51st International Conference on Environmental Systems
10-14 July 2021, St. Paul, Minnesota
ICES-2022-133

Human-in-the-Loop Evaluations: Process and Mockup Fidelity

Jackelynne Silva-Martinez[1], Gordon Vos[2], Jennifer Boyer[3], Robert Durkin[4], William Foley[5], Sarah Margerum[6], Kritina Holden[7], Victoria Smith[8], Leah Beebe[9], Chris Van Velson[10]
NASA Johnson Space Center, Houston, TX, 77058 & KBR Wyle Laboratories Inc, Houston, TX 77058

Human-in-the-loop (HITL) evaluations are iterative events used during design and development to identify issues with the implementation of human systems integration. HITL evaluations involve human test subjects who are user-representative participants performing activities with representative hardware, software, and procedures. The paper will describe the process by which HITL evaluations can be implemented in a program or project. This will include definitions for developmental and verification HITL evaluations, levels of mockup fidelity, and certification of HITL articles. The proposed process can be tailored depending on the type of HITL evaluation being conducted. It is highly recommended that findings are timely reported and incorporated in the hardware and/or software design.

Nomenclature

CDR	=	critical design review
CoFR	=	certification of flight readiness
CONOPS	=	concept of operations
DU	=	display unit
D&C	=	displays and controls
EDRO	=	exempt determination review official
EDU	=	engineering development unit
FEU	=	flight equivalent unit
FRR	=	flight readiness review
HCD	=	human-centered design
hi-fi	=	high fidelity
HITL	=	human-in-the-loop
HSI	=	human systems integration
IRB	=	internal review board
ISS	=	international space station
IVA	=	intra-vehicular activity
low-fi	=	low fidelity
med-fi	=	medium fidelity
PDR	=	preliminary design review
SME	=	subject matter expert

[1] Systems Engineering and Integration Lead & Gateway Human Systems Integration Lead, Human Systems Engineering and Integration Division, NASA, jackelynne.p.silva-martinez@nasa.gov
[2] Human Factors Technical Discipline Lead, Human Systems Engineering and Integration Division, NASA, gordon.a.vos@nasa.gov
[3] Orion Chief Health and Performance Officer Deputy, Human Health and Performance Directorate, NASA, jennifer.l.boyer@nasa.gov
[4] Human in the Loop Test Manager, Flight Operations Directorate, NASA, robert.j.durkin@nasa.gov
[5] Orion Human Engineering System Manager, Human Systems Engineering and Integration Division, NASA, william.a.foley@nasa.gov
[6] Orion Human Engineering Deputy System Manager, Orion Human Engineering, KBR, sarah.margerum@nasa.gov
[7] Human Factors Technical Fellow, Leidos, Kritina.l.holden@nasa.gov
[8] Gateway Human System Integrator, KBR, v.f.smith@nasa.gov
[9] Gateway Human Factors Task Lead, KBR, leah.n.beebe@nasa.gov
[10] Senior Medical Operations Integrator for Gateway and xEVA, NASA, christopher.m.vanvelson@nasa.gov

SRR	=	system readiness review
SR	=	safety review
TRR	=	test readiness review
VCN	=	verification closure notice
VISE	=	vehicle interface to suit equipment
VR	=	virtual reality

I. Introduction

HUMAN-in-the-loop evaluations (HITLs) are scenario-based events used during system design and development. Within the context of system engineering at NASA, HITL evaluations are leveraged to identify issues with human systems integration (HSI)[1], which seeks to ensure that systems are designed with the capabilities and limitations of the human in mind. They are used to demonstrate that the design and operational concept meets system requirements for effectiveness, efficiency, acceptability, and safety. HITL evaluations involve user-representative participants performing planned tasks with flight-representative hardware, software, and procedures. HITL evaluations are typically repeated with increasing system fidelity as part of a human-centered design (HCD) process[2]. Fidelity can be defined as the degree to which a simulation or physical mockup replicates a real-world, built-environment from the user's perspective, with respect to form, function, and user-interaction[3].

The history of HITL evaluations is a rich and storied one. Though informal evaluation of human designed systems has likely been going on from the times of prehistory (when an early hominid made the first stone tool and tested how well it worked), the formal and systematic evaluation of designed systems that are referred to as HITL dates back to the times of World War 2 and the modern beginings of human factors[4]. In those early days, human factors, a key domain within HSI, developed as a field in the evaluation of aviation systems, and quickly grew to include assessment of human systems of every sort. Test pilot evaluations of human interfaces (displays, controls, seats, cockpit layout, etc.) were a core part of this early work, and led down a direct path to the current work of HITL evaluation at NASA today. Many of NASA's current designs also include displays, controls, and workspace layout just as they did in early days, though the sophistication of the technologies included has certainly advanced over the past 80 years, and software and automation considerations are now front and center in NASA's focus as well.

Habitability, another HSI domain, is an important factor assessed during HITL evaluations, substantially supported by the field of space architecture, which covers how humans might occupy, use, and move within a space once it is built and habitable. Some examples of recent work at NASA are included in Figure 1, including evaluation of medical kits for the International Space Station (ISS), hand controller design for pressurized suit use (done in a low pressure glove box), and parabolic flight evaluation of the displays, controls, and seats for Orion. The use of HITL evaluations is a key aspect of the a mission/project design lifecycle, ensuring that human considerations are appropriately factored into system design.

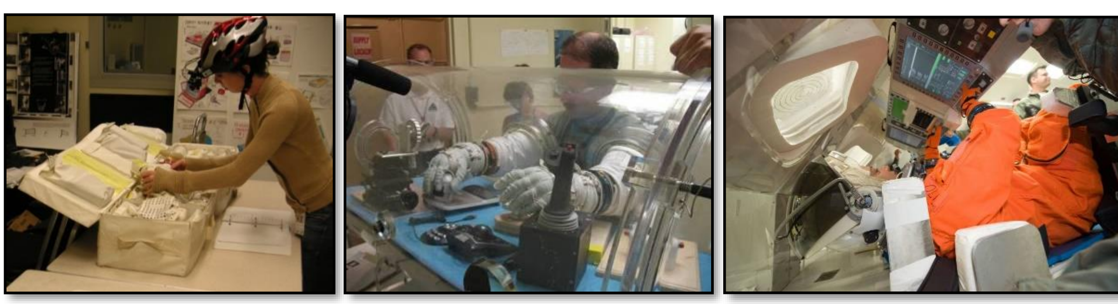

Figure 1. Examples of HITL evaluations from ISS Program (left, medical kit design evaluation) and Orion Program (center, glove-box evaluation of hand controller designs for use with pressurized gloves; and right, micro-gravity parabolic flight evaluation of display and control layout, window field of view, and seat design).

II. Types of Human-in-the-Loop Evaluations

A. Developmental HITL Evaluations

Developmental HITL evaluations are used to facilitate design and development work and are conducted to ensure a successful path to verification. They provide an opportunity to assist with technical trade decisions, refine the design and mature the concept of operations (CONOPS) based on human system performance before it is cost prohibitive to do so. These early HITL evaluations also serve to validate the System Readiness Review (SRR) requirements, as well as to flesh out issues associated with the vehicle human-system interfaces. Early in the project lifecycle, developmental HITL evaluations are done with low fidelity mockups or in virtual representations of early spacecraft design concepts[5]. Quick mockups are constructed using wood, foam-core, and/or with virtual reality (VR), allowing the Systems Engineering & Integration (SE&I) team and flight crew input on early concept design decisions.

Between Preliminary Design Review (PDR) and Critical Design Review (CDR), test procedures and mockups are incrementally updated to higher levels of fidelity, allowing for higher fidelity task representation. Whole cabin HITL evaluations can be conducted, looking at net habitable volume and other integrated tasks associated with the vehicle crew compartment. Part-task representations of critical items in the spacecraft interior, such as display/control elements are rendered in higher fidelity mockups. Developmental HITL evaluations help assess progress of the design toward meeting verification needs of requirements. Developmental HITL evaluations can be done from low fidelity to high fidelity mockups. Relating to displays, developmental HITLs are performed using the most up-to-date displays while crew walk through various scenarios and evaluate usability, handling qualities, or legibility aspects of the design. In development, this can be done with a laptop computer screen, progressing to higher fidelity as they become available. Integrated testing should be periodically done to ensure there are no glare or obstructions to displays with the physical hardware.

B. Verification HITL Evaluations

Verification HITL evaluations are conducted to confirm that the system complies with design requirements, and to provide confidence that the human will be capable of achieving the mission tasks required. Verification HITL evaluations are conducted to close requirements through verification closure notices (VCNs) and assess readiness for Certification of Flight Readiness (CoFR)[6,7,8]. Verification HITL evaluations occur in the highest fidelity mockup possible, including flight model if a prototype flight is delivered early enough in the design lifecycle to make changes should the verification HITL indicate that the design fails verification. Otherwise there is no schedule time allowed for fixing the design and achieving a passing verification. It is highly recommended that high-fidelity mockups be used for verification due to their lower cost and ability for the design team to make updates as needed prior to production of the flight model.

After CDR and up to Flight Readiness Review (FRR), verification requirements place increased rigor and restrictions on verification HITL evaluations. They typically have a required number of subjects based on applicable requirements. They must include a pre-test review of applicable requirements and whether test plans adequately meet these requirements. Verification HITL evaluations occur in the highest fidelity mockup possible with flight-like software and hardware. For more information on the fidelity validation refer to Section V. Relating to displays, verification HITLs should be performed on the most up-to-date hardware to ensure the performance of the hardware will support the operations of the crew for timing and response time. In addition, display screen fidelity and controller fidelity will impact the crew responses related to legibility or handling qualities. This is also usually coupled with a verification of the Avionics hardware.

C. Non-Test vs Test

Table 1 outlines test characteristics based on verification statements. It adds rigor to the criteria, especially for verification. Non-test evaluations are too informal for verification. They are done early in development or considered a demonstration.

Table 1. HITL Non-Test vs Test

Non-Test Evaluation	Test
Exploratory; discussion-driven, procedure-driven	Scripted; procedure-driven
Partial task	Scenario-based, realistic task
One or several participants (engineers, experts, crew-like, crew)	Minimum number of participants dependent on performance metric and fidelity (typically 5 or 10)
Primary data: comments, observations, recommendations	Primary data: human performance data [response time, errors (type, severity), usability, workload, acceptability] in addition to comments, observations and recommendations

III. Human-in-the-Loop Process

This process protects personnel in tests and training, ensuring evaluations are conducted in a safe manner which meet design and development needs, verification of program requirements, and training objectives. Each activity's risk is clearly defined and reduced to a manageable level. Accepting elevated risks, if necessary, requires institutional and program approval. This process promotes appropriate Subject Matter Experts (SMEs) participation on the HITL evaluations to provide feedback. It ensures time of the test participant is used efficiently. Figure 2 shows the process framework and timeframes for the strategic, tactical, and operational phases, and the deliverables at each stage[9]. The timeframe provided are guidelines, each program/project/mission may choose their own negotiated dates.

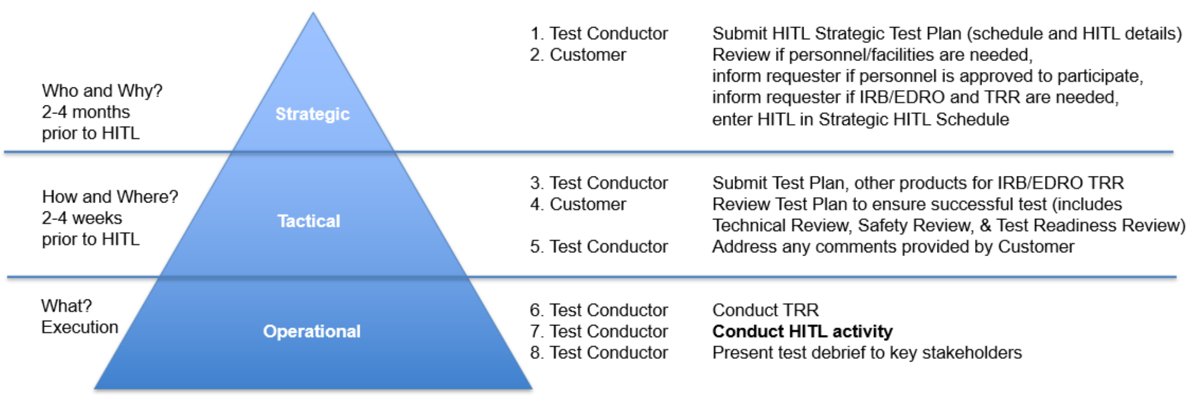

Figure 2. Proposed HITL Process Structure

A. Strategic

The Strategic HITL Test Plan describes the developer's plans for using HITL testing throughout their design process. The Strategy should include a description of test articles, test events, test objectives, test measures, and schedule. As detailed plans mature, individual test plans for each HITL event are written, which include specific test objectives, detailed descriptions of the test article(s), test setup, test subject selection and training, test instruments, and data collection and analysis methods. HITL test measures should include task errors (number and type), usability ratings, test subject comments, cognitive workload for assessment, and anthropometric analysis for the system design in order to identify effective design features and those that need changes or improvements.

Institutional Review Board (IRB) protocols need to be followed to ensure tests involving human subjects do not endanger the health, safety, or well-being of human subjects. All testing operations are conducted in an ethical manner. The NASA Exempt Determination Review Official (EDRO) provides a process to obtain a determination if a test is exempt from IRB review. IRB exemption does not mean Safety Review (SR) exemption nor Test Readiness Review (TRR) exemption[10].

B. Tactical

The customer reviews the test plan to identify concerns that might prevent test success, with three purposes: (a) The technical review is performed by appropriate cross-discipline SMEs who review the test plan technical content to ensure requirements compliance, and provide recommendations for test set-up and procedure, appropriate data collection, and analysis methods. (b) Medical Operations reviews the test plan for any medical needs and participates in the safety review. This includes an IRB Exempt Determination Review Official (EDRO) process and for determining need for health monitoring. (c) Flight operations reviews the test plan for any facilities readiness, and TRR needs.

The SR should be conducted prior to the TRR for complex and/or hazardous operations. The SR allows the appropriate SMEs to review the draft test plan and procedures allowing them to flag areas of potential concern from which the provider will need to provide additional information. This allows for level of risk mitigation agreement between the provider and the customer. It reduces the chance of rework or significant changes occurring at the TRR. To help with the process, there should be some level of coordination between the test conductor and the test facility owner/organization; and the facility owner should be involved in the SR and TRR processes.

C. Operational

The TRR verifies the testing environment is safe for the HITL test subjects and participants via a multi-program approved TRR process. The primary focus of the TRR is to ensure the safety of the test subjects and the personnel supporting the event. This includes a review of the hazard analysis to ensure all potential hazards associated with a test are identified and appropriate controls are in place. Additional mitigation controls, such as real time medical monitoring, may be required depending on the test activities. Once the TRR is approved, the test conductor can conduct the HITL evaluation.

The customer ensures proper personnel are involved to meet HITL objectives. This includes flight doctors, program or project stakeholders, observers, and test subjects. After completing each developmental HITL evaluation and/or failure of a verification HITL evaluation, the test conductor should check with test subjects to ensure: (a) data collected is accurate, (b) test subjects feedback provided during the test is consistent, and (c) team collecting the data is in sync with interpretation of results.

The operational phase final step is to present the results through a test report or test debrief to stakeholders. Test debriefs provide stakeholders the opportunity to provide technical feedback, design change concerns or other forward actions planned by the project that could prevent passing a requirement verification successfully.

IV. Human-in-the-Loop Mockup Fidelity

HITL testing relies upon having humans interact with concepts or prototypes that can be evaluated, including hardware, software, and/or functional integrated systems. As such, there are varying levels of fidelity in the concepts and prototypes being evaluated, typically defined using categories of low, medium, and high-fidelity mockups or flight simulators. Additionally, in recent years virtual reality (VR) simulations have been added to the collection of source material for HITL evaluations, and are often used when the concept is very early in the design lifecycle, or when the cost of physical mockup production is prohibitive, or in the case of minor changes which need to be evaluated in a very short timeframe.

A. Low Fidelity Mockups

The purpose of a low-fidelity (low-fi) mockup is to allow for testing and evaluation of the design, early and often across the systems engineering lifecycle, including the review of CONOPS for the physical use of the spacecraft. Factors assessed with a low-fidelity mockup may include net habitable volume, anthropometric accommodation, placement and orientation of components/stowage/interfaces, intra-vehicular (IVA) activities and habitability in general.

Low-fidelity mockups are expected to be full-sized (a scale of 1:1 with the design). The mockup should be spatially similar to the flight vehicle for initial CONOPS testing for physical activities in the mockup, with accurate dimensions and volumes of the space and equipment. If the volume includes multiple areas, these areas would ideally be contiguous and connected as in the flight model, but due to 1g or size considerations it is acceptable to have these spaces separately represented. ‘Part task mockups’ are often included in addition to a primary mockup in order to support specific functions (particularly when those functions are performed in an orientation that are not possible in 1g in the normal location within the design). Elements of the design that are also represented in a part task mockup may be included in the primary mockup as representative printouts, stickers, labels, or only volumetric models, as the stand-alone units allow for greater interaction. NASA may also use low-medium fidelity suit prototypes for integration testing in the mockups. Mockups should be able to accommodate the load of crew in prototype suits, as well as 2 test conductors, who will be managing the test and collecting data. The number of crew included in a given evaluation within the mockup will be defined based on mission CONOPS and vehicle requirements.

A low-fidelity mockup does not need to have flight like or flight similar appearances (finishes) on the inside or the outside. However, for evaluations of viewing or lighting, the color or material of finishes may be a factor. Materials in a low-fi mockup can range from foam core and wood for purely volumetric representations, to plastic and metal for load bearing components, depending on needs (such as handling the weight of a person), allowing the integrated HSI team and flight crew to weigh in on early concept designs. However, areas of focus such as crew interfaces should match closely to the flight design or current design under consideration to allow for evaluation and testing.

Greater fidelity elements can be included over time (e.g., replacing foam core or rough 3D printed models of display units or hand controllers with initially non-functional but representative units). A mockup plan will be used for the evaluation of the design evolution and specify the increase in fidelity as the design evolves. It is acceptable not to have the final design completed right away, as design changes are expected through the design lifecycle, which is one reason why a spectrum of mockup fidelity is implemented across the lifecycle. To this end, a low-fidelity mockup should be readily reconfigurable as the design changes or to accommodate feedback to try different techniques for an operation.

An initial low-fidelity mockup should minimally meet the following criteria and contain the following:

- Overall inner volume
- Location and function of crew interfaces
- Location, printed images, or volumetric models (e.g., computer-human interfaces, displays, buttons, hand controllers, stowage).
- Window locations represented by labels, stickers, or printed foam core
- Location and general motion of hatches (foam core is sufficient for a low-fidelity mockup, but with a plastic hinge or something that allows for examination of its intrusion into the crew occupied volume)
- Location and volumetric representations of crew accommodations (e.g., sleeping, waste management, food/water, etc.)
- Location and representations of seats and any other structures that impact interior volume
- Location and representations of any crew restraints (e.g. seat restraints)
- Volumetric representations of any other interface the crew is expected to manipulate (valves, switches, etc.)
- Ingress/egress aids (e.g. ladders or ascending/descending hardware) and interior mobility aids

Figure 3. Example of a low-fi mockup of the Orion capsule being used to evaluate design eye-point, viewability, and reach related to seat design, location, display and control panel layout, and launch-entry-ascent suit design impacts.

B. Medium Fidelity Mockups

The purpose of a medium-fidelity (med-fi) mockup is to allow for continued testing and evaluation of the design as it has matured beyond the usefulness of a low-fidelity mockup. These continued evaluations may include review of CONOPS for the physical use of the spacecraft and functional interactions with all human interface components. Factors to be assessed may include visibility, legibility, physical aspects of workload, usability, reach, range of motion, net habitable volume, anthropometric accommodation, placement and orientation of components, stowage, interfaces, IVA activities, habitability in general, ingress/egress, and emergency response operations.

A medium-fidelity mockup is expected to be full-sized (a scale model of 1:1), with the focus being on the interior. It is typically not necessary to have a mockup of the entire exterior of the vehicle, unless extensive EVAs on the exterior are planned as part of the CONOPS (as is the case with the ISS). Exterior areas of particular interest for evaluation (e.g., exterior ingress/egress systems, hatch operations and contingency scenarios) can be separate mockups as ‘part task mockups’ and included in evaluation in parallel with a primary medium-fidelity mockup.

The mockup fidelity should evolve as the design progresses. It should be expected to have something that has evolved at later design stages from what was provided at earlier design stages. While it is often advantageous while within a particular portion of the NASA engineering lifecycle for the maturation and upgrading of lower fidelity mockups, at critical junctures of the lifecycle (after key design reviews such as PDR or CDR), NASA strongly prefers production of a full new mockup of greater fidelity. This may include, for example, maturing a low-fidelity mockup to a certain point, but then producing a new medium-fidelity mockup, in order to surpass the limitations of the lower-fidelity construct, such as lower grade building materials and lower degrees of representativeness. This also allows for a low-fidelity mockup to continue to be used while the separate medium-fidelity mockup is being built (thus the organization always has a mockup available for testing, rather than having one completely out of commission while being upgraded). Fidelity of separate components or areas of the mockup may be tailored to support specific functions[11]. It is also beneficial for mockups to have the flexibility to be relocated for integration with other mockups or training environments, as sometimes a given program may have multiple systems being developed by different contractors, requiring cross-mockup integration for integrated testing.

A medium-fidelity mockup should minimally meet all of the requirements of the low-fidelity mockup, while superseding it with the following criteria:

- Fidelity should be sufficient to provide equipment interfaces and layout, with some functional interfaces existing, depending on testing needs.

- Include cabin lights of approximate size in the intended locations for assessments of lighting performance and glare issues.
- Lights should have similar performance to intended solution, allowing for design feedback on light positions relative to crew eye-points, displays and windows.
- Use of egress/ingress equipment/hardware in an EVA training facility where the crew may practice suited ops.
- Any ingress/egress hardware or ascent and descent hardware needs to accommodate suited personnel.

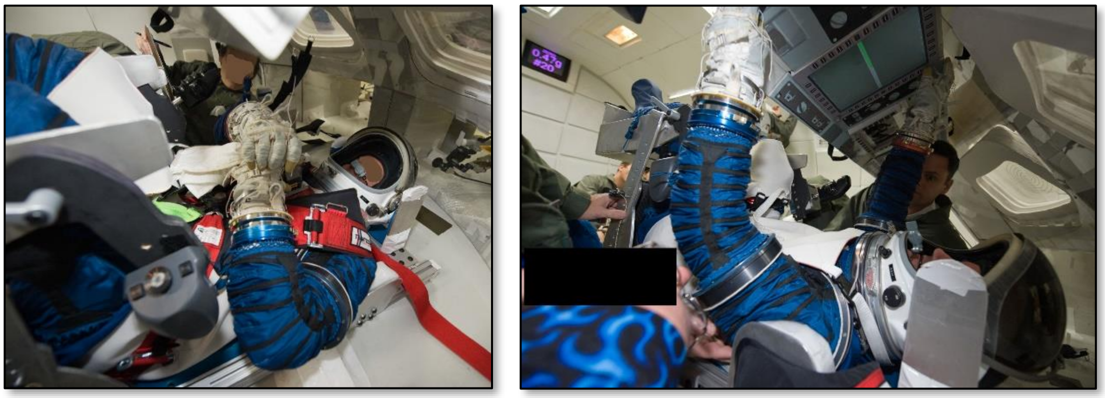

Figure 4. Example of a med-fi mockup designed for use in a microgravity parabolic flight to evaluate suit design, seat design, window field of view, D&C panel viewability and reach, as well as seat ingress and egress under microgravity conditions. Mockup was flown on three flights of roughly 60 parabolas each, with two crew exploring multiple operations associated with the seats, windows, and D&C panel.

C. High Fidelity Mockups

The purpose of a high-fidelity (hi-fi) mockup is to allow for testing and training of all aspects of the design, including hardware, software, and all human interaction components. A high-fidelity mockup is expected to provide a full-sized crew compartment (at a scale of 1:1) and be fully flight representative. The mockup should have a fidelity that is representative in form, fit, and function of the vehicle or interface simulated in order to support procedure development and verification, as well as training objectives[12].

A high-fidelity mockup should minimally meet all of the requirements of the medium-fidelity mockup, while superseding it with the following criteria:

- Engineering development units (EDU) for all components, at a minimum, though Flight equivalent units (FEU) are preferred
- High-fidelity software integration with the displays and controls
- Engineering unit lights installed in intended locations with flight like controls to allow for final evaluation of task lighting
- Structures and mechanisms (e.g., hatches)
- Habitation systems (e.g., sleep, food handling, waste management)
- Vehicle Interface to Suit Equipment (VISE) systems, for use in evaluation in suit donning and doffing as well as cabin preparation
- Emergency response hardware, software, and support for related training
- Ability to interface with other training simulators as required (e.g., integrated with Mission Control Center for mission simulation and training)

Figure 5. Examples of a hi-fi mockup from the Orion capsule at the NASA JSC Space Vehicle Mockup Facility (SVMF), depicting its use during a crew survivability HITL test.

D. Simulation and VR HITL Test Fidelity

In addition to physical mockups, software-based simulations, integrated hardware and software simulators, as well as virtual reality (VR) simulations are also used extensively during the systems engineering lifecycle. As with the purely physical mockups, flight simulators also come in varying levels of fidelity (low, medium, and high), with varying levels of both physical and software fidelity, and require an integration of software and hardware systems. These simulators make it possible for the crew involved in a HITL to respond in closed-loop scenarios with the habitat surroundings. The fidelity of the simulations can vary based on test objectives, and are generally expected to increase in maturity and fidelity as the engineering lifecycle progresses[13].

VR simulations, a more recent addition to the mockup toolset, can provide a method of evaluating an environment, system, or habitat in full 3D without producing physical hardware. These simulations can generally be developed and iterated much faster and with less expense than their physical counterparts[14]. The ability to share models among projects also provides cost-savings. Although there have been significant improvements in VR hardware and software capabilities in recent years, digitally immersive technologies vary greatly in their level of realism and capabilities[14]. Some aspects of human-system interactions are hard to manage or accurately represent in a VR environment, such as the test subjects field of view, how their perception of their physical form occupies the available net habitable volume, or physical workload or strength requirements to operate a system or a component. Hand interactions can be very limited and feel unrealistic or unintuitive (e.g., actuating switches), and judgment of distance within a volume can also be inaccurate.

For these reasons, the best use-cases for VR are currently considered to include quick assessments, early low-fidelity assessments of the layout of a volume, or minor tweaks to design that do not justify the cost of a new or modified mockup test. VR is thus considered valuable early in the design and development stages of a program, but is not currently deemed an acceptable method for HITL verification activities on its own. However, VR and CAD modeling in general may be a part of a mixed-model verification using a real-world data combined with human and system CAD models, for example in the assessment of anthropometric analyses.

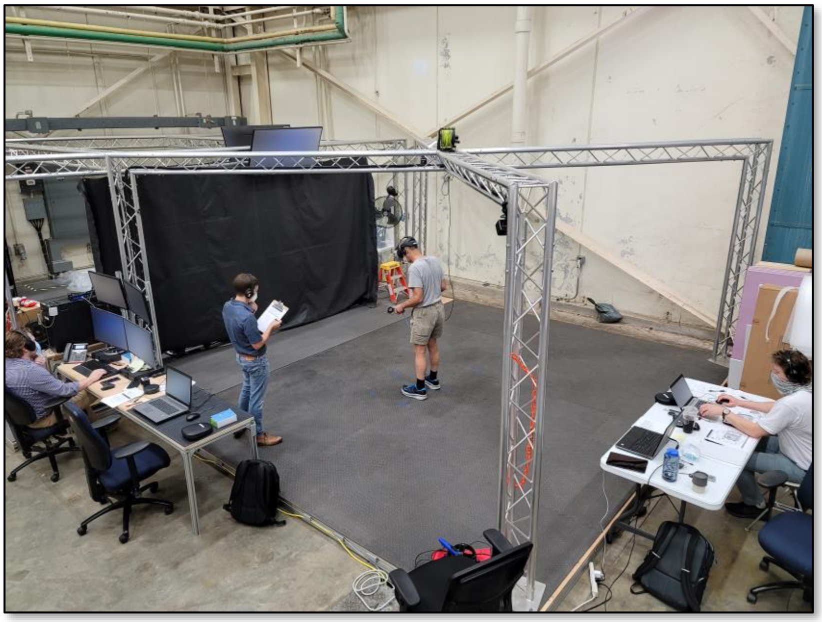

Figure 6. Example of Gateway Program VR HITL evaluations conducted at the Center for Design and Space Architecture Visualization Lab at NASA Johnson Space Center.

V. HITL Test Article Certification

HITL evaluation is a valuable tool for the design, development and verification activities that are core portions of the NASA systems engineering lifecycle. For HITL testing to be done, concepts, prototypes, and mockups are needed. The fidelity of these test articles varies, based on the stage of the lifecycle that the testing is being done for. Early in design and development, low-fidelity is usually the norm and quite acceptable as trades are being evaluated and different options considered. However, as the design matures, and design and development begin to move past the early program milestones and nears determination of whether a given design is likely to meet system requirements, the need for greater fidelity increases.

Once the need for greater fidelity approaches the higher end of the fidelity spectrum, the needs to assess the accuracy of the test articles increases, to the point that when testing is done for verification, it is of paramount importance that the test article's accuracy as compared with the official design is well documented[12]. This assessment of accuracy involves measurement of the test article along parameters that have been identified as important for a given test. These parameters will often involve physical dimensions, environmental considerations (e.g. lighting), material considerations (e.g. materials representative of the flight design), and more.

The following certification process shows what is used at NASA to validate and certify test articles (i.e., mockups) for use in HITL testing, based upon the identified parameters of interest[12]. This is not to be confused with contractual certification of a test article by a builder of the article (ensuring it meets production contract requirements) and formal acceptance of the deliverable by a federal contracting officer or the contracting officer technical representative. That is a separate topic well documented through Federal Acquisition Regulations (FAR) for acceptance of formal contract deliverables. Instead, though the validation and certification process described here is certainly relevant when a test article is delivered post-production, it also applies whenever individual components or systems are replaced with upgraded versions, or modifications are made to the article. This process is focused on the technical impact of any deviations on HITL testing.

Validation and certification are done to ensure that the article adequately represents the flight design prior to a particular HITL test, and to document any deviations from the official design, so that the deviations may be corrected or taken into consideration during results interpretation post-test. These deviations should be listed as caveats both in the test plan and in the final test report and associated presentations. This is highly significant for verification activities, as significant deviations from the flight design may invalidate results for verification closure.

A. Certification Process Overview

The certification and validation process is composed of several steps[15], including:

1) Determination of the official design/CAD model version that is being used as the baseline
2) Assessment of relevant system requirements for the test (e.g., anthropometric accommodation, reach, range of motion, net habitable volume, lighting, workload, usability, error rates, etc.)
   a) A table including relevant requirements, parameters of interest from those requirements, and how each relates to the test at hand
   b) The criteria for acceptance of the test article for the test of interest. What deviations are acceptable but need to be noted, versus the threshold for when update/modification of the test article is necessary
   c) Identification of how deviations from the design may impact interpretation of results
3) Measurement of the test article's accuracy related to the current flight design for each parameter of interest (e.g., offsets of a given test article feature/landmark in x, y, and z coordinates and related units of measure)
4) Comparison of the collected accuracy measures against the pre-determined critical thresholds, and suggesting of a course of action (acceptance of the deviation as something to be addressed in interpretation of results, or a deviation that requires correction pre-test). This will likely require that the teams conducting the HITL and those responsible for the test article coordinate together with program management to determine which deviations need to be mitigated to provide sufficient fidelity for testing, justifying any cost considerations.

Each step of this process should be presented in the relevant test plans for a given HITL, and completed prior to the test being performed. How any deviations are accommodated should be provided as part of the final test plan, and the outcomes of the validation should be included in the final report and presentations associated with the HITL.

B. Assessment and Measurement Methods

As noted in the process overview, some requirements parameters are associated with measures, while others are more of a pass/fail criterion. As an example for measurement methods, if the test article is being assessed for anthropometric accommodation or related factors such as reach and range of motion, then the XYZ coordinates of various landmarks associated with the test article (e.g., location of edges, D&C console, seat bolsters, buttons and switches, etc.) should be measured, and done so with a level of accuracy sufficient to certify/validate the test article for the HITL test of interest. Measures such as this may be collected physically using measurement instruments (e.g., meter sticks, micrometers, calipers, etc.) or using 3D scanning technology. The fidelity needs of the test should be used when determining the required methodology.

As an example of pass/fail criteria, take for instance display unit resolution: is the resolution of the test article's display unit the same as that of the flight design? If the test article's display unit (DU) is lower resolution than the design model, it may not be sufficient to adequately assess legibility of onscreen text. If the test articles DU is higher resolution than the design model, then it may provide more legibility/discriminability than the design model. In early design and development this may not be much of an issue, however it would not be acceptable for verification activities, and would invalidate the results of a test.

C. Matrix Template

The following matrix is an example template of how the assessment matrix should be composed. The needs on the left-hand column would be dictated by the CONOPS for the tasks being assessed, relevant evaluations to be performed, and the specific elements included in the design. Note that "relevant requirements" should include as many columns as needed, and should represent all of the requirements for which the design will be assessed in the HITL. The inclusion of workload, usability, and legibility in the matrix below is just provided as an example of some potential requirements that would frequently be assessed in an integrated HITL evaluation.

Table 2. Matrix Template

	Relevant Requirements			Test Article Assessment	Forward Work for Resolution
	Usability	Workload	Legibility		
Facility Fidelity Needs	*high level needs for meeting requirements*			*compliance assessment*	*resolution notes*
Display Unit	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Cursor Control Device	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Rotational Hand Controller	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Translational Hand Controller	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Switch Integration Panel	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Cabin Lighting	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Communications And Sound	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Window Views	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Seats	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Subject Orientation	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Data Collection - Video	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Data Archival	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*
Software	*relevant criteria for this component for each requirement*			*compliance assessment*	*resolution notes*

VI. Conclusion

The design and development of large complex systems includes many systems engineering processes and several methods for coordinating the efforts of a sizable number of designers, engineers, programmers, and subject matter experts from numerous fields. One of the key aspects of the engineering lifecycle for these complex systems is the need for Human Systems Integration (HSI) for ensuring that human considerations (both capabilities and limitations) are on par with hardware and software considerations, to ensure a safe, effective, and efficient system design. A key element of HSI and the systems engineering processes is the use of human-in-the-loop evaluations to facilitate design concept assessment, maturation of the design over time, and eventually verification of system requirements. Because of their importance, HITL evaluations are approached at NASA with a level of rigor and process management as befits their key involvement in the assessment of human-space-flight systems. This publication serves to document the overall structure of how HITL activities are managed at NASA, how mockups and test articles are implemented in these activities, and how they are certified for use in HITL evaluations.

Acknowledgments

Thank you to the Human Health and Performance Directorate, Flight Operations Directorate, and representatives from various Programs and Subject Matter Experts who have contributed to the development of this HITL evaluations process, providing lean processes while maintaining the needed safety regulations that these evaluations deserve as they shape the success of human spaceflight missions.

References

[1] NASA (2021). NASA/SP-20210010952 Human Systems Integration Handbook HSI Handbook. Retrieved from https://ntrs.nasa.gov/api/citations/20210010952/downloads/HSI%20Handbook%20v2.0%20092121_FINAL%20COPY.pdf

[2] NASA (2014). NASA/TP-2014-218556 Human Integration Design Processes (HIDP). Retrieved from https://www.nasa.gov/sites/default/files/atoms/files/human_integration_design_processes.pdf

[3] Acemyan, C.Z., and Adolf, J. (2018). Low-, Medium-, or High-fidelity? Defining Level of Fidelity for the Design and Evaluation of Environments. Internal NASA Report: NASA Johnson Space Center: Houston, TX.

[4] Stuster, J. (2006). The human factors and ergonomics society: Stories from the first 50 years. The Society.

[5] NASA (2021). NASA-STD-3001 NASA Space Flight Human System Standard Volume 2: Human Factors, Habitability, and Environmental Health. Retrieved from https://standards.nasa.gov/standard/NASA/NASA-STD-3001-VOL-2?check_logged_in=1

[6] NASA (2015). MPCV 70024 Orion Multi-Purpose Crew Vehicle Program: Human Systems Integration Requirements (HSIR), 4 March 2015.

[7] NASA (2015). CCT-REQ-1130 ISS Crew Transportation Certification and Services Requirements Document, Revision D1, 23 March 2015.

[8] NASA (2019). DSG-RQMT-001 Gateway System Requirements.27 June, 2019.

[9] NASA (2021). GP 10149 Gateway Human-in-the-Loop (HITL) Evaluations Process. 1 October 2021.

[10] NASA (2022). Institutional Review Board. Retrieved from https://irb.nasa.gov/

[11] Silva-Martinez, J., Othon, W., Humble G., Bellant, C., Litaker, H., Newton, C., Rich, T. (2018). NextSTEP Phase-2 Ground Test Overview and Flight Operations Support**.** 69th International Astronautical Congress 2018, IAC-18-E5.1.13x47254, Bremen, Germany.

[12] Gernhardt, M. (2018). Development of a Ground Test & Analysis Protocol for NASA’s Next STEP Phase 2 Habitation Concepts. Retrieved from https://ntrs.nasa.gov/api/citations/20180005479/downloads/20180005479.pdf

[13] Holden, K., Boyer, J., Ezer, N., Holubec, K., Sándor, A., Stephens, J. (2012). Human Factors in Space Vehicle Design. Acta Astronautica, doi:http://dx.doi.org/10.1016/j.actaastro.2012.10.020.

[14] Banerjee, N.T., Baughman, A.J., Lin, SY. et al. Development of alternative reality environments for spacecraft habitat design evaluation. Virtual Reality 25, 399–408 (2021). https://doi.org/10.1007/s10055-020-00462-6.

[15] NASA (2021). CWG-ORION-MOCKUP-CERT Orion Mockups Certification for Verification Human in the Loop (HITL) Testing. April 2021, Revision B.

Photos used in this paper are in the public domain.