbook of Simulator Based Training, Ashgate Publishing Limited, Vermont USA.