Mugla Journal of Science and Technology

FACE VALIDATION IN LARGE-SCALE SIMULATION PROJECTS

Savaş ÖZTÜRK*, Department of Digital Game Design, Muğla Sıtkı Koçman University, Türkiye, savasozturk@mu.edu.tr
(https://orcid.org/0000-0002-6259-6132)

Received: 25.09.2022, Accepted: 27.04.2023 Review
*Corresponding author DOI: 10.22531/muglajsci.1179857

Abstract

Face validation is an informal type of validation, usually applied in the early stages of a simulation project or model development process. Face validation is used to determine the reasonableness of a model or simulation when advanced validation methods such as statistical analysis or sensitivity analysis are not applicable. In large-scale simulation projects such as warfare simulation applications, which consist of a large number and variety of models that cannot be evaluated independently and contain high levels of uncertainty, face validation should be utilized, but no standard method for face validation can be proposed. In this paper, we review the literature and practical applications of face validation. We propose a process to improve the effectiveness of face validation. The proposed process is expected to help to realize a measurable, concrete and practical face validation, especially in large-scale projects involving high-fidelity complex models and constructive simulations.

Keywords: Verification, Validation, Face Validation, Simulation Model

BÜYÜK ÖLÇEKLİ BENZETİM PROJELERİNDE GÖRÜNÜŞ GEÇERLEME

Özet

Görünüş geçerleme, bir benzetim projesinin ya da model geliştirme sürecinin genellikle erken safhalarında uygulanan, resmi olmayan bir geçerleme türüdür. İstatistiksel analiz veya hassasiyet analizi gibi ileri seviye geçerleme yöntemlerinin uygulanamadığı durumlarda bir modelin ya da benzetimin makul olduğunun tespitinde görünüş geçerleme uygulanır. Çok sayıda ve çeşitte, bağımsız olarak değerlendirilmesi mümkün olmayan modellerden oluşan ve yüksek seviyede belirsizlikler içeren harp benzetim uygulamaları gibi büyük ölçekli benzetim projelerinde, görünüş geçerlemeden faydalanılması gerekmekte ancak görünüş geçerleme için standart bir yöntem önerilememektedir. Bu çalışmada görünüş geçerlemenin uygulandığı pratik çalışmalar incelenmiş ve görünüş geçerlemenin etkinliğinin artırılmasına yönelik süreç önerisi sunulmuştur. Önerilen sürecin, özellikle yüksek sadakatli karmaşık yapıda modeller içeren büyük ölçekli projelerde ve yapay(constructive) türde benzetimlerde, ölçülebilir, somut ve pratik bir görünüş geçerleme gerçekleştirilmesine yardımcı olacağı değerlendirilmektedir.

Anahtar Kelimeler: Doğrulama, Geçerleme, Görünüş Geçerleme, Benzetim Modeli

Cite

Öztürk, S., (2023). "Face Validation in Large-Scale Simulation Projects", Mugla Journal of Science and Technology, 9(1), 24-33.

1. Introduction

Simulation is the application of modeling and analysis methods using technology to make decisions and solve problems [1]. Simulation models are the abstraction of real-world physical entities and systems according to their intended use. A rigorous Verification and Validation (V&V) process must be applied to ensure that simulation models are valid and reasonable. Verification seeks to answer the question "Are we building the model right?" and validation seeks to answer the question "Are we building the right model?" [2].

Verification is usually related to the follow-up of the model development process and the rate at which customer requirements are met. Document and code reviews, unit tests, and functional tests, demonstrating the minimum level of operability of functions are some of the verification activities [3, 4]. Validation activities may include rigorous and intensive testing and analysis to prove that the developed model meets the need and that the real-life application of the model is sufficiently abstracted. While whether a laparoscopic surgery simulator to be used in the training of trainee surgeons is capable of suturing a surgeon is a function that even a non-surgeon user can understand through validation methods, the decision that this simulator is sufficient and reasonable to be used in surgical suturing training can only be made by surgeons comparing the simulator with their real-life suturing experience.

Validation can be carried out with a large number and variety of methods depending on the type of the model, the nature of the developing team, the data available, and the time allocated for development. Balcı conducted a detailed taxonomy of verification, validation, and testing techniques and listed 77 methods [5]. Among the methods listed in four main groups, the first one is informal methods and face validation is included in this group. Face validation is the examination of a simulation model or system by experts who have experience with that model or similar models to determine whether the model is reasonable enough. Face validation is one of the most preferred validation methods when models are interdependent, data is limited and there is little information about the system. In face validation, the project team, potential users of the model, and people with knowledge of the system under study make predictions and inferences, subjectively compare the model and system behavior under certain input conditions, and make judgments about whether the model and its results are acceptable. The example of a surgical simulator is a good example of face validation; whether a simulator with three-dimensional visuals and haptic elements can provide a realistic experience of an operation such as surgery or endoscopy can be decided not only by mathematical models, software tests, or statistical data, but also by the experience of users, i.e. doctors who have years of experience in surgical operations. However, some systems are not as single-purpose and not as well abstracted as the training simulation given as surgical simulator example. On the other hand, they may not be used by a human (expert) and face validation may not be easy. Face validation of distributed, complex and interacting simulation systems with many models and no human in the loop involves many uncertainties and there is no standard methodology proposed for this purpose.

This paper investigates the use of face validation in the simulation literature and proposes a face validation process that is particularly applicable to large-scale simulation projects. In the second part of this paper, necessity of face validation is provided in terms of lack of data and lack of expertise. The third part presents a review of the literature on face validation. The fourth part presents practical examples from industrial applications. In the fifth section, the proposed face validation process model is described based on examples of use in academia and industry. The article concludes with the results of the review study and information about future work.

2. Necessity of Face Validation

The extent to which the simulator or simulation, which is expected to replace its real-life counterpart, meets the expectations and resembles the real one can be evaluated by applying both user experience and some formal validation methods. In order to apply formal methods and to express the similarity rate numerically, a large amount and variety of data produced with the physical original of the model is needed. Obtaining all the data on which validation is based is in most cases impossible or limited. Face validation, which is one of the informal methods, is to ask some questions about the validity of the system to faces who have knowledge and/or experience with the system and to analyze their answers with statistical methods. Face validation is used in highly efficient areas such as surgical operation simulators, as well as in the validation of autonomous systems that contain a high degree of uncertainty. The situations that require face validation are as follows:

Lack of data: Comparing data from the real system with simulated data is one of the most preferred validation methods [6]. If there is not enough real-world or field data from the real system, it is not known exactly what the system will be compared to. For example, there are many factors affecting the performance of a dynamic motion model of a surface ship: Sea state, precipitation, bathymetry (bottom nature), and salinity affect the dynamic model indirectly. A large number of experiments are required to reveal the effect of these factors on the model. The data collected on the ship will be valid in the simulation model under the same conditions. The data obtained with the scenario of maneuvering at maximum speed in Saros Bay with a wind force of 7 according to the wind Beaufort scale will not be suitable for comparing the maneuvering at maximum speed in the simulation environment with the modeled environment in Marmaris with a wind force of 1 according to the Beaufort scale. In addition, it is necessary to obtain the parameter matrix from the manufacturer for the motion model [7, 8].

In this case, the preferred method is to synthetically reproduce the limited data available for different situations or to use it with certain assumptions. For example, in recent years, it is claimed that hundreds of millions of miles of test drives are required to collect the necessary data for the development, training, and validation of artificial neural networks in autonomous vehicle studies. Since this is not possible in the short term, the use of virtual environments, simulating vehicles and roads in a computer environment, has become widely preferred for validation [9].

Lack of Expertise: In the absence of sufficient data, it is necessary to ask one or more experts. Face validation is the act of consulting people and institutions that are trusted and whose word is valid in the field. However, in order to increase objectivity, in addition to those who know, experts and experienced people, there are also cases where judgments are made by consulting less knowledgeable, not yet specialized faces. When the bibliography on face validation is examined, it is seen that the studies in the field of medicine stand out and yield quite successful results. An example of face validation in the medical field is the investigation of the closeness of the feeling obtained as a result of the experience of materials, devices, and software, which are generally developed for educational purposes, to a mixed group in terms of expertise level. For example, training doctors to use a surgical robot using synthetic organs produced with a 3D printer instead of a real patient would allow for more practice and eliminate potential operational risks (Figure 1a). For this purpose, the experiences of a group of doctors who specialized in robotic surgery and a group of doctors who have not yet gained experience in the relevant experimental set were examined to investigate whether synthetic organs could be tried instead of real patients. The surgeons in this study were asked questions similar to the examples in Table 1 to determine whether the system was valid or not [10].

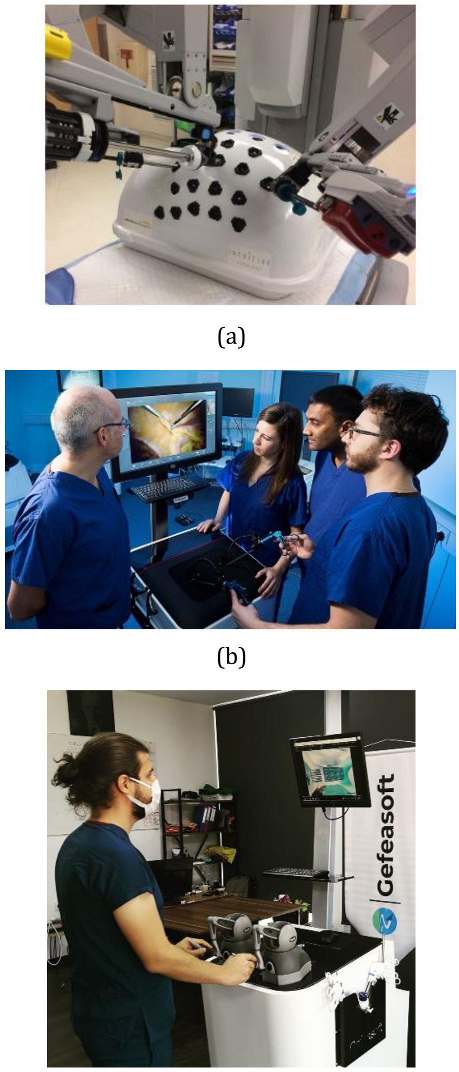
In another study similar to the one shown in Figure 1b, organs were transferred to the computer environment with three-dimensional modeling, and the difference between laparoscopic operation equipment and the surgery performed by looking at the computer screen was investigated. Surgeons' opinions on the realism of the system were obtained through survey questions prepared in accordance with a five-point Likert scale [11]. A similar example of the use of face validation in the field of medicine is implemented in the "COSimO - Surgical Oncological Simulator and Gamification Tool Development Project" (see Figure 1c) supported by TÜBİTAK TEYDEB within the scope of 1512 - Techno-Entrepreneurship Capital Support Program [12]. It is obvious that these three studies provide an important benefit in surgeon training: a) Since it is not tested on a real patient, any harm is eliminated. The error can be compensated and can be practiced as much as desired. b) The number of operating rooms and equipment is limited, more surgeons were trained in a short time with a simple setup. c) The quality of training can be increased by modeling situations that are rarely encountered in real life.

Table 1. Questions Asked to Surgeons Using Surgical Simulators [10]

Question No	Survey Question
1	Rate the aesthetics of this model between 1 and 5.
2	Rate the overall feel of this model between 1 and 5.
3	Rate the realism of this model between 1 and 5.
4	Rate the usability of this model between 1 and 5.

Royal [13] claims that face validation is not a uniform method that can be used for the validation of surgical simulators and robotic systems and argues that validation should be done in accordance with the "Standards for educational & psychological tests" published by the American Psychological Association(APA) [14].

Complexity and interdependency of models: In wargaming or platform (air, land, naval and joint) training simulators, a large number of simulation models are expected to be interoperable at different levels of fidelity and resolution [15, 16]. In such cases, separating a model from the whole system and validating it independently may not produce meaningful results. In these cases, "faces" with expertise in the system need to consider and evaluate the system from a holistic perspective.



(a)

(b)

(c)

Figure 1. Face validation practices in surgical simulations and simulators.

For real-time and virtual simulations, face validation can be applied and can be efficient due to the presence of human factors in the loop. However, it is not clear how to utilize face validation for constructive simulations where there is no human in the loop. One of the areas where face validation is most needed is in distributed and complex simulations where a large number of environments, platforms, sensors, and weapon models are expected to work together. In these simulations, both continuous and discrete simulation models are expected to be interoperable at different fidelity levels. However, validation of individual models does not guarantee the performance of the system as a whole. The people who will make the best decision in the acceptance of such systems are experts who have experienced this in real life, in other words, faces. Because all models contain imperfections, the model is not expected to be perfect, and it is up to the experts to decide "good enough".

3. Literature Survey

In the modeling and simulation literature, there seems to be no contradictory approach to the term face validation. However, examining the validity of questionnaires and tests is a separate issue and the terms face validation and face validity are often confused. Investigating the measurement success of a test or questionnaire is face validity. Content validity investigates the appropriateness of the test for learning/operational purposes [17, 18]. In short, most studies that use the term face validity have nothing to do with face validation, which is a test evaluation technique.

One of the fields where the term face validity in most commonly used is social research, and in this field, it is often confused with the term content validity [19-21]. The use of questionnaires has an important place in the evaluation of abstract and uncountable qualities. Although it is reported that it cannot be a concrete criterion in terms of validity because it is a daily, temporary, weak method, it is also frequently used in developing countries [22].

The first studies mentioning verification and validation date back to the 1940s. It is understood that the issue of model validation in this field, which falls within the scope of operations research, started to gain importance at that time [23].

Face validation is known to be based on participant reports and verbal feedback. Although face validation cannot explicitly test the simulation model, it is an invisible factor in the acceptance of the model. Face validation is difficult to do justice to, but it contributes significantly to the performance of training simulators [24].

In the simulation literature, face validation is considered as a validation method that can be applied throughout the project lifecycle but is not recommended to be used alone. The reason for this is that it cannot be quantified and involves subjective inference.

Sargent [25] proposed a flow that enables the development of simulation models through an iterative process. Accordingly, simulation models are validated in two stages. In the first stage, the conceptual model is validated, and in the second stage, the model that is turned into a computer program is validated (operational validity). It is suggested that conceptual model validation should be ensured by at least face validation [25, 26]. Sargent pointed out that face validation may also be used to evaluate graphical outputs of the simulation at operational validation stage [26].

Balcı has shown face validation as a method that can be applied in all process steps together with document review and inspection in both the US Department of Defense (DoD) and his own proposed verification and validation lifecycle models [2].

According to Law [27], it is not necessary to have a fully developed, working system to perform face validation. Analysts and domain experts review the simulation outputs (numerical results, animations, etc.) for suitability. The important point in face validation is to ask the experts before validation what kind of output the model should give against which input, and therefore to determine validation scenarios suitable for the purpose. If the simulation results are consistent with the expectations of the experts, the model is face validated.

In the MITRE (*Massachusetts Institute of Technology Research & Engineering*) Guide to Systems Engineering, face validation is mentioned as one of the five validation methods and, unlike the other four, its independence from data is emphasized [28]. Therefore, it is stated that it can be used when data-dependent model validation is not possible or practical. Moreover, even if data-dependent methods can be used, face validation is recommended at the first stage. If there is already a similar and validated model, a comparison of these will give a preliminary idea. Following this preliminary validation, more comprehensive approaches such as predictive validation can be taken.

In the MITRE Guide, face validation is defined as asking people with knowledge of the system whether a model or its behavior makes sense (e.g., is the logic in the conceptual model correct, and do the model's input-output relationships make sense?)

In another study, the Software Engineering Institute (SEI) proposed a 6-level validation quality maturity model inspired by CMMI (Capability Maturity Model Integration), which ensures the quality of the software development process [29]. According to this model, at Level 0, no validation is performed, no validation criteria are defined, no real-world model to compare against is defined, and only some validation is performed to assist development.

Level 1 is based entirely on face validation and is the most principled of the validation types. The domain expert sets the validation criteria and the reference model, and the domain expert compares the simulation results with the reference model data. The acceptability of face validation depends on the faces (domain experts) doing it. Many systems with complex scenarios, such as war games, have only been validated by face validation. At Level 2, the domain expert is gradually phased out. The domain expert is used to formulate user requirements.

Wang and Lehmann [30] proposed a maturity model similar to the Harmon and Youngblood validity maturity model, but more centered on independent verification and validation. According to this model represented in Figure 2, validation activities can be graded on a 5-point scale from 0 to 4 depending on whether they are performed by an independent team/organization, whether the model is developed from scratch or similar to a previously validated model, and the objectivity of the validation methods. Although face validation is not mentioned in this study, since it is a subjective method, it is classified as Level-1, regardless of the level of independence and maturity (intensity). In other words, since expert opinion is utilized, face validation is completely separated from dependent validation and is not considered as the lowest level of validation, but it cannot be considered as a higher level of validation. In Figure 2, it can be seen that face validation fits the subjective characteristic of salience and therefore corresponds to Level-1, which is indicated by the red color.

Klügl considers face validation for a social simulation domain with a large number of agents capable of autonomous decision-making and proposes a multifaceted validation method that incorporates human evaluation [31]. In this study, a clear and concise validation process is described. As can be seen in Figure 3, face validation is an important validation step before sensitivity analysis, calibration, and statistical analysis. If a model passes face validation, it is considered plausible and subjected to sensitivity analysis. If the model is not deemed valid enough by experts, implementation and verification are repeated and the executable model is updated.

In Louloudi and Klügl [32], the authors propose three techniques for face validation. First, they say that watching animations will give a bird's eye view of the accuracy of the model. The second is to examine the generated simulation data. In the third and newly proposed technique, the real user is involved in the simulation as an agent. However, the constraint here is whether the simulation interface is suitable for this. The authors state that even if data is available and statistical validation can be performed, face validation will contribute.

Pace [33] lists qualitative validation methods and face validation among the three main topics that are administratively confronted in validation activities. He emphasizes that qualitative validation has a bad reputation in the simulation world due to the distrust of domain experts' inferences based on unstructured, haphazard, and incomplete information. According to Pace, if applied well, qualitative validation can be as successful in simulations as it is in the medical field. According to him, the first major problem in face validation is that experts do not clearly report the rationale and basis of their judgments. Although there are efforts to provide expert opinions in formal ways and to improve the quality of the qualitative evaluation, it is also stated that there is no technique that can be recommended for general use.

Bharathy and Silverman [34] also critically review the literature on validation and propose their own taxonomy. Accordingly, validation is examined under three main headings. Internal, methodological, and external. Qualitative, causal, and narrative validation is a subtype of external validation. Face validation is also considered under this heading. It is mentioned that face validation can be done in two ways: In a very descriptive way with some statistics or in a way that is obtained with the help of checklists and has few definitions. However, it is possible to increase the contribution of the field expert for holistic and scientific validation.

Holistic evaluation can be achieved if the outputs of the simulation can be summarized in a way that the domain expert can understand and evaluate. In order to make the tests rigorous and unaffected by biases, experts sometimes make the model run outside the planned flow.

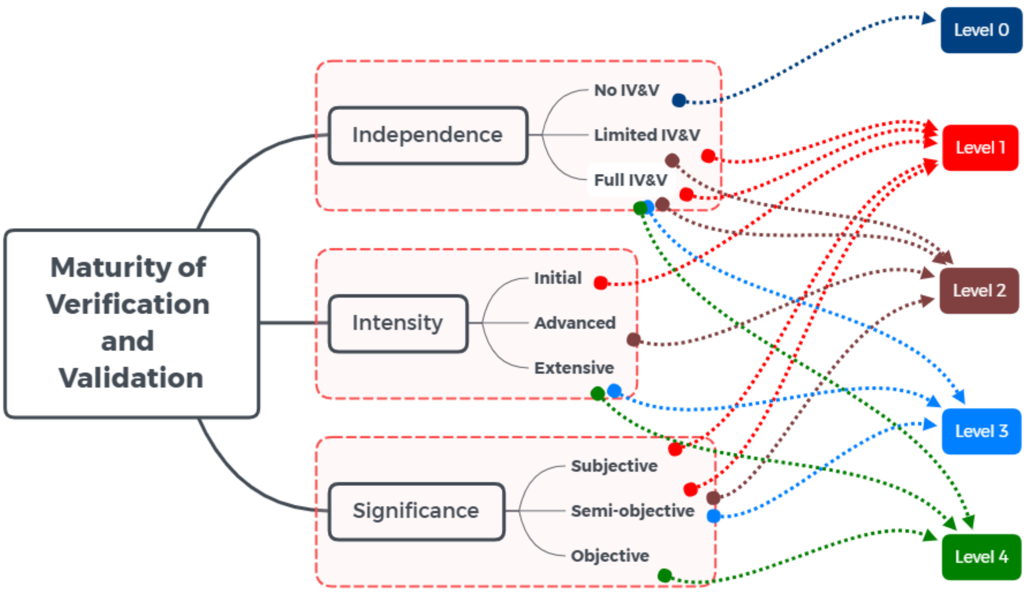


Figure 2. Wang and Lehmann Verification and Validation Maturity Taxonomy (IV&V stands for Independent Verification and Validation) [30].

.

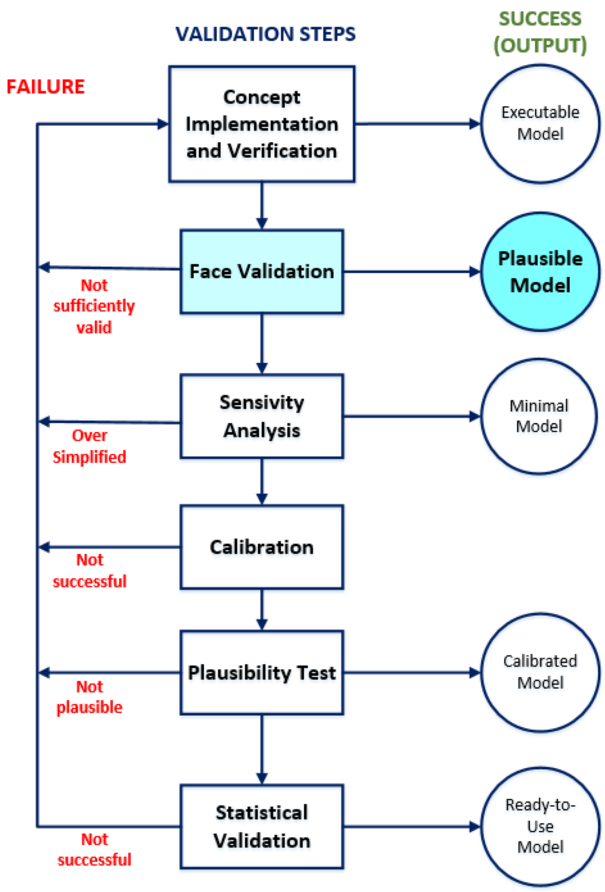


Figure 3. Validation Steps Defined For Agent-Based Simulations [31].

This paper also proposes an adapted Turing test for face validation purposes. A model is valid if the evaluator is unable to distinguish the differences between the results and the real system when shown the simulation results. One of the most important activities in qualitative validation is to increase the participation of domain experts through information-gathering meetings attended by all stakeholders.

4. Face Validation Practices in Large-Scale Projects

While there are definitions, stages, and examples of face validation in the literature, there is not enough information about the method. This is because there is no generally accepted, standardized method for applying face validation. For this reason, this section will first present several studies that have applied face validation independently of face validation.

MITRE, one of the world's leading research organizations, has drawn practical examples from lessons learned on validation [35]:

Quantifiable, objective model validation criteria should be established: If the model depends on stochastic processes, statistical tests are usually conducted comparing the behavior of the model with the real system. If hypothesis testing is performed, tests should be planned to anticipate Type 1 (false positive) and Type 2 (false negative) risks. As an alternative to hypothesis testing, a certain confidence interval can be determined and analyzed. In both cases, validation will be an activity that requires time, resources, and experience.

Trade-off curves should be established for the models: Increasing the level of fidelity and acquiring the data needed can be costly, so a trade-off analysis should be done early on for each model.

Not every model has to be validated: Some models are included in the system only to add certain functionalities, for those models validation is sufficient.

Simplification of models should be allowed: Some models can be simplified without affecting the optimal operation of the system. For example, if the goal is to generate numbers, sampling can be done using a limited number of parameters.

If possible, more than one validation approach should be used: Face validation is usually applied when data is limited. Face validation is recommended even when data-driven validation is possible.

Partial validation can be used when necessary: When it is not possible to validate a model as a whole due to a lack of data or other reasons, selected parts of the model can be validated, provided that they are reported instead of shelving the model completely.

A model working group could be established: For each model or group of models, a group could be established with participants from the sponsoring/client government organization and the model developer. The function of this group is to produce and review the assets needed for validation. In the case of a data-driven validation, a major effort of the group will be on data acquisition. The existence of such a group will improve the quality of the models.

Invest in analytical skills and resources: From the very beginning of the verification and validation process, it should be planned to invest in the manpower, tools, and process assets required for data analysis. It would be appropriate to appoint at least two experts: One is a domain expert who is familiar with the overall system, and the other is an expert who is competent in data analytics, data extraction, statistical analysis, and graphical capability for validation activities.

As a result of their literature review, Hatip and Durak [30] found that the most commonly used method for validating models is face validation [36, 37]. In this study, which gives importance to expert opinion, it is suggested that experts should perform the following checks.

- Whether there are inaccuracies/deficiencies in the description of the modeled system,

- Whether there are inconsistencies in model requirements,

- Whether the assumptions made are appropriate and accurate for the purpose of the model/simulation,

- Whether all assumptions used in the modeling are defined,

- Whether there are unclear or incorrect points in the model algorithms and formulations,

- Whether there is a mismatch in the model,

In the same study, the authors also performed validation with two practices. The first one is to implement the same conceptual model by two different tools or two different developers and compare the consistency of the results, and the second one is to produce scenario-based tests and compare the output results determined by the expert with the validation results.

In one of the most concrete practices on face validation presented in military simulation projects, the validation process is completed in 5 steps [38].

1. Determining the models to be validated
2. Establish a detailed validation plan
3. Operation of the detailed validation plan
4. Reporting and analyzing operational validation
5. Validation, defect resolution, and re-validation of model components

The most emphasized issue in this study is the plan formulation phase. The V&V team develops such a detailed and nuanced plan that it both facilitates the work of the experts who will conduct the validation and prevents them from making subjective decisions. For this purpose, details such as which platforms will be equipped with which sensors, which sensor will detect which target, and which target will be hit with which weapon are presented to the evaluators in checklists. The evaluators decide whether the results are realistic and feasible and whether the resulting product is reasonable. One of the most important contributions of this study is the presentation of lessons learned:

- Before starting validation, the design and code should be reviewed, if necessary, to see what the simulation does.

- Validation should be done where and under what conditions (operating system, amount of memory, etc.) the simulation will be run.

- Sufficient time should be allocated for validation. If sensitive information will be used, the time required to add and remove the information should be calculated.

- It should be ensured that the simulation is compatible with the real-world limits of the modeled entity (e.g. an automobile cannot travel at 500 km/h).

- The recorded data should be reviewed daily.

5. A Process Recommendation for Face Validation

Verification and validation, as a concrete block, is an important phase of simulation project management. A review of the literature and practice shows that there is a lot of information about the definition of face validation and its position and importance among validation methods, but no practical and standardized method for its application in large-scale and distributed systems has been proposed. Almost all of the face validation application examples in the literature have been successful in virtual simulations, that is, when the operation is performed by humans and the environment is virtual. However, in cases where statistical experts, data, resources, time, and budget are not sufficient, face validation, which is one of the subjective methods, is considered to make an important contribution to decision makers. On the other hand, it is emphasized in the literature that face validation is a technique that is not respected and avoided due to its subjective and informal nature. War games and military simulations are considered to be in the class of discrete event simulations, which are used extensively in decision making, but for which there are not many practical examples for validation [39]. Project management practices have been applied to the development of discrete event simulations and visual validation has been applied in addition to statistical validation [40]. The need arises to propose a process for this type of project that is easy to implement, reduces subjectivity by including a large number of faces, is supported by concrete and measurable criteria checklists, and is compatible with formal software and simulation development processes (CMMI, ISO 25010, etc.), is well documented, and can be inspected when necessary.

Defense Modeling and Simulation Coordination Office (DMSCO) conducts defense modelling and simulation projects in the USA. DMSCO operates the collaboration model for the purpose of registering the models obtained in projects for reuse [5]. In this coordination model, V&V Agent manages all auditing and enforcement activities of the models and ensures that the models are approved by the institutional accreditation body. This model will work very effectively in the presence of a sufficient number of qualified and independent verification and validation authorities. One of the most important contributions of this study is the preparation of a roadmap for auditing verification and validation to be carried out by independent and accredited institutions, even though it cannot be done in person, and also applying it on the face validation model, which is one of the most used validation methods in large-scale projects.

In order to minimize the problems arising from lack of validation, neglect or misapplication of validation in simulation projects, the foundations of an independent verification and validation practice were laid for the projects organized jointly by TÜBİTAK Defense Technologies Research Group (SAVTAG), Undersecretariat for Defense Industries (SSM) and Turkish Naval Forces Command, and an independent Verification and Validation Organization was assigned to each project for audit purposes. This model has been successfully applied and maintained for the first time in a project called Warfare Effectiveness Analysis Model (GEMED) [41]. Within this business model, face validation proved to be the most widely used validation method. The steps of the face validation process proposed in this study for implementation in projects are as follows:

1. Models to be face validated are identified.

2. Preliminary checks are made that the models are ready for validation: Have tests been completed, improvements made, and successfully integrated into the system?

3. If the readiness rate of the model is above a certain level (e.g. 90%), planning for validation without waiting for other models will ease the schedule.

4. A validation team is formed for each model: It is recommended that the team includes at least one domain expert, one client organization staff, and staff from contractor organizations. The presence of personnel from the test teams of different contractors will also increase the number of "faces", which will positively affect the reliability of the results. Developers of similar models should be involved in each other's validation studies. For example, the team developing the surface motion model should be included in the teams validating the underwater motion model, the torpedo model, and even the motion model of fixed-wing and rotary-wing platforms.

5. Experts should provide their input on the tolerance ranges of the model to be validated, if any, as far in advance as possible and propose test scenarios. This contribution increases the objectivity and consistency of the validation.

For example, "it is acceptable for two fixed-wing aircraft flying parallel to each other at a distance of 80 meters or less to be seen as a single detection on the radar."

Or; "it is acceptable for the sonar calculated bearing to differ by +- 3 degrees from ground truth".

6. In addition to the validation scenarios, a questionnaire form is prepared in which team members who are not experts in the field can also give their opinions. This form can consist of a list of questions in a 5-point Likert scale format that can be used to assess the participants' sentiments. The questions in Table 2 are suggested as an example, the set of questions should be determined according to the model. The criteria list in Table 2 is inspired by the work of Balci [42] where requirements, design, architecture, implementation and evaluation of the results of simulation models are associated with a number of quality factors. This relationship is actually very similar to the McCall model [43], which allows software quality to be expressed numerically and where subjective judgment is transformed into objective inference. McCall model has evolved over the years into an ISO (International Organization for Standardization) standard [44]. As a result, the list of criteria is based on the assessment experience gained on software quality, modeling and simulation systems.

7. The team convenes on the specified day and time and performs the tests within the specified scenarios. Scenarios missed before the test can also be tried.

8. The form containing the opinions of each team member is attached to the minutes of the validation activity.

9. At the meeting to be held at the end of the validation activity, no definitive conclusion is drawn as to whether the model is valid or not, but the participants can make a joint decision that the model should be validated again if necessary.

10. Documentation produced during validation activities is included in the Verification and Validation Report.

11. Experiences gained during validation activities are documented as lessons learned and published for use in other studies.

12. The independent validation authority, if any, is informed at each stage of validation activities and validation planning, reports and results for each model are shared.

This proposed process is managed, and monitored by the team responsible for verification and validation of the project or the client organization. The project's verification and validation documents (such as plan, design, monitoring report and results) will be the project-adapted version of this process.

Table 2. Sample Evaluation Items of Experts Participating in Face Validation

No	Evaluation Item
1	Evaluate the ease of use of the model (if possible, if it has a module reflected in the interface).
2	Evaluate the performance of the model (is there a significant difference in system performance when the model is not running and when it is?)
3	Evaluate the consistency of the model (if the same scenario is run twice in a row, do you get the same results?)
4	Evaluate the accuracy of the values produced by the model (Does the model produce the expected values and exhibit the expected behavior?)
5	Evaluate the integrity and completeness of the model (are there any missing or incomplete parts?).
6	Evaluate the quality of the model documentation (tests, user manuals, readability and understandability of the conceptual model).
7	Evaluate the maintainability of the model (How descriptive, readable and properly designed is the model's code and documentation? Can a new model developer easily adapt the code?)
8	Evaluate the portability and reusability of the model (Can the model be easily added to another simulation?)

6. Conclusion

In this study, the face validation technique, which is a subjective method that is widely used in the validation of simulation models, is explained and how this technique can be applied in simulation projects is discussed. Face validation is not considered a reliable validation method in some sources due to its subjective nature. However, it seems to give successful results in visual and tactile training simulations where surgical expertise cannot be tested with a machine or in complex systems consisting of a large number of models such as a war game. By taking advantage of practical applications used in industry, a process is proposed that will enable an effective face validation process to be operated in large-scale simulation projects. With this proposed method, it is aimed to introduce the face validation technique, which is avoided and disregarded as subjective, unofficial, and useless in defense and technology projects with high external dependency, data and expert shortages and difficulties in applying formal validation methods, and to obtain the highest benefit. Future work will be to examine the proposed process in the context of test automation in order to gain a formal and valid identity and to investigate the use of intelligent methods in the selection of experts, determining the questions to be asked to the experts, and evaluating the answers.

7. References

[1] Greenwood AG, "A Specification for Effective Simulation Project Management", *2020 Winter Simulation Conference (WSC)*, pp. 2482-2492, doi: 10.1109/WSC48552.2020.9383982, 2020.

[2] Balci O, "Verification, validation, and accreditation'', In Proceedings of the 30th conference on Winter simulation (WSC '98), IEEE Computer Society Press, Washington, DC, USA, 41–4, 1998.

[3] Kaya MC, Karamanlioglu A, Çetintaş Ç, Çilden E, Canberi H, Oğuztüzün H, "A configurable gateway for DDS-HLA interoperability", *Simulation Series,* vol. 2019-July, 2019.

[4] Hocaoğlu MF, Kolsuz H, Reşber E, "Simulation Conceptual Model And Scenario Verification Based On Adsif", *USMOS 2015- National Defense Applications Modeling Simulation Conference*, Ankara, 2015.

[5] Balci O, "Verification, Validation, and Testing of Models", In: Gass S.I., Fu M.C. (eds) Encyclopedia of Operations Research and Management Science. Springer, Boston, MA.

[6] Öztürk V, Arar ÖF, Rende FŞ, Öztemel E, Sezer S, "Validation of railway vehicle dynamic models in training simulators", *Veh. Syst. Dyn.*, vol. 55, no. 1, pp. 41–71,

[7] Haklidir M, Guven AF, Eroglu O, Aldogan D, Tasdelen I, "High fidelity modeling and simulation of submarine in a commercial computer generated forces toolkit", *In Proceedings of the 2009 Summer Computer Simulation Conference (SCSC '09). Society for Modeling & Simulation International*, Vista, CA, 299–306, 2009.

[8] Senyurek L, Koksal S, Genc HM, Aldogan D, Haklidir M, "Implementation Of Fuzzy Control For Surface Platforms In A Computer Generated Forces Toolkit", *International Conference on Computational Intelligence for Modelling, Control and Automation (CIMCA08)*, Vienna, Austria, 2008.

[9] Kalra N, Paddock SM, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?", *Transp. Res. Part A Policy Pract.* vol. 94, pp. 182–193, doi: 10.1016/j.tra.2016.09.010, 2016.

[10] Bendre HH, Rajender A, Barbosa PV, Wason SEL, "Robotic dismembered pyeloplasty surgical simulation using a 3D-printed silicone-based model: development, face validation and crowdsourced learning outcomes assessment", *J. Robot. Surg.*, vol. 14, no. 6, pp. 897–902, doi: 10.1007/s11701-020-01072-9, 2020.

[11] Ayodeji ID, Schijven M, Jakimowicz J, Greve JW, "Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum", *Surg. Endosc. Other Interv. Tech.*, vol. 21, no. 9, pp. 1641–1649, doi: 10.1007/s00464-007-9219-7, 2007.

[12] Gefeasoft, https://gefeasoft.com/?page_id=142, *COSimO- Surgical Oncological Simulator and Gamification Tool*, (Last Access: 29.03.2023).

[13] Royal K, "'Face validity is not a legitimate type of validity evidence!", *Am. J. Surg.*, vol. 212, no. 5, pp. 1026–1027, doi: 10.1016/j.amjsurg.2016.02.018, 2016.

[14] American Psychological Association (APA), *Standards for educational & psychological tests*, 1974.

[15] Kılıç E, Öztürk S, "'Büyük Ölçekli Projelerde Entegrasyon Testleri"'. *Yazılım Kalitesi Ve Yazılım Geliştirme Araçları Sempozyumu (YKGS)*, İstanbul, 2010.

[16] Öztürk S, Kılıç E, Erkan Ö, Dikenelli O, "'Dağıtık Benzetim Sistemleri İçin Entegrasyon Ve Test Yaklaşımları"'. *Yazılım Kalitesi Ve Yazılım Geliştirme Araçları Sempozyumu (YKGS)*, İstanbul, 2008.

[17] Parrott AC, "Performance tests in human psychopharmacology: II. Content validity, criterion validity, and face validity", *Human Psychopharmacology: Clinical and Experimental*, 6(2), 91–98., doi:10.1002/hup.470060203, 1991.

[18] Leijte, E., Arts, E., Witteman, B., Jakimowicz, J., De Blaauw, I., & Botden, S. (2019). "Construct, content and face validity of the eoSim laparoscopic simulator on advanced suturing tasks", *Surgical Endoscopy*, *33*(11), 3635-3643.

[19] Bharathy GK, Silverman B, (2003) "Holistically evaluating agent-based social systems models: A case study", *SIMULATION,* 89(1):102-135.

[20] Louloudi A, Klugl F, "Immersive face validation: A new validation technique for agent-based simulation", *Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2012*, pp. 1255–1260, 2012.

[21] Bolarinwa O, “Principles and methods of validity and reliability testing of questionnaires used in social and health science researches”, *Niger. Postgrad. Med. J.*, vol. 22, no. 4, p. 195.

[22] Sangoseni O, Hellman M, Hill C, “Development and Validation of a Questionnaire to Assess the Effect of Online Learning on Behaviors, Attitudes, and Clinical Practices of Physical Therapists in the United States Regarding Evidenced-based Clinical Practice”, *The Internet Journal of Allied Health Sciences and Practice*, 11(2), Article 7, 2013.

[23] Sargent, R.G., Balci, O., "History of verification and validation of simulation models," *2017 Winter Simulation Conference (WSC)*, Las Vegas, NV, USA, 2017, pp. 292-307, doi: 10.1109/WSC.2017.8247794.

[24] Harris, D. J., Bird, J. M., Smart, P. A., Wilson, M. R., & Vine, S. J. (2020). A framework for the testing and validation of simulated environments in experimentation and training. *Frontiers in Psychology*, *11*, 605

[25] Sargent,R.G., "An introduction to verification and validation of simulation models," *2013 Winter Simulations Conference (WSC)*, Washington, DC, USA, 2013, pp. 321-327, doi: 10.1109/WSC.2013.6721430.

[26] Sargent, R.G., "Verification and Validation Of Simulation Models: An Advanced Tutorial," *2020 Winter Simulation Conference (WSC)*, Orlando, FL, USA, 2020, pp. 16-29, doi: 10.1109/WSC48552.2020.9384052.

[27] Law AM, “How to build valid and credible simulation models”, *In Proceedings - Winter Simulation Conference*, vol. 2005, pp. 24–32.

[28] MITRE, “Verification and Validation of Simulation Models”, *MITRE Systems Engineering Guide*, pp 461-469, 2014.

[29] Harmon SY, Youngblood SM, “A Proposed Model for Simulation Validation Process Maturity”, *J. Def. Model. Simul. Appl. Methodol. Technol.*, vol. 2, no. 4, pp. 179–190.

[30] Wang Z, Lehmann A, “A Generic Maturity Model for Verification and Validation of Modeling and Simulation Applications'', In: Tan G., Lehmann A., Teo Y., Cai W. (eds) Methods and Applications for Modeling and Simulation of Complex Systems. AsiaSim 2019. Communications in Computer and Information Science, vol 1094. Springer, Singapore. doi: 10.1007/978-981-15-1078-6_6, 2019.

[31] Klügl F, “A validation methodology for agent-based simulations,” *Proc. ACM Symp. Appl. Comput.*, no. January 2008, pp. 39–43, doi: 10.1145/1363686.1363696, 2008.

[32] Louloudi A, Klugl F, “Immersive face validation: A new validation technique for agent-based simulation”, *2012 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2012*, pp. 1255–1260, 2012.

[33] Pace DK, “Modeling and Simulation Verification and Validation Challenges”, *Johns Hopkins Apl Technical Digest*, Volume 25, Number 2 (2004), 2004.

[34] Bharathy GK, Silverman B, “Validating agent based social systems models”, *Proc. - Winter Simul. Conf.*, no. 2002, pp. 441–453, doi: 10.1109/WSC.2010.5679142, 2010.

[35] Hatip A., Durak U., “Simülasyon Projeleri İçin Model Doğrulama Geçerleme Pratikleri“, *USMOS 2012- National Defense Applications Modeling Simulation Conference*, Ankara, 2012.

[36] DoD, “Technical Report: VV\&A Recommended Practices Guide: Conceptual Model Development and Validation (RPG Reference Document 11/30/00)“. “Validation Methods”, *https://vva.msco.mil/Ref_Docs/Val_Lawref/Val-LawRef-pr.pdf*, 2000.

[37] DoD, “Technical Report: VV\&A Recommended Practices Guide: V\&V Techniques (RPG Reference Document 8/15/01). “Informal V\&V Techniques”, *https://vva.msco.mil/files/ST05_VV_Techniques.pdf*, 2001.

[38] Dunn C, “Face Validation : From Concept To Concrete Process”, *The Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, Paper no. 12218, pp. 1–9, 2012.

[39] Tako AA, “Model development in discrete-event simulation: Insights from six expert modelers”, *Proceedings of the 2011 Winter Simulation Conference (WSC)*, 2011, pp. 3923-3934.

[40] Pereira TF, Pugliese LF, Montevechi JAB, Oliveira MLM, Robinson S, Banerjee A, "Methodology For The Management Of Discrete Event Simulation Projects Based On Pmbok®: Action Research In A High-Tech Company", *2018 Winter Simulation Conference (WSC)*, pp. 4002-4013.

[41] GEMED Warfare Effectiveness Analysis Model Project, *https://www.stm.com.tr/uploads/docs/1628678551_stm072urungrupfoylerisimulasyoneng.pdf*, (Last access: 29.03.2023).

[42] Balci, O., “Quality Indicators Throughout the Modeling and Simulation Life Cycle”, In: Yilmaz, L. (eds) *Concepts and Methodologies for Modeling and Simulation. Simulation Foundations, Methods and Applications.* Springer, Cham., 2005. https://doi.org/10.1007/978-3-319-15096-3_9

[43] McCall, J.A, Richards, P.K., Walters, G.F,”Factors in Software Quality”, *Final Tech Report, RADC-TR-77-369*, Rome Air Development Center,Griffith Air Force Base, 1977.

[44] ISO/IEC 25010:2011 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733 (Last access: 29.03.2023).