The generic methodology for verification and validation applied to medium range anti-tank simulation training devices

Jeroen Voogd
Q-tility
The Hague, The Netherlands
jeroen.voogd@q-tility.nl

Manfred Roza
Q-tility
Amsterdam, The Netherlands
manfred.roza@q-tility.nl

Keywords:
Verification, Validation, V&V, GM-VV, Simulation Training Devices

ABSTRACT: *The Dutch Ministry of Defense (NL-MoD) has recently acquired an update of its medium range anti tank (MRAT) missile system, called the GILL. The update to the SPIKE Long Range (LR) weapon system is accompanied with the acquisition of new simulation training devices (STDs). These devices are bought Commercial off the Shelf (COTS). The question arises whether the STDs are valid training means for the NL MoD intended training purposes.*

In this paper we present the application of the Generic Methodology for Verification and Validation (GM-VV) to the question above. First the intended purpose of the STD's is determined by executing a training needs analysis, then the verification and validation (V&V) areas of interest are selected based on how the training curriculum depends on the usage of an STD and the uncertainty about its quality. During the V&V study it was found that specific tests would only be possible at a later time, e.g. due to unavailable validation reference data, outside of the time frame of the V&V study, and feasible substitute tests had to be defined. Many findings from the V&V tests indicate the usefulness of the STDs, while others indicate that changes are required, either to the training curriculum or to the STDs.

The GM-VV allows for adapting to the scope as well as the graceful degradation of the V&V tests. The NL MoD can build upon the current findings at a later time, e.g. by adding reference data, to further decrease uncertainty, and thus reducing the M&S use risk.

1 Introduction

The NL-MoD has recently acquired an update of their GILL anti-tank missile system it has used for the past ten years. This upgrade, the SPIKE weapon system, is a more powerful system, but its operation requires some difficult psycho-motoric skills that need to be mastered before a gunner can be send into a combat situation. The operational cost of the SPIKE system prohibits live training, leaving simulation as the only alternative. Therefore, to train their SPIKE team members the NL-MoD envisioned a training environment and program which deploys four different types training devices: the mechanical trainer (MT), the InDoor Trainer (IDT), the OutDoor Trainer (ODT), and the Spike Team Trainer (STT). The last three are simulation-based training devices.

The NL-MoD has acquired these STDs commercial off the shelf (COTS) from the SPIKE weapon system supplier, RAFAEL [1]. These COTS STDs have been developed based on common and general applicable user requirements. For the NL-MoD this results in a significant M&S use risk that the STDs are not be able to completely meet the NL-MoD training needs in terms of learning goals, doctrine, typical mission area's and other training requirements. To address these M&S use risks the NL-MoD ordered a V&V study to determine the validity of the acquired STDS for their specific intended training purposes. Based on the findings of this V&V study the NL-MoD wanted to make well informed decisions regarding the STD acceptability, STD change proposals and modification of the SPIKE training program(s) to maximize the training effectiveness and efficiency.

The V&V study towards the SPIKE STDs has been conducted by Q-tility, the Dutch expertise center for V&V of M&S systems. Q-tility is established in 2012 and powered by the Dutch National Aerospace Laboratory NLR and the Netherlands Organization for Applied Scientific Research TNO. Q-tility was selected by the NL-MoD for their knowledge and expertise of the Generic Methodology for Verification and Validation (GM-VV) [2] [3] [4]. The GM-VV was developed under the auspices of the Simulation Interoperability Standards Organization (SISO) and NATO in an international consortium supported, among others, by the NL-MoD.

This paper discusses how the GM-VV has been tailored and applied by Q-tility to suit the V&V needs of the NL-MoD SPIKE STDs. In Section 2 first an overview of the GM-VV is given, which is necessary to understand the tailoring and application of the GM-VV for the V&V of the SPIKE STDs. Next an overview of SPIKE weapon system and simulation training devices the GM-VV is given (Section 2). The paper continues with a description of the V&V study activities and outcomes (Section 3 and 6). The paper ends in Section 7 with lessons-learned and recommendation regarding the application of the GM-VV and V&V in general.

2 GM-VV Overview

The optimal VV&A method depends on the individual needs and constraints of an organization or application domain. However, common principles and best practices are clearly recognizable, and this was the key driver behind the development of the Generic Methodology for Verification and Validation (GM-VV) by the SISO VV&A community [2] [3] [4]. The GM-VV offers a general baseline and guidance for VV&A of M&S that is applicable and tailorable to the individual VV&A needs of a wide variety of M&S technologies and application domains. The GM-VV consists of three interrelated parts, the conceptual, implementation, and tailoring framework (Figure 1).

The GM-VV conceptual framework provides essential VV&A terminology, semantics, concepts, and principles. The framework facilitates communication, understanding, and implementation of VV&A across and between different M&S contexts. In contrast to many views on V&V—namely, that it starts with the M&S requirements and ends with the developed M&S asset—this framework is premised on the idea that models and simulations are always developed and employed to fulfil the specific needs of their stakeholders (e.g., trainers, decision makers).

The GM-VV implementation framework consists of the interrelated products, processes, and organization. The product dimension contains information-based VV&A products that can have multiple instances, and representational and documentation formats. These VV&A products are produced by the processes, activities, and tasks defined by the process dimension. They can be executed recursively, concurrently, and iteratively. The roles defined in the organizational dimension are involved in the execution in one or more of the VV&A processes, activities, and tasks.

The GM-VV tailoring framework provides ways to tailor the implementation framework for each individual M&S organization, project, or application domain. The tailoring allows for modification of the building blocks in the GM-VV product, process, and organizational dimensions to satisfy the specific VV&A requirements and constraints in the M&S environment in which the GM-VV is applied. During the execution of the V&V work, risk-based tailoring is used to find the optimum cost-benefit ratio (e.g., distributing project resources based on M&S use risk).

3 SPIKE Weapon System and Simulation Training Devices Background

In 2001 the NL-MoD chose to equip its antitank units with the GILL weapon system (WS). The GILL is a

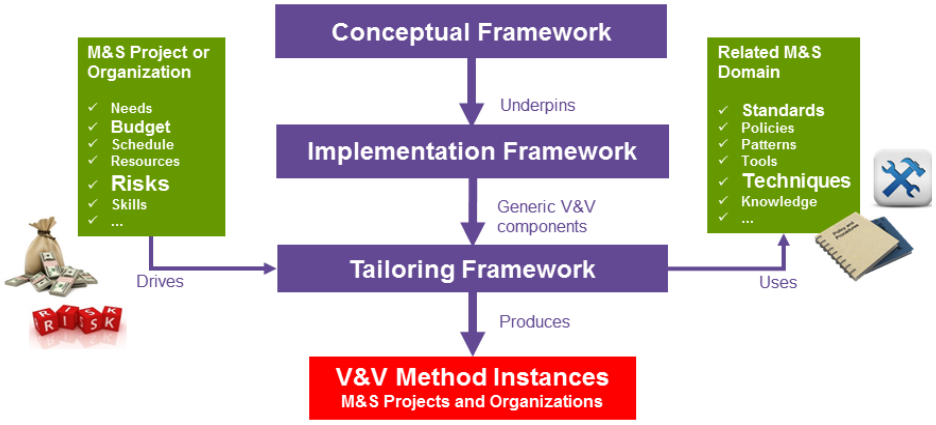

Figure 1 GM-VV Technical Framework Overview

medium range (effective range is 2500m), fire and forget (F&F) version of Rafael's SPIKE missile system [1]. F&F means that after making a lock on a line of sight (LOS) target and firing the missile, the missile will automatically guide itself to the target. However, over the years the type of operations of the antitank units has changed. Nowadays, these units are deployed more broadly than originally meant: they are no longer only used as a medium range antitank weapon in the open field, but now also for extended ranges and non-line of sight (NLOS) targets. In these situations it may become necessary for the gunners to manually guide or steer during missile flight. Therefore this year the GILL version will be upgraded to the full SPIKE LR version.

3.1 SPIKE Weapon System Description

The SPIKE version is almost identical to the GILL. The key differences are the increased effective range to 4000m of the SPIKE LR missile and most important the addition of manual missile guidance modes that allows for NLOS operations (Figure 2). Besides the F&F the gunner now has two extra (manual) operation modes: Fire and Update (F&U), and the Fire and Steer (F&S) mode. In these modes the gunner can observe the target and its environment by means of either a Daylight Optic or IR images send through a fiber-optic wire from the missile sensors to the gunner. The gunner can give lock updates or steering commands via that same wire.

Figure 2 SPIKE Weapon System and Missile

These new technical capabilities also cause difference in the NL-MoD antitank unit's operational deployment and doctrines between the GILL and SPIKE which are:

- *Changing target, increase hit probability and/or avoid collateral damage*, the F&S and F&U provide the gunner with the option to change the position of impact during missile flight, either to a target of higher value, adjust the missile to react on counter measures or to avoid collateral damage.
- *Fire without lock on a target*, the SPIKE allows a missile to be launched without being locked onto a target. This is especially useful when there is no line-of-sight (LOS) with the target or target is at large distance and no lock can be made, but information of other parties indicates a target should be present.
- *Intelligence information*, The SPIKE missile sensors provide a good opportunity to observe the terrain during flight, allowing the option to use these images for gathering intelligence about the enemy.
- *Battle damage assessment (BDA)*, the possibility to follow the missile up until to the moment of impact, the gunner knows whether the target was hit.

3.2 SPIKE STDs Description

The technical and operational changes have a significant influence on the training program, means and costs. The new SPIKE capabilities are not covered by the current GILL trainers and thus no longer suitable for training the new required gunner's skills. At the same time live training with real missiles is too costly. This meant that the new NL-MoD SPIKE training program and environment had to rely more on simulation means. These conditions made the procurement of new simulators necessary. To support such simulation-based training the next three STDs are COTS available from RAFAEL:

- *InDoor Trainer (IDT):* The IDT is a mock-up with a high physical fidelity, simulating the weapon system's sensors and missile flight dynamics (Figure 3). It allows SPIKE gunners to train their procedural, psychomotor, cognitive and divided attention skills. It is typically used in an indoor training facility. The instructor console allowed editing scenarios and observing the trainees interaction with the mock-up.
- *OutDoor Trainer (ODT):* The ODT is for the most part the actual weapon system. The command and launch part is the same. But the canister with the missile does only contain those elements relevant for all procedures up to launch and the real missile sensors. Some additional differences are present in the power supply and cooling mechanism of the missile IR sensor.
- *Spike Team Trainer (STT):* The STT provides tactical training for a Spike combat team by integrating 2 times 2 IDTs within a single collaborative training environment. It can be configured to enable the training of single gunners, 2 sections or an integrated platoon. STT facilitates training of the unit's full operational sequence along with advanced debrief facilities.

Besides these STDs also *mechanical trainer (MT)* is available from RAFAEL. This is a physical dummy of the SPIKE WS without any simulation capabilities that is used for practicing (dis)assembly of the WS and mounting on the backpacks. To facilitate the training of their anti-tank units the NL-MoD acquired both the SPIKE IDT and STT from RAFAEL.

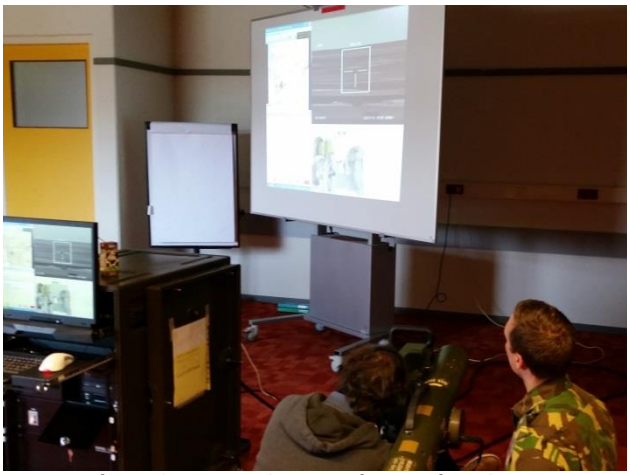

Figure 3 NL-MoD SPIKE Indoor Trainer Setup

4 V&V Project Initiation, Organization and Management

The V&V study of the SPIKE STDs has been performed using Q-tility's V&V life-cycle model. This V&V life-cycle model is a tailored compliant instance of the GM-VV that incorporates both the project and technical level components as specified by the GM-VV. The GM-VV project level has been instantiated to be able to perform V&V studies as a separate and independent project from an M&S development project. Q-tility's V&V life-cycle has successfully been applied in two other V&V studies [5] [6] and consists of three distinct phases (Figure 4): V&V project initiation (this section), V&V project execution (Section 5) and V&V project closure (Section 6). The blue processes reflect and integrate the GM-VV project level process components. The green and red processes reflect and integrate the GM-VV technical level process components. It should be noted that the Q-tility enterprise organization and memory are a tailored instances of the enterprise level components of the GM-VV, and incorporate the V&V personnel, best-practices, tools and techniques.

4.1 V&V Agreement Settlement

The V&V initiation phase started with settling a V&V agreement with the NL-MoD SPIKE STD stakeholders. This activity consisted of various interviews with these stakeholders, visits to NL-MoD anti-tank units training facilities to gain insight into their current operations and duties, as well as the current simulation-based training program and environments. Based on this information the V&V project needs, objectives and constraints were identified. To guide the elicitation and to structure this information the GM-VV V&V agreement document template was used [3]. This V&V agreement settlement was conducted by two V&V experts of Q-tility and resulted in the following four V&V study objectives:

1. To have sufficient evidence on to what extent the NL-MoD anti-tank units learning objectives can be achieved with the SPIKE STDs,
2. To prevent negative training from introducing the STDs in the SPIKE training program,
3. To establish trust and acceptance of simulation-based training among operational weapon system users,
4. To have sufficient insight in the STDs quality to make informed decisions on possible Equipment Change Proposal (ECP) towards the Rafael.

From the discussions with the NL-MoD stakeholders it became clear that they did not have a well-defined formulation of the intended purpose (i.e. learning objectives) and use (i.e. deployment in the training program) of the new STDs. A clear formulation of the M&S intended purpose and intended use is mandatory pre-condition for any V&V study. Therefore, prior to start the actual V&V study, it was recommended by Q-tility to first perform a training needs analysis (TNA).

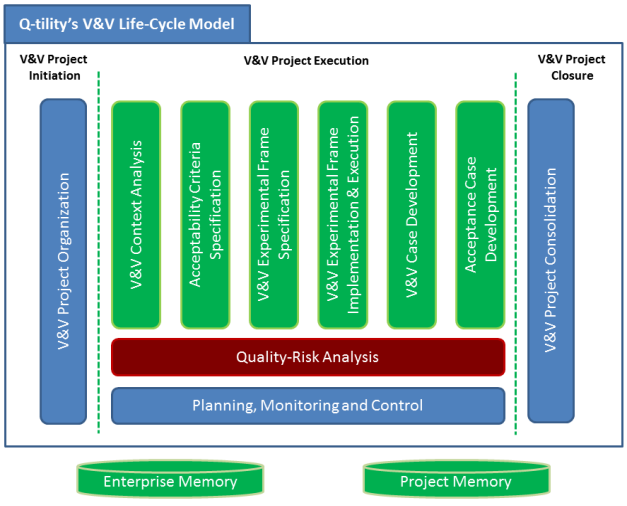

Figure 4 Q-tility V&V Life-Cycle Model

The TNA resulted in an overview of the new capabilities and resulting operational tasks and situations in which the SPIKE WS is expected to be used. A list of new competencies that both the SPIKE unit gunners and commanders should possess to effectively deploy the SPIKE WS in these situations was also missing. The competencies identified in the TNA are grouped into five clusters, containing in total 29 competencies. The five groups are: *Taking and leaving fire position* (e.g. Take and set-up the fire position), *Observing and engaging targets* (e.g. detecting targets), *Management and maintenance* (e.g. handle ammunition), *Analysis and planning* (e.g. analyze the terrain conditions), *Command* (e.g. maintain shared situational awareness). Based on these competencies the learning objectives were defined as well as two high level training programs. Moreover, the TNA resulted in a specification of possible targets (e.g. (un)armored vehicles), possible terrains (e.g. open field, urbanized areas, mountainous), possible environmental conditions (e.g. warm, cold, wet, dry), possible partners (e.g. other national and international units, support units), and required attitudes (e.g. independence, perseverance, proportionality). As the final step of the TNA all the acquired information regarding the M&S intended purpose and use was translated into general M&S requirements for the STDs also provided an important input for the V&V study.

4.2 V&V Project Organization

After the V&V agreement was settled a V&V project manager was assigned by Q-tility enterprise manager, who formed a V&V team of 10 experts. Several of these experts came from independent research institutes (e.g. TNO and NLR) and the NL-MoD itself. Collectively this team had all the skills and domain knowledge to properly accomplish the V&V project objectives.

To manage and document all information artifacts produced during the V&V project, a project memory was instantiated. For confidentiality reasons the project memory has been implemented as an internal and protected Sharepoint. This project memory file system was organized according the key V&V management and technical activities as presented in Figure 4, and contained instances of the GM-VV information templates as a basis to document the V&V information items [3][4]. Furthermore, the freeware yEd tool in combination with the Q-tility Goal-Claim Network Modeling Language has been used to implement the GM-VV argumentation structure [4].

4.3 V&V Project Planning, Monitoring and Control

From the V&V agreement a more detailed V&V plan was created at the beginning of the project with specific goals, tasks and deadlines, which have been allocated to one or more of the V&V team members. It was also decided to have regularly scheduled meetings with the NL-MoD stakeholders to discuss the status and progress of the project. In these planning and management activities an M&S use risk-based approach was used to collectively make decisions regarding the focus, rigor and allocation resources of the V&V study. Some of the most important decisions and underlying arguments were the following.

Two of the three SPIKE STDs are an upgrade or new version of GILL STDs already in use for the past ten years. Only the team trainer, STT, is a new STD for the NL MoD. Therefore, it was decided to focus on the STDs itself, and not on the overall setting in which also the complete training of the anti-tank units is validated.

Since there were three different STDs to be validated, it had then to be determined how to distribute the V&V effort and resources over the each STD. It was finally decided, in cooperation with the NL-MoD stake holders to primarily focus on the IDT (75% of the effort) and secondary on the STT (20% of the effort), and very limited to the ODT (5% of the effort) based on the following decisive arguments:

- The SPIKE IDT will be the main training system for learning to operate the SPIKE WS.
- The STT is basically a layered network of 2 times 2 IDTs, with some additional functionality for training commanders and support instructors.
- The STT would not become available to the V&V team during this V&V project lead time. Only the five year old Israeli Defence Forces (IDF) version of the STT, which is different than the purchased NL-MoD version, would be available to the V&V team.
- The ODT is basically a SPIKE WS without a real live missile with a warhead that could be fired. Except for the power and gas sources, all other hard- and software is identical to the real SPIKE WS.

Since Rafael has been producing the IDT for many years, the new IDTs are likely to be delivered without obvious flaws for normal training purposes. However, not much is known about the fidelity of models within the IDT, the scenario tool (i.e. VR-Forces)used in the IDT is new, and the application of the IDT to the specific NL-MoD operational tasks, competencies, training program and

tasks have not been studied before. Therefore it was decided to dedicate extra effort to the following aspects of the SPIKE IDT:

- Fidelity of the sensor models,
- Fidelity of the missile flight model,
- Utility of the instructor stations,
- Utility of VR-Forces as scenario tool,
- Use of existing and future terrain DB,
- M&S of exceptional situations (e.g. weather and counter measures) behavior of the SPIKE WS,
- The M&S of features not previously available on the GILL version of the SPIKE WS,
- Utility for training the NL-MoD specific doctrines and missions.

In order to be able to make claims about the validity of the STDs with a high degree of certainty, insight into the underlying models and/or V&V referent data of high quality is required. One of the major constraints that was foreseen was the difficulty of obtain information on to the STDs internals such as the missile flight model and sensor models and access to V&V referent data of actual flights (e.g. telemetry data). During the V&V project it proved that it was impossible or difficult to obtain such V&V referent information from the SPIKE WS and STDs supplier RAFAEL. Reason for this is that such information has been classified by RAFAEL as both commercial and military confidential and could not be disclosed to any customer. This had significant impact on V&V experimental frame design and implementation, and the results and their associate residual uncertainties.

5 V&V project execution

The V&V project has been performed according the seven technical processes (green and red) of the V&V life-cycle instance depicted in Figure 4. All technical processes were executed mostly concurrently and in several iterative cycles to incrementally develop the V&V report for the NL-MoD stakeholders. This is the result of both the applied quality-risk and explorative V&V strategy. In the next sections a summary description of the conducted activities inside these seven V&V technical processes and their outcomes will be given.

5.1 V&V context analysis

The objective of this activity to acquire and analyse all relevant information regarding both the real SPIKE WS, the STDs in combination with the training environment in which the STDs are intended to be used. For this purpose Q-tility applied a context information discovery and analysis technique that has been developed based on the GM-VV four-world model (Figure 5). The first step of this technique is that the V&V team identifies an initial set of possible information sources (people and media) in each of these four worlds. Starting point for this first step was the information collected during the TNA (Section 4.1). This information was combined with historical and readily available information on the SPIKE WS, STDs, their usage and V&V efforts, which were acquired from stakeholder interviews (e.g. hand written notes), documents provided and e-mail discussions. All context information was analysed and stored in the V&V project memory until sufficient coverage (measured by a convergence of information) and understanding of the V&V context was obtained to proceed to the next V&V technical activity

Both historical, literature and open source information on anti-tank missiles design, performance and sensors systems in general available from previous projects executed by the V&V team experts or NL-MoD stakeholders were used. Furthermore, specific information of the GILL weapon systems and some of its V&V results were discovered during this activity. Though some of the information was already 10 years old, it was considered useful since the SPIKE missile and weapon system is for a large part still similar to the GILL. As a matter of fact the NL-MoD GILL missiles are upgraded to the SPIKE missiles by minor electronics and software modifications.

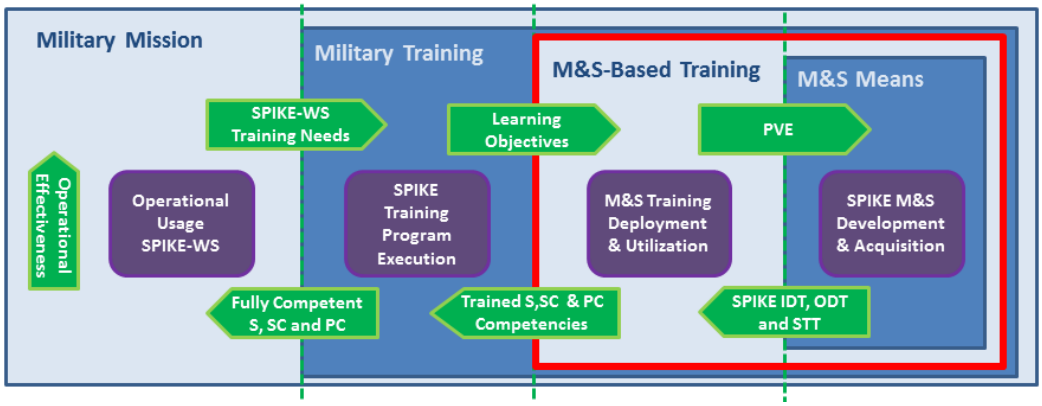

Figure 5 Q-tility Four World Context Information Discovery

On the other hand, due to confidentially considerations it proved to be very hard get actual information from RAFAEL on the SPIKE WS, STDs and past V&V efforts (4.3). Information provided by RAFAEL was very high level specifications, presentations and other advertisement based information. To gain more information from the supplier the NL-MoD managed to arrange a visit for two V&V team members to the RAFAEL premises. During that visit a bit more information was gained regarding SPIKE WS, STDs and the ways in which they perform V&V of their STDs:

- The Israeli Defense Force (IDF) has trained 1300 trainees with the STDs. RAFAEL regularly asks the trainees with operational experiences for their feedback on STDs simulation realism. These indicate that the realism is high (4.2-4.5 on a scale of 1 to 5). There was a score of 4 on the realism of simulated combat scenarios.
- When further investigated the trainees often see sometimes trivial differences between the real weapon system and the STD. RAFAEL considers these differences to have negligible impact on training outcomes.
- The STDs are regularly tested by experienced SPIKE WS gunners of the IDF.
- Visual comparisons of simulated and real missile trajectories were showed in comparable scenarios. The images indicated very similar trajectories.
- When asked if telemetry data was compared with a simulated trajectory RAFAEL indicated that this had indeed been done several years ago and the results showed that the trajectories deviated very little, but no data was actually produced during the visit.

Except for the information described above no other information regarding the SPIKE WS or the STDS has been received from RAFAEL.

5.2 Acceptance criteria specification

The objective of this activity is to specify appropriate acceptability criteria for the STDs. Acceptability criteria specify quality criteria that the STDs products should meet in order to be suitable for the defined intended training purpose and use (Section 4.1). GM-VV identifies three principle categories of acceptability criteria that can be assigned to M&S systems [2]:

- *Utility*, level of user functionality and performance
- *Fidelity*, level of simulation realism
- *Correctness*, level of M&S development quality

These quality criteria directly relate to the level of confidence (i.e. uncertainty) and associated residual M&S use risk that can be placed on STDs after the performing this V&V study. The definition of these acceptability and V&V quality criteria was an iterative effort of the V&V team in cooperation with domain SME and NL-MoD stakeholders. This activity leveraged the V&V context analysis results as its input (Section 5.1). These acceptability criteria were developed by the V&V team using Q-tility's specific implementation of the GM-VV goal-based argumentation network and the yEd tool. This developed network is referred as the Target of Acceptance (ToA) and was formally approved by the NL-MoD V&V project supervisor prior to move to the V&V experimental frame specification activity (Section 5.3).

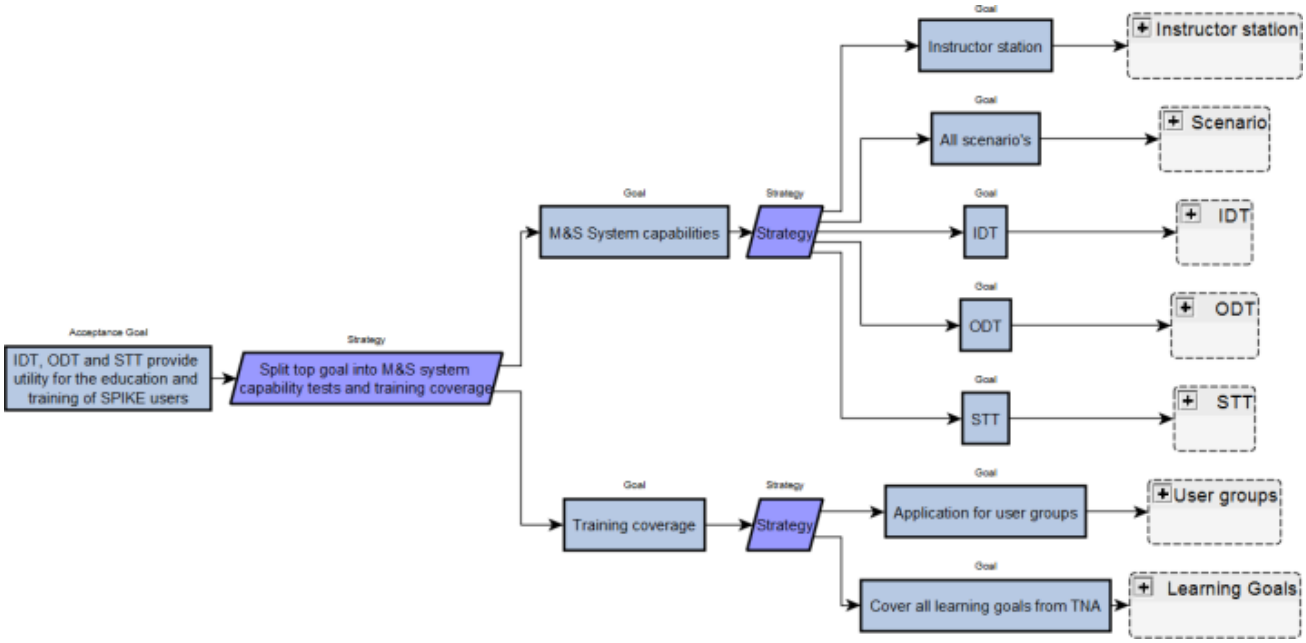

Figure 6 High Level View of the SPIKE Target of Acceptance

A high level view of the final ToA is depicted in Figure 6 and consists of two main branches. On the bottom the training coverage branch consists of STD-based training solution related criteria that were derived from the training objectives and competencies derived of the TNA. These were translated in a set of minimal training usage requirements on the STDs. The top branch specifies the M&S requirements regarding the (non)functional and fidelity capabilities of the STDs in order to support the NL-MoD training purpose and used (Section 4.3). Given the limited information given by RAFAEL on the STD internal models, simulation software/hardware structure and data, the V&V team used their own expertise in missile systems modelling and simulation to set-up a simulator component-based network with acceptability criteria. In here they also used a reusable Q-tility argumentation pattern for STDs, which from a high level considers for example the following aspects for the SPIKE IDT:

- *Simulation system (SS)*: This comprises the modelling and simulation part of the IDT, which mimics or simulates the real SPIKE WS in representational form, function and behavior (i.e. fidelity), and that of the real operational work environment (i.e. mission area) in which the Gunner / Commander deploys and perform his/her tasks with the SPIKE WS. Besides fidelity properties also the correctness properties that impact the simulation system quality are assessed.

- *Training and instruction support system* (TISS): This comprises all training and instruction support systems (e.g. instructor working station, scenario management tool, recording and after action review tools) that are part of the IDT and which the instructors can use to monitor, evaluate, control and guide the training of the Gunner / Commander on the IDT. Besides the utility properties (e.g. functionality and performance) also the correctness properties that impact the TISS utility are assessed.

- *Technical sustainment support system* (TSSS): This comprises all support systems of the IDT or associated with the IDT, that are used by technicians to technically sustain the IDT. In order to keep the IDT technically operational and up-to-date, and in-line with the actual SPIKE weapons system upgrades. Besides the utility properties (e.g. functionality and performance) also the correctness properties that impact the TSSS utility are assessed.

5.3 V&V experimental frame specification and implementation

These activities respectively specify and implement the V&V experimental frame that will produce adequate evidence to demonstrate with a sufficient level of confidence that the STDs and the STD-based training solution meet the acceptability criteria from the ToA (Section 5.2). The experimental frame consists of a set of evidence solutions that specify what evidence and how the evidence for each acceptability criteria should be obtained [2] [3]. Specifying evidence solutions involves the design of experiments and for each evidence solution the following information should be specified: a reference base with the expected results (i.e. V&V referent), specification of measurement or experiments with the STDs and real system, the required equipment and analysis techniques, and initial or boundary conditions, evidential strength [3]. The definition of the evidence solutions was also an iterative V&V team effort and leveraged the V&V context analysis results and ToA as its input (Sections 5.1 and 5.2). Again Q-tility's specific implementation of the GM-VV goal-based argumentation network and the yEd tool was used to develop the underlying argumentation. This developed network is referred as the ToVV.

From the ToVV several extensive STD sub-system specific V&V test guides were developed that served as the basis for the implementation and execution of all V&V evidence solutions. However, during the implementation of the experimental frame two major issues were encountered that strongly limited the scale and rigor in which strong objective and quantitative V&V evidence could be acquired.

The first issue was the limited information regarding the STDs but most important was the very limited availability of the STDs to do dynamic and explorative testing. This was caused by the fact that RAFAEL did not supply the STDs on time to the NL-MoD. As a result the SPIKE ODT and STT were not available to the V&V to perform V&V testing. The SPIKE IDT that could be tested, with a major delay of the V&V project, was a beta-version of the IDT that would be delivered to the NL-MoD. Moreover, the SPIKE IDT is a COTS product and therefore didn't provide the dedicated V&V test interfaces and tools. Though they were available, RAFAEL didn't make them available to the V&V team. As a result the controllability and observability of the SPIKE IDT that were needed to properly execute the foreseen V&V tests provided to be insufficient to acquire strong objective and quantitative V&V evidence.

The second issue encountered was the unavailability of V&V reference data for the SPIKE WS from RAFAEL (Section 4.3). Different other options were conceived and tried to acquire best possible V&V reference data. These comprised, in a decreasing evidential strength order, having RAFAEL perform the specified V&V tests, having NL-MoD performing live firing trials, acquire access to similar SPIKE STDs and WS in other countries and get SPIKE experienced personnel for subjective V&V. Despite many management and negotiation efforts this eventually none were successful. As a result the V&V tests and ToVV had to be gracefully scaled down in size and rigor. This meant that most of the tests had to rely on subject matter expert opinion, inspections and evaluations (i.e. subjective qualitative and quantitative V&V results), and that, where possible, some V&V tests had to be reworked into the most optimal and feasible objective quantitative alternative V&V tests.

For instance, for the STT it was determined by the V&V team that no additional information would come available, besides the information the V&V team acquired from their visit with RAFAEL to the IDF STT. Here they had observed a demonstration of training scenario running on the STT and given the opportunity to asked questions. Though, the IDF STT is from 2010 and differs from the one the NL-MoD would receive in combination with some information from RAFAEL on the differences, this was the best available V&V referent data and STT data to be used as evidence. Moreover, it leverages and integrates the findings of the IDT V&V tests. This means that the V&V of the specific STT aspects (e.g. not of the individual IDT) could only comprise static V&V methods or techniques, such as comparative desktop analyses and walkthroughs of the available information sets.
Another example was the lack of V&V reference and IDT data to objectively and quantitatively validate the simulation of the four SPIKE WS's sensors performance and dynamics. Instead two SME from TNO in weapon system sensors were employed by the V&V team to observe and qualitatively assess (e.g. face-validation) the simulated IDT sensor performance and dynamics. They based their assessment on commonly known basic technical and engineering characteristics in several typical sensor system test scenarios. As a V&V referent the GILL ODT was used. This was justified by the fact that the GILL WS sensors are not significantly different than the ones of SPIKE WS, and these differences were known as well as to what the effects would be on the sensor performance.

5.4 V&V test execution

Since the V&V of STT only comprised static V&V tests based on a fixed set of written information, the V&V could be executed separately from the IDT V&V tests by two V&V team members. Later these findings could be combined with the findings of the IDT V&V tests into a consistent acceptance case for the STT (Section 5.5).

The V&V tests for the IDT were grouped and executed according to the areas of interest defined in the ToVV. In total five full V&V testing days with the beta version of the SPIKE IDT were scheduled at the NL-MoD premises in the following V&V test sessions order:

1. V&V testing of the SPIKE IDT Weapon system control logic, functions and malfunctions simulation (*duration 2 days*). This session consisted of large set of functional testing, assertion checking and special (e.g. extreme or boundary conditions) input testing V&V techniques [4] [7]. The focus was primarily on checking the correctness of the SPIKE IDT mock-up hard and software, and secondary the level of fidelity of the underlying control logic and functionality models. The V&V referent consisted of all information on the GILL and SPIKE WS that was made available by RAFAEL and the NL-MoD. This referent information such as reference, user,

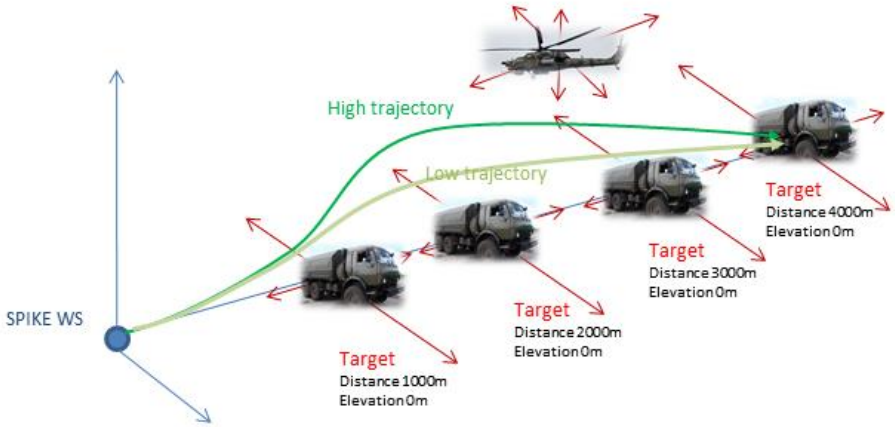

Figure 7 Missile flight model tracking tasks control response V&V test range setup for the IDT

operational and maintenance manual. These tests were conducted by two NL-MoD V&V team members experienced with the both the GILL and SPIKE. All tests were performed under supervision of a senior V&V engineer of Q-tility.

2. V&V testing of the SPIKE IDT Missile flight simulation model (*duration 1day*). This session consisted of extensive set of explorative dynamic model fidelity, model assertion checking and model special (e.g. extreme or boundary conditions) input testing V&V techniques [4] [7]. The focus was on explorative assessment of the level fidelity of the SPIKE IDT missile flight simulation model for manual steering (i.e. F&S) or lock updates (i.e. F&U) and possible anomalies in known model assertions (e.g. range or turn radius) and under extreme operational conditions (e.g. sharp turns and steep climbs, cross wind, fast moving (Figure 7) and non LOS targets). The V&V referent consisted of all available documented GILL and SPIKE missile performance information from RAFAEL and the NL-MoD, as well as missile performance characteristics and data from public sources and a parametrised first-order flight simulation model that was developed by the V&V team. These tests were conducted by two V&V team members with an aerospace engineering background and experienced with missile model development, supported by a NL-MoD expert GILL / SPIKE trainer.

3. V&V testing of the SPIKE IDT Sensor simulation models (*duration 1 day*). This session consisted of extensive set of explorative dynamic model fidelity, model assertion checking and model special (e.g. extreme or boundary conditions) input testing V&V techniques [4] [7]. The focus was on explorative assessment of the level fidelity of the SPIKE IDT sensor models (i.e. DL, CCD, TS and IIR) and possible anomalies in known model assertions (e.g. terrain light and thermal conditions) and under extreme operational conditions (e.g. tracking of targets with counter measures or altering LOS/NLOS visible targets). The V&V referent consisted of all available documented GILL and SPIKE missile sensor information from RAFAEL and the NL-MoD, as well as the GILL ODT (Figure 8) which had similar DL, CCD, TS and IIR sensors as the SPIKE WS. These tests were conducted by two V&V team members with sensor systems engineering background and experienced with sensor model development, supported by a NL-MoD expert GILL/SPIKE gunner.

4. V&V testing of the SPIKE IDT Training and support systems (*duration 1 day*). This session consisted of series V&V tests that comprised walkthroughs of the training programs in which the STDs are used, inspection of the IDT and its training support system (e.g. scenario tools and instructor working station), and reviews of the training competencies and objectives to assess to what extend they could be covered by the STDs and the IDT in particular. Input to this V&V test session were the other V&V test findings regarding the SPIKE IDT level of fidelity, which are necessary to assess the impact on training program and objectives. During this V&V test session also the SPIKE IDT and the GILL ODT (Figure 8) were available and used to validate these STDs in several representative training use-cases. These tests were conducted by three V&V team members, one being a simulation-based training expert and two V&V team members that participated in the previous three V&V sessions allowing them to provide their input on the IDT level of fidelity. Furthermore, this session was supported by a NL-MoD expert GILL/SPIKE trainer who is involved in the training design, implementation and execution for the anti-tank units.

Figure 8 GILL ODT V&V test set-up

The order of the V&V test sessions was explicitly chosen because the V&V findings of one V&V test runs were preconditions to start the other V&V test runs or served as a necessary input. For instance, in case the SPIKE WS control logic simulation and hardware did not work to a certain quality or fidelity level, it would be hard if not impossible to properly test the missile flight simulation model that uses the control logic as input. Likewise, V&V testing (test session 4) of the SPIKE IDT training utility

against the training objectives and intended use within the anti-tank units training programs, requires knowledge about the level of fidelity of flight and sensor simulation models.

5.5 V&V and acceptance case development

These activities respectively integrate and document the V&V findings of the executed V&V tests (Section 5.4) and develop the acceptance case for the each of the three verified and validated STDs. This acceptance case has been developed by the V&V team using Q-tility's specific implementation of the GM-VV goal-based argumentation network and the yEd tool. Besides the V&V findings also the ToA (Section 5.2) and ToVV (Section 5.3) were used as input for this activity to help build the acceptance case structure and assess to what extent the acceptability criteria have been met by each of the three STDs. This acceptance case has been classified by the NL-MoD stakeholders as confidential and therefore cannot be disclosed in this paper.

The acceptance case served as the basis for the V&V report that provided written acceptance recommendations according the following template developed by Q-tility:

- *M&S requirements satisfaction*: which comprises a list of all M&S requirements for each STD that were derived from the intended training purpose and uses, along with the argument for its inclusion and level of importance, and based on the V&V findings to what extend the M&S requirement is met. For the latter the following coding is used:
  - 👍 + green text = can be satisfied, no real issues
  - ☞ + black text = can possibly be satisfied, but with some explaining remarks,
  - 👎 + red text = cannot be satisfied, there are serious issues.
- *Additional V&V findings*: which comprises a list of V&V findings that were found during the V&V tests and may have an impact on the satisfaction of one or more M&S requirements or have unforeseen side effects on the M&S intended purpose and use. The impact is indicated with the same coding as for the M&S requirements.
- *Terms of STD usage*: for the application of the STD to the intended training purpose and use. The next four terms are defined:
  - Unacceptable: expresses for which parts of the intended purpose the STD is not acceptable.
  - Limitation: expresses for which parts of the intended purpose the STD can be used with certain restrictions.
  - Unserviceability: expresses functionality of the STD that are either temporarily inoperative or performing below the nominal level in order to be used for the intended training purpose
  - Reservation: expresses for which parts of the intended purpose it cannot clearly be proven that the STD is unacceptable or acceptable.
- *M&S use risks*: expresses any (residual) risk factors of using the STD for the intended training purpose. For each factor the likelihood and impact is expressed on a five point scale.
- *Recommendations for M&S use and improvement*: expresses recommendation regarding possible STD improvements and its use, in order to optimize its utility for the intended training purpose and use. These recommendations are prioritized using the four MoSCoW categories [8].

This V&V report has been classified by the NL-MoD stakeholders as confidential and therefore no examples from this report can be included in this paper.

6 V&V Project consolidation and ROI

According the GM-VV standard the V&V project consolidation phase consists of two major activities. First the delivery of the V&V report to the V&V client, i.e. NL-MoD stakeholders, and the official project sign off (Section 6.1). In this chapter also the NL-MoD usage and RoI of the V&V study is high-lighted (Section 6.2 and 6.3). The second activity comprises the archiving of the V&V project memory and capitalization of lessons-learned and other reusable V&V knowledge from the SPIKE V&V study to improve the Q-tility's V&V service provision (Chapter 7).

6.1 V&V report delivery

The first step after all technical activities is to compile a concept V&V report that summarizes on one side the technical V&V activities performed for justification and establishing confidence in the quality of V&V work itself. Second it provides as summary presentation of the V&V results, acceptance claims and recommendations from the acceptance case, along with the residual (use) risks that remains after the V&V study (Section 5.5). Prior to the delivery the V&V report was distributed to the NL-MoD stakeholder for a review with written feedback. At the moment of writing this paper the review was still in progress. Next a delivery meeting will have to be

scheduled at one of the NL-MoD premises with all relevant SPIKE STDs stakeholders that might read or use the V&V report and the inhered acceptance recommendation. During this meeting a brief introduction on V&V will be given, followed with a presentation how the V&V study was conducted by the V&V team. Next the resulting acceptance case plus the associated recommendations will be presented. After this presentation the audience will be given the opportunity to provide feedback on the acceptance recommendation and the V&V project execution. The meeting will be closed with a discussion and agreement on the terms of usage of the produced V&V project information artifacts and results. The stakeholder feedback and other results from discussion during this meeting will then be summarized and added to the V&V report as a separate section. Moreover, it is expected that some changes need to be made throughout the report to enhance the clarity and understandability of the V&V report. This will help assure that relevant stakeholder information regarding the V&V project or the SPIKE STDs are consolidated, facilitates the usage of the V&V report by the NL-MoD stakeholders, gives input for possible recurrent V&V of the SPIKE STDs in the (near) future (Section 6.2), and help formalize the agreements made on the terms of usage of the V&V project outcomes. After this the final V&V report will be sent to NL-MoD V&V project supervisor for final approval and sign-off of the V&V agreement.

6.2 NL-MoD V&V results utilization and RoI

The V&V report served as input for a gap-analysis that was performed by the NL-MoD with support of training experts and M&S engineers from TNO. The objective of this analysis was to optimize the SPIKE training program utility by developing solutions to close the gaps in accomplishing the training objectives that were identified by the V&V project. Solutions that have been considered ranged from: differently deploying the STDs (e.g. modify the training program design), with or without other training means, making ECPs for the STDs or a combination thereof. In this gap-analysis the findings and acceptance recommendations from the V&V report were considered by NL-MoD and the TNO team as very useful and indispensable to make well-informed decisions on the most cost-effective solution to fill these gaps.

In the end the resultant training program and application of the STD is expected to lead to a significant reduction of the number of required live-training shots with the real SPIKE WS. For some parts of the training these live training shots could be safely reduced to zero with little or no impact on the training outcomes. Given the fact that a real SPIKE missile costs about the costs of two luxury cars, this is a large cost saver for the NL-MoD. The V&V project costs, equivalent to about three times the cost of a SPIKE missile, are far more outweighed by the returns the for the NL-MoD.

The V&V report also served as a useful reference for the NL-MoD factory/site acceptance phase of the SPIKE STD. It helped the NL-MoD acquisition team to both structure and focus these acceptance tests and provided convincing argumentation for (future) ECPs towards RAFAEL. The V&V project also uncovered several important shortcomings and errors in the SPIKE IDTs; shortcomings and errors not previously detected by RAFAEL themselves. Fixing these issues proved to be a general improvement of RAFAELs COTS SPIKE STDs quality and capabilities that would benefit their other or new customers as well. Driven by these commercial benefits RAFAEL has already implemented the recommended solutions for these shortcomings and errors for free in the final version of the STDs delivered to the NL-MoD. Hence this SPIKE V&V project resulted in a win-win situation for both parties, and an additional V&V return on investment for the NL-MoD.

Besides the V&V report, the project also delivered a set of reusable V&V test guide document for the missile flight simulation model, the sensor models, the STDs mock-up controls, logic, functions, and malfunctions, and the training support systems. These V&V test guides are intended to be used by the NL-MoD to develop and execute additional live test runs with SPIKE WS missile and sensors next year. These live test runs aim to acquire new V&V referent data for more rigorous V&V of some parts of the STDs that will help in reducing the remaining residual uncertainties (i.e. M&S use risks). Moreover, these live tests will provide more insight in the actual SPIKE WS capabilities and help improve its operational deployment and doctrines.

The V&V test guides are also intended to be reused by the NL-MoD for recurrent V&V of the SPIKE IDT when new updates and upgrades come available from RAFAEL for either the weapon system itself or the IDTs. Furthermore, the V&V test guides may be used by the NL-MoD in the sustainment phase of the STDs as a quality assurance tool to periodically check whether the STDs still perform conform to their acceptability criteria. Due to the reuse of these V&V test guides, these recurrent V&V efforts can be conducted will less cost (i.e. time and money) than the initial V&V efforts performed as described in this paper.

6.3 Other V&V results spin-off

Apart from future utilization on the SPIKE IDT/STT STDs NLD-MoD plans to re-use the V&V data on two other simulation systems that incorporate SPIKE functionality. These simulation systems are the live training simulator from SAAB and the battalion level trainer TACTIS.

7 Conclusions and lessons learned

Overall this V&V study for NL-MoD proved again that GM-VV is a very versatile and general applicable V&V methodology; a methodology that was easily tailored to scope, resources and timeline of the V&V project. Due to the implementation of the GM-VV enterprise layer by Q-tility [2] [3], it was possible to benefit from a reusable V&V life-cycle model along with previously developed or used V&V argumentation design patterns, V&V methods, techniques and tools. It resulted in a more effective (i.e. better outcomes) and efficient (i.e. less cost) V&V project execution. The V&V test guides set-up and ToVV structure developed in this SPIKE V&V project proved not only be useful for the NL-MoD for recurrent V&V but also for two other V&V projects that are currently conducted by Q-tility.

This V&V project encountered the regular occurring issue within the M&S domain of not having well defined M&S intended purpose and use [9] [10]. Having a well-defined M&S intended purpose and use is of great importance, not only for performing effective V&V but also for specifying M&S systems (e.g. requirements) [11]. The application of structured TNA for the SPIKE V&V proved to be a very useful tool in defining the M&S intended purpose and usage profile for the STDs. Moreover, it seamlessly integrated which each other. Initiatives for developing standard practices and templates for M&S intended purpose and user specifications are currently proposed [9].

This SPIKE V&V project demonstrated that performing V&V on COTS M&S systems is strongly limited by the availability of V&V referent data as well as the willingness of the M&S supplier to cooperate with the V&V team and provide insight or access to the internals of the M&S system. This should not only be taken into account when planning V&V efforts for COTS M&S systems. In fact provisions for getting access to essential weapon system data from both the weapon system supplier as well as the M&S supplier should be a prerequisite in any procurement document, whether it being a COTS product or a user defined system. These provisions are necessary in the attempt to get best possible V&V referent data and evidence possible. However, the SPIKE V&V project made clear that despite the initial V&V tests had to be gracefully scaled down in scope and rigor for practical constraints and considerations (Section 5.3), the remaining downscaled V&V tests provided more findings than were anticipated. Moreover, they proved to be of unexpected value to both the V&V team and the NL-MoD stakeholders that had to make well-informed decisions for the SPIKE STDs (Section 6.2).

The lesson learned from this V&V project is that when next to nothing is known about the M&S system or its referent (i.e. an information/data poor environment), a V&V study can still deliver very valuable evidence about the M&S system which was not known before. This is in line with the general conception that the first observations or measurements usually provide the highest return on investment in reducing uncertainty [12].

GM-VV recommends the application of a rigorous (e.g. objective and quantitative) use risk and uncertainty approach as drivers for selecting and setting acceptance criteria, V&V methods and techniques. However, for the SPIKE V&V project this proved to be difficult due to the lack of pragmatic methods, techniques and tools to support this assessment in a cost-effective manner. The practical approach that was adopted instead was to discuss risk and uncertainty during meetings with the various domain SME and NL-MoD stakeholders. These discussions are mandatory to determine the focal points that should be used during derivation of the acceptability criteria. Though this proved to work for the context and constraints of this V&V project, it won't completely work for every V&V project such as large scale or high risk-critical M&S systems [12]. Currently, the NATO task group MSG-139 is developing and evaluating practical, but more rigorous, M&S user risk methods to address this issue [13].

8 References

[1] SPIKE-LR. http://www.rafael.co.il/Marketing/343-997-en/Marketing.aspx, accessed June 2015.

[2] SISO, *GM-VV Vol. 1: Introduction and Overview*, SISO-GUIDE-001.1-2012, October, 2012.

[3] SISO, *GM-VV Vol. 2: Implementation Guide*, SISO-GUIDE-001.2-2013, August, 2013.

[4] SISO, *GM-VV Vol. 3: Reference Manual*, SISO-REF-039-2013, October 2013.

[5] Lemmers et al, *V&V study of an F-16 familiarization training simulator*, 13F-SIW-016, *Simulation*

Interoperability Workshop, Orlando, FL, 16-20 September 2013.
[6] Voogd J, *Verification and Validation of the Public Order Management Serious Game*. NATO NSMG LS-123, Januari 2014.
[7] US-MSCO, VV&A RPG http://www.msco.mil/VVA_RPG.html
[8] http://en.wikipedia.org/wiki/MoSCoW_method, accessed May 2015
[9] Piersall C H., Grange F.E.. The Necessity of Intended Use Specification for Successful Modeling and Simulation of a System-of-Systems CrossTalk. October 2014.
[10] Zeigler P., Nutaro J., *Towards a framework for more robust validation and verification of simulation models for systems of systems*. SCS JDMS, February 26, 2015.
[11] Jamshidi, M., 2008. *Systems of Systems – Innovations for the 21st Century*, Wiley, June 6-8, 2012, Bordeaux, France.
[12] Hubbard D., *How to measure everything, 3rd edition*, Wiley, March 17, 2014.
[13] Roza. M., Towards an M&S use risk management methodology for simulation based training, NMSG 126 Conference, Washington D.C., October 2014.

Acknowledgement

First of all the authors wish to thank the V&V project supervisors, Ad van Lier of the NL-MoD Joint IV Command – Section Simulation (JIVC/KIXS) and Patrick den Engelsman of the NL-MoD Defence Material Office (DMO/MATLOG) for their support from the NL-MoD side for this project, by e.g. making documents available and making the visit to RAFAEL possible. Furthermore, many thanks to Adj. van Mier of the NL-MoD anti-tank weapons training school (IGPaB) for his cooperation and providing access to their facilities for executing the V&V tests. Last, but not least, we are very grateful to Sgt1 van Dalen, sr. weapon instructor at the IGPaB, for his very energetic support in executing the V&V tests. His knowledge of the SPIKE weapon system and the STDs was indispensable for this V&V project.

Author Biographies

JEROEN VOOGD is a member of the scientific staff at the Modeling, Simulation and Gaming Department of TNO Defense, Security and Safety. He holds a Ph.D. in Computational Physics from the University of Amsterdam in the field of modeling and simulation biophysical systems on parallel distributed computation platforms. His current research interests include decision support with simulation, validity and agent technology. Currently he is the TNO technical principle for V&V within the Dutch V&V expertise center Q-tility.

MANFRED ROZA is a senior R&D Engineer at the Training, Simulation and Operator Performance department of the NLR. He holds a M.Sc. and Ph.D. in Aerospace M&S Technology from the Delft University of Technology. He specializes in requirements, design specification quality assurance, verification and validation of models and simulation, in particular for simulation-based training systems. Currently he is the NLR technical principle for V&V within the Dutch V&V expertise center Q-tility. Besides this he is also a civil (EASA) and military aviation authority (NL-MAA) accredited technical inspector for flight simulation training device qualification and certification.