The Simulation Fidelity (SiFi) Scale: A Task Centric Approach to Defining Fidelity Requirements for Research and Acquisition

Dylan Bush, Christopher Lamb
Georgia Tech Research Institute
Atlanta, GA
dylan.bush@gtri.gatech.edu,
christopher.lamb@gtri.gatech.edu

ABSTRACT

The question of the relationship between level of fidelity and training effectiveness has been repeatedly posed in the training literature over the previous decades. Specifically, to what extent does increasing the fidelity of a training system improve or not improve training effectiveness. As research aimed at answering this question is aggregated in order to conduct larger meta-analyses, a problem arises that affects both researchers and acquisition professionals alike. That is, the training community lacks an accepted, standardized approach to decomposing and defining fidelity in such a way that can be generalized across different training effectiveness evaluations and different training systems. For the researcher, this gap creates an issue when attempting to aggregate the results of multiple training effectiveness evaluations when each of the various authors is using their own operationalized definition of *high* and *low* fidelity. For the acquisition professional, this gap creates an issue when attempting to develop fidelity requirements for a future training system or when attempting to define the fidelity of an existing training system. This paper outlines the development of a novel approach to defining the fidelity of a training system using the Simulation Fidelity (SiFi) scale. The SiFi scale is a 6-point scale used to rate the fidelity of a training system with respect to its real-world referent. Ratings from the SiFi scale can be used to construct a *descriptive* fidelity model (i.e. the level of fidelity of an existing system) or to construct a *prescriptive* fidelity model (i.e. the required level of fidelity for a future system). In addition to real-world examples, previous empirical evaluations of the scale are discussed. The SiFi scale and approach offers a standardized means to both defining training system fidelity in the context of training effectiveness research as well as defining and developing requirements for training systems.

ABOUT THE AUTHORS

Dylan Bush *Mr. Bush* is a Human Systems Engineer at the Georgia Tech Research Institute. With previous experience researching training system effectiveness for NAVAIR and current work with both the U.S. Air Force and Army, Mr. Bush's primary focus of research is related to improving warfighter performance through the strategic and science-based implementation of virtual and augmented reality technologies in both training and operational contexts. Mr. Bush holds a MS in Human Factors Psychology from Embry-Riddle Aeronautical University.

Christopher Lamb *Dr. Lamb* is a Principal Research Scientist at the Georgia Tech Research Institute. Dr. Lamb's decades of experience spans military, industry, and academia in roles ranging from technology research and development to leading product and engineering teams in large enterprise environments. More recently, Dr. Lamb's research focus has been centered on developing methodologies to identify usage contexts that would objectively benefit from the integration of virtual and augmented reality technologies. Dr. Lamb holds a PhD in Cognitive Psychology from the University of North Carolina at Greensboro.

The Simulation Fidelity (SiFi) Scale: A Task Centric Approach to Defining Fidelity Requirements for Research and Acquisition

Dylan Bush, Christopher Lamb
Georgia Tech Research Institute
Atlanta, GA
dylan.bush@gtri.gatech.edu,
christopher.lamb@gtri.gatech.edu

INTRODUCTION

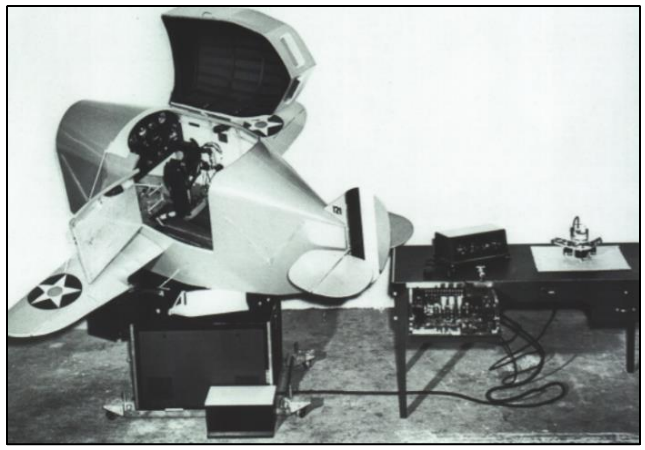


Figure 1: Link Trainer from American Society of Mechanical Engineers (2000)

The origins of modern training simulation are often credited to the aviation community. Specifically, the debut of Edward Link's flight trainer (Figure 1) in 1929 would go on to revolutionize not just the world of aviation training, but the training community at large (Allen, 1993). The success of the Link Trainer in the decades that followed led to significant interest in improving the technology and engineering realism of training simulators. Moreover, the desire for modern, realistic training simulation has not remained limited to the aviation industry. The legacy of Link's innovations can be seen in a range of higher risk industries including public safety, military, medicine, and process control (Bennell and Jones, 2004; Champney et al., 2017; Okuda et al., 2009; Patle et al., 2019). As the prevalence and capabilities of training simulations have continued to advance, a question has persisted in the training literature: that is, what level of realism is enough? And more specifically, given the associated cost increase that generally comes with higher realism, when does more realism begin to result in diminishing returns with respect to training effectiveness (Lapkin & Levett-Jones, 2011; Miller, 1954)? Answering this question requires an examination that goes beyond the engineering goal of higher and higher realism and instead focuses on the training goal of identifying what level of realism is best suited to simulating the tasks at hand given the overall context of the training. In addition, consensus must be reached in regards to how simulation realism is ultimately defined.

The American Psychological Association (APA) defines *simulation training* as "a technique in which trainees learn a complex or hazardous task by practicing with a replica of the task. This may involve the use of computer simulations, mechanical training aids, or plausible but fictitious scenarios (e.g., business games or case methods)." Two terms that stand out in this description are "replica" and "plausible." Both terms imply the necessity for some degree of realism or congruency between the simulation context and its associated real-world context. What is not explicitly defined, however, is how one would determine the degree of realism required for effective simulated training or how to quantify that requirement. Moreover, this definition suggests "complex" or "hazardous" training tasks as being ideal candidates for simulated training while offering examples of potential mediums for the delivery of the simulated training (e.g. computer simulations or mechanical training aids). Once again, missing from the definition is a description of the means with which one would determine what tasks are suitable for simulated training and more specifically, what medium of simulation is most appropriate for the specific training task. To be clear, the authors' intent is not to critique or refute the APA's definition of *simulation training*, but rather identify a gap in the literature with regards to how training tasks and training simulation technology is described, defined, and ultimately integrated in the most effective manner.

SIMULATION FIDELITY

Defining Fidelity

The term “fidelity” carries slightly different meaning depending on the context with which it is used. For example, the term could refer to loyalty to a person or cause, accuracy of information, or the degree to which one thing replicates another (Merriam-Webster). In the context of simulation, this means that terms like “replica” or “plausible” used in the aforementioned APA definition of *simulation training* are ultimately referring to the level of fidelity of the training simulation. That is, the relationship between the simulated elements in the training context and their associated elements in the real-world context (Hays & Singer, 1989). In general, this relationship is often described as being either *high fidelity* or *low fidelity.* While not objectively defined, *high fidelity* is generally used to describe a relationship between simulation context and real-world context that is more realistic in nature with less incongruencies in form or function. Conversely, *low fidelity* is generally used to describe a relationship that is less realistic in nature with more incongruencies in form of function. As such, simulation fidelity as it is traditionally viewed is very much a relative and, by extension, subjective term (see Liu et al., 2009; Schricker et al., 2001).

Categorizing Fidelity

Categorizing fidelity into two, discrete groups (i.e. *high* and *low*) is certainly convenient, but unfortunately the concept of simulation fidelity is likely too complex for such a basic categorization (Beaubien & Baker, 2004). A more accurate decomposition of simulation fidelity requires acknowledgement of the multi-dimensional aspects of simulation fidelity and more specifically, the multi-modal nature of human perception and subsequent cognition (see Meyer et al., 2012). An early attempt by Miller (1954) to define these dimensions includes the concepts of *engineering simulation* and *psychological simulation.* In Miller’s view, *engineering simulation* refers to the physical properties of the simulation (i.e. the appearance of cockpit gauges or the feel of the throttle) while *psychological simulation* refers to the more abstract concepts, knowledge, or skills that are being trained in the simulation (Miller, 1954). This distinction is further echoed in the perspectives presented by Hays and Singer (1989). The authors note the importance of emphasizing the psychological perspective when designing simulation systems rather than only the engineering perspective. This is largely due to the fact that the engineering perspective is almost always focused on attaining the highest levels of fidelity possible rather than the focusing on the level of fidelity that is best suited to the goal of the simulation (Hays & Singer, 1989). Work by Allen et al., (1991) further categorizes fidelity into two groups; *physical fidelity*, or how well the simulation replicates the appearance of the real-world system and *functional fidelity*, or how well the simulation replicates the response, output, or behavior of the real-world system. Further, this work also introduces a means to define different levels of *physical* or *functional* fidelity as described in Table 1.

Table 1: Levels of Physical and Functional Fidelity adapted from Allen et al., (1991).

		Physical Fidelity		
		High	**Medium**	**Low**
Functional Fidelity	**High**	All physical components present / All output devices function	Includes physical components and mock-ups / All output devices function	Includes mock-ups only / All output devices function
	Medium	All physical components present / Some system feedback provided	Includes physical components and mock-ups / Some system feedback provided	Includes mock-ups only / Some system feedback provided
	Low	All physical components present / No system feedback provided	Includes physical components and mock-ups / No system feedback provided	Includes mock-ups only / No system feedback provided

Later work by Talbot and Walker (1996) suggests that the traditional paradigm of viewing fidelity purely as the physical or functional match between the simulation and the real-world is overly simplistic. The authors argue that each training task should be analyzed according to the importance of each fidelity area and corresponding sub-elements (i.e. *spatial*, *tactile*, and *appearance* for *physical fidelity*; *content, function,* and *response* for *functional fidelity; sound, motion*, and *ambience* for *environmental fidelity*). The relative importance of the fidelity sub-elements is rated on a scale of 0-3 indicating the importance of full replication of a particular element and its impact on achieving the overall goal(s) of training (Talbot & Walker, 1996). This concept is examined further by (Meyer et al., 2012) who contend that emphasis should be placed on the multi-modal aspect of human sensory perception and its relationship to human performance during the execution of specific tasks. In other words, fidelity should not be solely focused on simply replicating all physical and functional aspects of the real-world. Rather, focus should be placed on replicating the sensory components of the real-world in such a way that supports successful task execution within the simulation and ultimately supports the overall intent of the training (Meyer et al., 2012). By focusing on what sensory components are required for task execution combined with the ultimate intent of the training, it is possible that certain sensory components are replicated with a higher degree of fidelity while others are replicated to a lower degree or not at all without sacrificing training transfer (Cooper et al., 2018; see Grierson 2014). More recently, Ye et al., (2020) proposed a fidelity framework called the General Conceptual Framework of Fidelity (GCFF). Developed within the application context of Virtual Reality (VR) based Serious Games, the GCFF consists of two overarching categories of fidelity; *objective fidelity* and *subjective fidelity* (Ye et al., 2020). Each overarching category is then further decomposed into sub-categories as illustrated in Table 2. A noted difference between this approach and previous approaches is the inclusion of fidelity dimensions that go beyond just the physical and functional features of the simulation to include more abstract, human-centric components such as *conceptual* and *emotional fidelity* (Ye et al., 2020).

Table 2: GCFF adapted from Ye et al., (2020)

Fidelity Category	Fidelity Sub-category	Description
Objective fidelity	Physical fidelity	All internal properties of an SSG system that can be described through diverse disciplines, including physics, engineering, chemistry, and biology.
	Functional fidelity	Various aspects related to learner-system interaction.
Subjective fidelity	Sensory fidelity	Various human senses resulting from physical characteristics of an SSG system.
	Conceptual fidelity	Further understanding, reflection, description, definition, and imagination of human beings toward an object or a process, which usually can be described by disciplines like linguistics, mathematics, and philosophy.
	Emotional fidelity	Holistic feelings of learners when they are experiencing the process in an SSG, as well as their responses to the system with different emotions.

Defining Levels of Fidelity

In the 30+ years separating the work by Allen et al., (1991) and the publication of the present paper, the training and simulation research community has yet to progress towards a standardized means to define levels fidelity beyond highly subjective and difficult to generalize terms such as *low*, *medium*, and *high* fidelity (see Burnett, 2008; Grierson, 2014; Schricker et al., 2001; Tun et al., 2015). However, this is not due to a lack of acknowledgement or attempts at resolving the issue. For example, Talbot and Walker (1996) propose a ranking structure based on the assignment of a Fidelity Factor to each sub-task defining the level of replication required to accomplish the goals of the training. While not a direct measure of different levels of fidelity, this approach does present the importance of defining fidelity requirements based on the needs of the specific task (Talbot & Walker, 1996). An alternative scale-based approach is found in the Model Fidelity Scale proposed by Burnett (2008). In an effort to add standardized meaning to terms like *high* or *low* fidelity, the Model Fidelity Scale consists of 11 levels (MFS0 – MFS10) indexed against the Live, Virtual, and Constructive (LVC) simulation paradigm. For example, a fidelity level MFS0 would correspond with a fully *Live* simulation while a fidelity level of MFS10 would correspond to a fully *constructive* simulation (Burnett, 2008). More recently, Tun et al., (2015) attempt to reduce the ambiguity of *high, medium, and low* fidelity by defining minimum simulation elements across three categories of fidelity identified as critical to medical simulation training. For example, a simulation with high patient fidelity would require inclusion of a fully functional patient simulator while a low patient fidelity simulation would require a limited anatomical and/or physiological representation of a patient (Tun et al., 2015).

Research Problem

The ambiguity in terminology used to describe fidelity as *high, medium, or low* presents a challenge to researchers interested in the domain of training effectiveness. This issue has become especially evident in the context of virtual reality (VR) based simulation as interest in the technology as a training mechanism continues to grow (see Lewis & Livingston, 2018; McCoy-Fisher et al., 2019; Mishler et al., (2022). For example, suppose a researcher is interested in conducting a comprehensive meta-analysis investigating the effectiveness of virtual reality-based training similar to the analysis conducted by Kaplan et al., (2020). The difference in this case being that the researcher is interested in the effect of different levels of fidelity on training effectiveness rather than type of immersive technology (i.e. virtual reality vs. augmented reality). The researcher would certainly be able to categorize results from the extant literature using standard terms like *high, medium, or low* fidelity. However, unlike Kaplan et al., (2020), who could organize their data based on standard definitions of immersiveness and hardware configurations, the researcher would have to rely on the assumption that every author who describes an experimental group as *high* or *low* fidelity is using the same operational definition of the term. In other words, the researcher (and conversely, the authors whose work they are referencing) have no means to define what *high, medium, or low* fidelity actually means in a way that can be generalized across studies. A simulation configuration defined as *high fidelity* by one author may actually be more congruent with the simulation configuration defined as *medium fidelity* by another author. This makes answering questions such as "Is high fidelity virtual reality training more effective than low fidelity virtual reality training?" almost impossible to answer in a comprehensive manner.

Acquisition Problem

The problematic nature of ambiguous terms such as *high, medium, or low* fidelity is not limited to the training effectiveness researcher. The acquisition professional must be able to define requirements for the development of new simulation systems as well as for the purchase of existing simulation systems. When it comes to fidelity, the acquisition professional needs a means to not only define what level of fidelity is expected in their requirements documentation, but also need a way to measure what level of fidelity is supported in an existing simulation system. For example, suppose an acquisition professional needed to develop fidelity requirements for a VR aviation trainer targeted for brand new flight students. The acquisition professional could conclude that a *low-medium* fidelity configuration is the best option (see Alessi, 1988), but how do they define that level of fidelity in a way that can be understood and leveraged by both procurement officials and developers? Ultimately, the acquisition professional arrives to the same road block as the researcher in the previous example. That is, the ambiguous terms often used to describe fidelity require modification that is standardized and generalizable in order to better define what is actually meant by *high, medium,* or *low* fidelity.

SIFI APPROACH

The SiFi approach consists of both a framework and associated scale to categorize and define the fidelity of an existing simulation system (i.e. *descriptive* model) as well as to rate the fidelity required to simulate a set of training tasks (i.e. *prescriptive* model). Implementation of the SiFi approach occurs in two phases. Phase 1 consists of the researcher or acquisition professional analyzing a set of training tasks and categorizing the associated task elements according to the fidelity components and sub-components included in the SiFi framework (see table 3). For the *descriptive* model, phase 2 consists of the researcher or acquisition professional applying ratings to each categorized task element using the SiFi scale (see table 4) based on the level of fidelity supported by the simulation system being investigated. For the *prescriptive* model, phase 2 would instead consist of the researcher or acquisition professional applying ratings to each categorized task element using the SiFi scale based on the level of fidelity required of the simulation system to be developed or acquired.

The process for calculating the final fidelity ratings using the SiFi scale is the same for both the *descriptive* and *prescriptive* model approaches. Once the researcher or acquisition professional has finished applying a fidelity rating to all task elements, they will be left with a series of values ranging from 0-5 across all 5 fidelity sub-components (i.e. visual, auditory, tactile, user interactions, and system behaviors). While each level of the SiFi scale is presented as a numerical value, the levels themselves are ultimately discrete categories. As such, the authors recommend that scoring the fidelity ratings for each sub-component be done by calculating median scores (Stevens, 1946). Use and interpretation of the final fidelity ratings is dependent on which model approach is being used. For the *descriptive model*, the fidelity ratings are a measure of the level of fidelity a particular simulation system is capable of supporting with respect to the specific tasks to be trained. For the *prescriptive* model, the fidelity ratings are a measure of the level of fidelity a simulation system would be required to support with respect to the specific tasks to be trained. An important point to note in both cases is that the fidelity ratings are not a rating of the overall simulation system, but rather are a rating of the simulation system within the context of the tasks to be trained. This is due to the fact that the fidelity of a simulation system should ultimately be defined within the confines of the training tasks it is intended to support (see Meyer et al., 2012).

Table 3: SiFi Framework

Fidelity Component	Fidelity Sub-component	Description
Physical	Visual	Any visual stimuli associated with a task element and required for task completion
	Auditory	Any auditory stimuli associated with a task element and required for task completion
	Tactile	Any tactile stimuli associated with a task element and required for task completion
Cognitive	User Interaction	Any user manipulation of a task element
	System Behavior	Any output or response of a task element

Table 4: SiFi Scale

Generic Term	Scale Level	Level Description
Low fidelity	L0	Element not present
	L1	Element recognizable as placeholder
Medium fidelity	L2	Element recognizable with noticeable incongruencies
	L3	Element recognizable with some incongruencies
High fidelity	L4	Element realistic with minimal incongruencies
	L5	Element indistinguishable from real-world form

REAL-WORLD APPLICATION

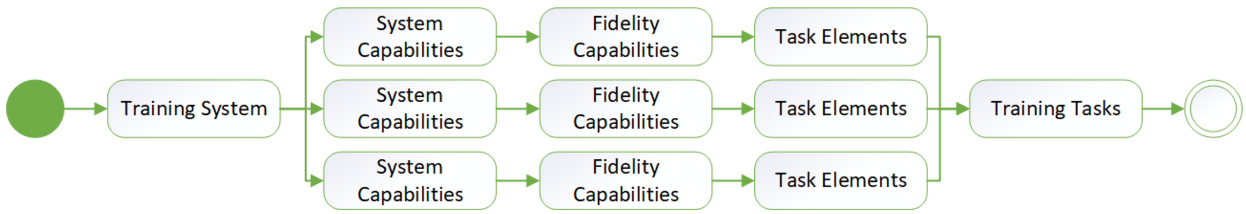
In order to more adequately present and further refine the proposed concept of the SiFi approach, the authors developed both a *descriptive* and *prescriptive* model using a VR-based F16 simulation system as well as a set of hypothetical training tasks relevant to the F16. The sections below describe the methods and results from the analyses to produce both models as well as data from an empirical evaluation of the interrater reliability of the SiFi scale.

Descriptive Model

The concept of the *descriptive* model (figure 2) is centered around the analysis of an existing simulation system with respect to its ability to support the level of fidelity required by the to-be-trained tasks of interest. That is, the *descriptive* model serves as a means to define the level and type of fidelity an existing simulation system is capable of supporting. The benefit of this approach is that it offers a means to define fidelity in a way that is generalizable across different simulation systems and training tasks.

The *descriptive* model developed in this analysis was based on a series of hypothetical training tasks executed using a VR-based F16 simulation system. As previously discussed, the intent of the *descriptive* model is to define the level of fidelity of an existing simulation system with respect to the training tasks it is intended to support. As such, the hypothetical training tasks used in the analysis were compiled before the authors had the opportunity to use the simulation system being evaluated. This was to ensure that tasks weren't accidently or intentionally chosen that would highlight particular strengths or weaknesses of the system. The output of this analysis was a complete model of the fidelity of the simulation system in a way that would allow for determination of the suitability of the system for supporting the intended training tasks. In addition, the model of the fidelity of the simulation system could also be used to better define the fidelity of the system for the purposes of training effectiveness evaluation.

Figure 2: Descriptive Model Concept



Method

In order to develop the list of hypothetical training tasks to be executed in the VR simulation, the authors consulted with an F16 Subject Matter Expert (SME) having both operational experience in the F16 as well as knowledge of simulation in aviation training. The hypothetical training tasks ultimately selected for inclusion in the analysis consisted of a series of low-level flight tasks, navigation tasks, and basic fighter maneuvers. These tasks were chosen based on their applicability to most military aviation tasks across different aircraft as well as requiring the participant to reference a variety of different instruments, gauges, and objects in the cockpit. The final task list consisted of a total of 5 primary tasks each decomposed into a series of sub-tasks. The 5 primary tasks included a straight and level flight task, a level turn, a slice back, a pitch back, and a split s. Each of these 5 primary tasks was decomposed into sub-tasks in order to identify required task elements. An example of this decomposition for a slice back maneuver is shown in table 5. Note, the list of task elements associated with each sub-task was not intended to be exhaustive for this initial analysis. Rather, the authors worked with the SME to identify key task elements that would be required to complete each maneuver and would generally be understood and recognized by individuals with at least some degree of aviation experience.

Table 5: Slice back task decomposition

Task 3: Slice Back		
Sub-task	***Description***	***Task element***
3.1	Proceed to heading 160	HSI
		Stick
3.2	Visually scan terrain	Obstacles in environment
		Terrain slope
3.3	Confirm adequate altitude (>10K ft.)	Altimeter
3.4	Roll into 135 deg. of bank and pull	Bank angle indicator
		Attitude indicator
		Stick
3.5	Adjust power as needed	Throttle
		ASI
3.6	Adjust G's as needed	G-force indicator
3.7	Roll out at opposite heading	HSI
		Stick

Once the primary tasks and sub-tasks were decomposed and their corresponding task elements identified, the next step consisted of categorizing each task element within the SiFi framework. For example, if completing sub-task 3.1 ("Proceed to heading 160") requires the user to visually reference information in the HSI, then that task element would be categorized into the visual fidelity sub-component of physical fidelity and the system behavior fidelity sub-component of cognitive fidelity (see table 3). In addition, the authors only categorized task elements into the corresponding components of the SiFi framework that were explicitly required for task completion. For example, the HSI in a real F16 cockpit includes both tactile and interaction sub-components (i.e. the pilot can touch and manipulate the instrument). However, execution of sub-task 3.1 only requires use of the visual and system behavior sub-components. As such, only those sub-components were assigned to the HSI task element for that sub-task. This distinction is crucial to maintaining the task-centric intention of the SiFi approach. Meaning, the assessment of fidelity is specifically based on what is required for task execution. The process of assigning task elements to their respective SiFi framework components was repeated for all task elements identified within the sub-tasks. An example of this categorization is shown in table 6 for task 3. Of note, certain task elements were repeated across different tasks and sub-tasks. While this may seem redundant, the authors intention in doing so was to include a "weighting" of the importance of certain task elements into the final fidelity score. For example, if a task element appears only once in a series of 50+ sub-tasks then its fidelity rating should likely not influence the final fidelity rating to the same degree as a task element that appears a dozen times.

Table 6: Slice back task element categorization

Task 3: Slice Back						
Sub-task	***Task Element***	***Visual***	***Auditory***	***Tactile***	***User Interaction***	***System Behavior***
3.1	HSI	Appearance of HSI	-	-	-	Heading data displayed
	Stick	-	-	Feel of the stick	Manipulate stick	-
3.2	Obstacles in environment	Appearance of outside environment	-	-	-	-
	Terrain slope	Appearance of terrain	-	-	-	-
3.3	Altimeter	Appearance of altimeter	-	-	-	Altimeter data displayed

Once all task elements were identified and categorized into the SiFi framework, two F16 SME's were enlisted to execute the tasks in a VR F16 simulator. The simulation was developed using the Unity game engine and presented in an HTC VIVE Pro headset. System components included an Alienware Aurora R12 desktop with an NVIDIA GeForce RTX 3090 graphics card. All system interactions were supported using a Thrustmaster hands on throttle-and-stick (HOTAS). In order to reduce bias, the SME's used to execute the training tasks were not involved in the creation of the tasks or categorization of the task elements into the SiFi framework. Following execution of the tasks in the VR F16 simulation, the SME's independently provided ratings for all identified and categorized task elements using the SiFi scale. In instances where a task element was repeated across different tasks, the SME's were instructed to provide a rating based exclusively on their perception of the task element during execution of each specific sub-task individually (i.e. rating a certain task element as a '1' for one sub-task doesn't automatically mean that element is a '1' for all sub-tasks). Once all task elements were rated, a final fidelity rating was calculated for each fidelity sub-component of the SiFi framework.

Results

The final results for the *descriptive* model for the VR F16 simulation were based on a total of 117 ratings of task elements across a total 5 primary tasks. Final fidelity ratings for SME 1 include visual: *1*, auditory: *n/a*, tactile: *2*, interaction: *2*, and behavior: *0*. Final fidelity ratings for SME 2 include visual: *1*, auditory: *n/a*, tactile: *2*, interaction: *2*, and behavior: *0*. A further breakdown of these results is presented in table 7. In addition, further analysis of interrater reliability (IRR) indicates strong reliability of the SiFi scale based on the data collected (Cohen's *k* of 0.81).

Table 7: VR F16 *Descriptive* Model

Tasks	**Visual**		**Auditory**		**Tactile**		**Interaction**		**Behavior**	
	SME 1	SME 2	SME 1	SME 2	SME 1	SME 2	SME 1	SME 2	SME 1	SME 2
Task 1	1.5	1	-	-	2	2	1	1	0	0
Task 2	1.5	1.5	-	-	2	2	2	2	0	0
Task 3	1	1	-	-	2	2	2	2	0	0
Task 4	1	1	-	-	2	2	2	2	0	0

Task 5	1	1	-	-	2	2	2	2	0	0
Total	**1**	**1**	**-**	**-**	**2**	**2**	**2**	**2**	**0**	**0**

Interpretation of the VR F16 simulation *descriptive* model suggests the system is capable of supporting tasks that require:

- *visual elements* that are recognizable as placeholders for their real-world referent (L1);
- *tactile elements* that are recognizable, but with noticeable incongruencies from their real-world referent (L2);
- *user interactions* that are recognizable, but with noticeable incongruencies from their real-world referent (L2);
- *system behaviors* that are non-functional or not present in the simulation (L0).

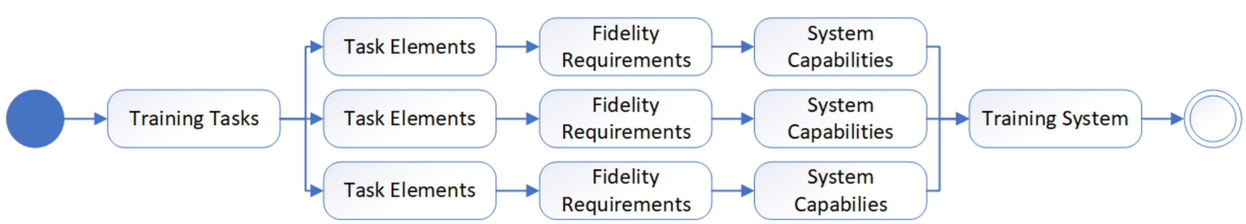
In aggregate, this implies the VR F16 simulation system would be most suitable to training tasks that do not require a significant level of realism in order to execute the training tasks or accomplish the indented training objectives. For the acquisition professional, this data is useful in determining the suitability of the system in supporting intended training requirements and objectives. For the researcher, this data is useful in better defining the fidelity of the simulation system when used in training effectiveness evaluations so that other researchers can replicate and/or aggregate results into larger meta-analyses.

Prescriptive Model

The concept of the *prescriptive* model (figure 3) is essentially the opposite perspective of the *descriptive* model. That is, the *prescriptive* model begins with a set of training tasks that are to be analyzed according to the level of simulation fidelity required to accomplish the tasks and overall training objective(s). In practice, the *prescriptive* model serves as a means to define the level and type of fidelity required by a simulation system before the system is developed or acquired. The benefit of this approach is that it seeks to ensure a match between the requirements of the to-be-trained tasks and the capabilities of the to-be-developed or to-be-acquired simulation system before development resources are allocated.

The *prescriptive* model developed in this analysis was based on a decomposition of tasks required for completing Pre-flight through Departure procedures. As previously discussed, the *prescriptive* model can be thought of as a requirements document of sorts that defines the level of fidelity required of the simulation system across each dimension included in the SiFi approach (i.e. Visual, Auditory, Tactile, etc.). Given the task-oriented nature of the *prescriptive* model, the analysis does not include any references to specific systems or system components. Rather, the model describes the fidelity requirements of the specific to-be-simulated tasks.

Figure 3: Prescriptive Model Concept



Method

In order to develop the *prescriptive* model, the research team leveraged support from the same SME who supported the *descriptive* model development effort. In contrast to the *descriptive* model, the *prescriptive* model does not include an existing simulation system as part of the model development. Rather, it relies on a set of to-be-trained tasks that are to be analyzed according to their associated training context (i.e. training requirements, learning objectives, learner characteristics, etc.). As such, the research team developed a hypothetical narrative to outline the primary user of the model, training context, and training constraints. This narrative is outlined below:

- ***Model User***: Acquisition Professional
- ***Training Context***: Part-Task Trainer to support practice and rehearsal of Pre-flight and Departure procedures before graded events in the real aircraft (PreD-PTT)
- ***Training Constraints***: Minimum effective fidelity for student aviators to accomplish training context

Following creation of the narrative, the research team worked with the SME to develop a realistic list of procedures starting from initial cockpit checks through engine start, taxi, takeoff, and climb. Once the procedures were finalized, the tasks were decomposed and analyzed in the exact same manner as described in the previous section (see table 5 and table 6 in the *descriptive* model section). Following the decomposition of tasks into their subsequent task elements and categorization according to the SiFi framework, the research team and the SME rated the required level of fidelity of each element in the SiFi framework using the SiFi scale. The results of this analysis are presented in the following section.

Results

The final results for the *prescriptive* model of the PreD-PTT were based on a total of 841 ratings of task elements across a total 9 primary tasks. Final fidelity ratings include visual: 4, auditory: 5, tactile: 2, interaction: 3, and behavior: 2. A further breakdown of these results is presented in table 8.

Table 8: PreD-PTT *Prescriptive* Model

Tasks	Visual	Auditory	Tactile	Interaction	Behavior
Task 1	4	-	2	3	2
Task 2	4	-	2	3	2
Task 3	4	-	2	4	4
Task 4	4	-	2	4	2
Task 5	4	5	3	4	4
Task 6	4	-	3	4	4
Task 7	4	-	2	3	2
Task 8	4	-	4	4	2
Task 9	4	-	3	4	4
Total	**4**	**5**	**2**	**3**	**2**

Interpretation of the PreD-PTT *prescriptive* model suggests the system must be capable of supporting tasks that require:

- *visual elements* that are realistic with minimal incongruencies from their real-world referent (L4);
- *auditory elements* that are indistinguishable from their real-world referent (L5);
- *tactile elements* that are recognizable, but with noticeable incongruencies from their real-world referent (L2);
- *user interactions* that are recognizable, but with some incongruencies from their real-world referent (L3);
- *system behaviors* that are recognizable, but with noticeable incongruencies from their real-world referent (L2).

In aggregate, the interpretation of the *prescriptive* model suggests the need for a simulation system that is capable of supporting a higher degree of physical fidelity (i.e. Visual and Auditory elements) with a moderate degree of cognitive fidelity (i.e. Interactions and Behaviors). Based on this result, one can conclude that the *descriptive* system model described in the previous section would not be a suitable candidate for acquisition based on the requirements defined in the present *prescriptive* model.

SUMMARY

The need for a standardized method to categorize and rate simulation fidelity is by no means a new concept. Miller (1954) notes the significance of distinguishing between physical elements of a simulation (i.e. engineering simulation) and the psychological stated elicited by the simulation (i.e. psychological simulation). Later work by Hays and Singer

(1989) further echo and expand this sentiment with the acknowledgement that distinction must be maintained between the engineering and psychological perspectives of simulation design in order to prevent the goals of the former from overshadowing the needs of the latter. In the following decades, numerous attempts were made to refine the distinction between "physical" and "psychological" simulation into multi-dimensional frameworks of simulation fidelity. From early attempts by Allen et al., (1991) to categorize fidelity as "physical" and "functional" and assign ratings such "high," "medium," or "low" to more recent attempts by Ye et al., (2020) to expand the definition of fidelity across 5 distinct dimensions, there is still no uniformly accepted, standardized approach to categorizing and ranking levels of simulation fidelity. In an effort to close this gap, the authors have presented a proposed approach for categorizing and rating simulation fidelity using the SiFi framework and companion scale. The presented approach consolidates and expands upon previous efforts to categorize fidelity by presenting a framework that views fidelity through the lens of *physical* and *cognitive* fidelity. Additionally, each fidelity component is decomposed into across 5 sub-components (i.e. visual, auditory, and tactile for physical; interaction and behavior for cognitive). In tandem with the companion, 6-point rating scale (i.e. SiFi scale), The SiFi approach offers a standardized method to both define and rate simulation fidelity. The examples provided in this paper present two uses of the SiFi approach that offer benefit to both the researcher and acquisition professional alike. That is, the *descriptive* model, which provides a description of the fidelity an existing simulation system is capable of supporting as well as the *prescriptive* model which provides requirements for the level of fidelity a simulation system should be capable of supporting.

CONCLUSION

The presented approach offers a method to categorize and rate the level of simulation fidelity across 5 sub-components of fidelity. The authors contend that this approach can be leveraged as a standardized means of defining fidelity for both research and acquisition purposes. That said, developing a scale, such as the one proposed in this paper, requires more than just a review and consolidation of previous literature. The scale should also be empirically evaluated for both its reliability and validity to ensure the scale actually measures what it purports to measure. To date, the authors have conducted an analysis of Interrater Reliability for the *descriptive* model with favorable results (Cohen's k of 0.81). While promising, the SiFi approach still requires rigorous examination of validity in order to fully gauge its overall utility. The authors intend to examine validity in a similar manner to the approach outlined by Meyer (2012) using overall task performance as the primary variable of interest.

ACKNOWLEDGEMENTS

The research team would like to sincerely thank our colleagues who participated in the development and analysis of the SiFi framework described in this paper. Their subject matter expertise was crucial in ensuring the tasks included in the analysis were as accurate and realistic as possible.

REFERENCES

Allen, J., Buffardi, L., & Hays, R. (1991). *The relationship of simulator fidelity to task and performance variables*. US Army Research Institute Office of Basic Research. Alexandria, VA.

Allen, L. (1993). Evolution of flight simulation. In *AIAA Conference on Flight Simulation and Technologies*, Monterey, CA. https://doi.org/10.2514/6.1993-3545

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40–47.

American Psychological Association. (n.d.). Simulation training. In *APA dictionary of psychology*. Retrieved January 27, 2022, from https://dictionary.apa.org/simulation-training

American Society of Mechanical Engineers (2000). *The Link Flight Trainer* [Image]. Roberson Museum and Science Center, Binghamton, New York.

Bennell, C., & Jones N., (2004). *The Effectiveness of Use of Force Simulation Training Final Report*. Psychology Department, Carleton University, Ottowa, Canada.

Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: how low can you go?. *Quality & safety in health care, 13*(Suppl 1), i51–i56. https://doi.org/10.1136/qhc.13.suppl_1.i51

Burnett, E. (2008). A Proposed Model Fidelity Scale. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*. Honolulu, HI. https://doi.org/10.2514/6.2008-6689

Champney, R. K., Stanney, K. M., Milham, L., Carroll, M. B., & Cohn, J. V. (2017). An examination of virtual environment training fidelity on training effectiveness. *International Journal of Learning Technology*, *12*(1), 42-65.

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018) The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS ONE 13(2):* e0191846. https:// doi.org/10.1371/journal.pone.0191846

Grierson, L., (2014). Information processing, specificity of practice, and the transfer of learning: considerations for reconsidering fidelity. *Adv in Health Sci Educ 19***,** 281–289. https://doi.org/10.1007/s10459-014-9504-x

Hays, R. T., & Singer, M. J. (1989). Simulation fidelity as an organizing concept. In *Simulation fidelity in training system design* (pp. 47-75). Springer, New York, NY.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2020). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Human Factors*. doi:10.1177/0018720820904229

Lapkin, S., & Levett-Jones, T. (2011), A cost–utility analysis of medium vs. high-fidelity human patient simulation manikins in nursing education. *Journal of Clinical Nursing*, *20*, 3543-3552. https://doi.org/10.1111/j.1365-2702.2011.03843.x

Lewis, J., & Livingston, J. (2018). Pilot Training Next: Breaking institutional paradigms using student-centered multimodal learning. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL.

Liu, D., Macchiarella, N. D., & Vincenzi, D. A. (2009). Simulation Fidelity. In D. A. Vincenzi, J. A. Wise, M. Mouloua, & P. A. Hancock (Eds.), *Human Factors in Simulation and Training* (pp. 61-73). CRC Press.

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2019). *Student Naval Aviation Extended Reality Device Capability Evaluation*. Naval Air Warfare Center Training Systems Division. Orlando, FL.

Merriam-Webster Dictionary (n.d.). Fidelity. In *Merriam-Webster Dictionary.* Retrieved January 27, 2022, from https://www.merriam-webster.com/dictionary/fidelity

Meyer, G.F., Wong, L.T., Timson, E., Perfect, P., & White, M. D. (2012). Objective Fidelity Evaluation in Multisensory Virtual Environments: Auditory Cue Fidelity in Flight Simulation. *PLoS ONE 7(9):* e44381. https://doi.org/10.1371/journal.pone.0044381

Miller, R. B. (1954). *Psychological Considerations in the Designs of Training Equipment.* Wright Air Development Center, Wright Patterson AFB, OH.

Mishler, A., Severe-Valsaint, G., Natali, M., Seech, T., McCoy-Fisher, C., Cooper, T., & Astwood, R. (2022). *Project Avenger Training Effectiveness Evaluation*. Naval Air Warfare Center Training Systems Division. Orlando, FL.

Okuda, Y., Bryson, E. O., DeMaria, S., Jr, Jacobson, L., Quinones, J., Shen, B., & Levine, A. I. (2009). The utility of simulation in medical education: what is the evidence?. *The Mount Sinai journal of medicine, New York*, *76*(4), 330–343. https://doi.org/10.1002/msj.20127

Patle, D. S., Manca, D., Nazir, S., & Sharma, S. (2019). Operator training simulators in virtual reality environment for process operators: a review. *Virtual Reality*, *23*(3), 293-311. https://doi.org/10.1007/s10055-018-0354-3

Ponnusamy, S., Albert, V., & Thebault, P. (2014). A simulation fidelity assessment framework. In *2014 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH)* (pp. 463-471). IEEE.

Schricker, B., Franceschini, R., & Johnson, T. (2001). Fidelity evaluation framework. In *Proceedings of the IEEE 34th Annual Simulation Symposium*, Seattle, WA.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, 103*, 677-680.

Talbot, N., & Walker, A. (1996). Behavioral fidelity requirements analysis, In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL.

Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining Simulation Fidelity for Healthcare Education. *Simulation & Gaming*, *46*(2), 159–174. https://doi.org/10.1177/1046878115576103

Ye, X., Backlund, P., Ding, J., & Ning, H. (2020). Fidelity in simulation-based serious games. *IEEE Transactions on Learning Technologies, 13(2)*, 340-353. https://doi.org/10.1109/TLT.2019.2913408