Validating Human Performance Models of the Future Orion Crew Exploration Vehicle

NASA's Orion Crew Exploration Vehicle (CEV) will provide transportation for crew and cargo to and from destinations in support of the Constellation Architecture Design Reference Missions. Discrete Event Simulation (DES) is one of the design methods NASA employs for crew performance of the CEV. During the early development of the CEV, NASA and its prime Orion contractor Lockheed Martin (LM) strived to seek an effective low-cost method for developing and validating human performance DES models. This paper focuses on the method developed while creating a DES model for the CEV Rendezvous, Proximity Operations, and Docking (RPOD) task to the International Space Station. Our approach to validation was to attack the problem from several fronts. First, we began the development of the model early in the CEV design stage. Second, we adhered strictly to M&S development standards. Third, we involved the stakeholders, NASA astronauts, subject matter experts, and NASA's modeling and simulation development community throughout. Fourth, we applied standard and easy-to-conduct methods to ensure the model's accuracy. Lastly, we reviewed the data from an earlier human-in-the-loop RPOD simulation that had different objectives, which provided us an additional means to estimate the model's confidence level. The results revealed that a majority of the DES model was a reasonable representation of the current CEV design.

INTRODUCTION

NASA's Orion Crew Exploration Vehicle (CEV) will provide transportation for crew and cargo to and from destinations in support of the Constellation Architecture Design Reference Missions. Crew performance is one of the major design factors of the CEV. Among the design methods NASA employs for crew performance effort are human-in-the-loop (HITL) evaluations and human Modeling and Simulations (M&S).

One of the human performance M&S methods used is Discrete Event Simulation (DES). DES can enhance design confidence by augmenting HITL design methods such as mockups and computer simulations. Models can be used to identify where there are operator-driven constraints in Orion system design and to evaluate the capability of the flight crew to effectively operate and accomplish the mission under expected situational and environmental stressors. DES can also address design issues that are more costly and time consuming to carry out with mockups and HITL simulations by generating workload profiles as well as providing partial and end-to-end response/reaction times, insight into high workload drivers, more efficient what-if simulation for workload analyses, and the basis for requirements verification. This process can also potentially expedite the requirements verification process by reducing the need for HITL verification. Finally, it can validate requirements for scenarios that cannot be safely or effectively validated under HITL conditions.

Despite the many advantages of employing crew performance DES models as a design tool, its model validation process can be complex and expensive. Consequently, during the early development of the CEV, NASA and its prime Orion contractor Lockheed Martin (LM) strived to seek an effective low-cost method for developing and validating human performance models. This paper focuses on the initial efforts in this development process.

MODEL DEVELOPMENT APPROACH

We immediately recognized early in the CEV design stage that early development of the DES models would allow us to iron out a model development process inexpensively and efficiently due to the lower complexity of the early vehicle design. Furthermore, having the model developed early in the design cycle enabled it to grow in maturity with the vehicle design. Our approach was to identify a vehicle design element that the human factors engineers were focusing on at the time that could substantially benefit from a DES model.

We ended up developing a DES model for the CEV Rendezvous, Proximity Operations, and Docking (RPOD) task. This model was the first in a series of models. For the RPOD task, the crew of the CEV has to successfully dock the CEV with the International Space Station (ISS). In the model, the crew controls the translation maneuvers manually while allowing the attitude to be controlled via auto-pilot. Besides using it for the CEV design, the model was used to develop a viable low-cost development process. This was done by tying the model's maturity level to the CEV Design Analysis Cycles (DAC).

MODEL DESCRIPTION

The RPOD model consists of two major components: a Master Task List (MTL) as the input and the human performance DES model itself. Both the MTL and human performance DES models are being developed based on continual inputs from subject matter experts (SMEs), archival data, space flight documentation, training materials, human performance

research literature, and standards such as the NASA-STD-3000 (1995). The model was iteratively updated as details about the CEV designs and functionalities became available. Throughout the process, the Orion requirements documents were used as a guide.

The MTL's data came from carrying out several task analyses. Although this was a time-consuming and costly process, its benefits far outweigh carrying out HITL analyses alone. The MTL was organized by mission phases, major functions within phases, and high-level tasks within functions. Data types included the functions and tasks, the human operator designations, the task flows, the decision logic associated with operator decision nodes, and the descriptive parameters associated with each of these model components. Data sources came from existing task lists, network logics and task allocations provided by the respective CEV mission area teams, task attributes provided by SMEs and research literature, as well as crew, environmental, displays/controls and other subsystem information.

The human performance model component was constructed using the Improved Performance Research Integration Tool (IMPRINT Pro v.3.0). IMPRINT is a dynamic, stochastic, discrete event modeling tool designed and maintained by the U.S. Army Research Laboratory's Human Research & Engineering Directorate to help assess the interaction of operator and system performance throughout the system lifecycle. Data derived from task analyses are used to lay out task steps in the form of a task network. Then task decision sequences and logics are added to tasks where there are alternatives for which task(s) may come next in the sequence. Each crew task in the model is described by multiple parameters: the operator(s) performing the task, the mean time and standard deviation required to perform the task, the appropriate distribution curve for task duration, equipment utilized to perform the task, the consequences of failing to perform the task, and the workload values associated with task performance. The workload values used are based on a modified Visual, Auditory, Cognitive, and Psychomotor (VACP) scale (McCracken and Aldrich, 1984; Szabo and Bierbaum, 1986). Model outputs include both workload graphs and performance indicators such as task completion rates and times for completion that reflect the degree to which crew members will be able to perform their assigned tasks in both nominal and off-nominal circumstances.

MODEL VALIDATION

Model validation was the most important portion of this activity. We adhered our validation method strictly to existing established M&S standards (US DOD, 2001; NASA, 2008). First, we prepared an Analysis, Validation, and Verification plan. The plan outlines the validation, verification, and accreditation activities that will eventually lead to establishing the model's accuracy and credibility. We also involved both the NASA and LM Orion M&S community by working directly with the Orion Modeling and Simulation Working Group, a NASA and Orion contractor community that oversees all the M&S developments for the Orion CEV, to ensure the correct procedures outlined in the aforementioned M&S standards were followed.

Orion CEV stakeholders, current NASA astronauts, and SMEs from NASA, LM, and its subcontractors were involved throughout the process in generating and validating the MTL data and inspecting model. To control cost while maintaining quality of the model, we employed a desktop evaluation approach through face, content, and output validation (Naylor and Finger, 1967; Law and Kelton, 1991) as well as low-fidelity HITL simulations.

Face/Content/Output Validity

For face validity, the SMEs reviewed the network flow diagram of the model to determine how well it represented the tasks the crew would perform in the CEV as it was understood. For content validity, the SMEs reviewed the performance times, error rates, and workload values for each of these tasks.

The SMEs determined the output validity by reviewing the output of the model (e.g., the workload charts, number of task failures, etc.). We then documented the feedback from the SMEs and modified our assumptions, input data, task flows, and model as necessary. Once the model was updated, it was re-executed and the new results were re-analyzed.

As part of the output validation activity, a sensitivity analysis was performed by systematically changing the values of model input variables and parameters over some range of interest and observing the effect upon model behavior (Shannon, 1975; Hermann, 1967; Miller, 1974a; Miller, 1974b; Van Horn, 1971). Based on suggestions from SMEs that some operator tasks may take two or even three times longer to perform in zero gravity (0-g) than in 1-g, an RPOD model sensitivity analysis was executed to evaluate these possibilities. In order to address this question, mean task performance times for all relevant operator tasks (those tasks that contained a physical element) were first doubled for one series of runs, and then tripled for another set of runs to emulate the possible prolonged task durations in 0-g.

HITL Handling Qualities Simulation

In terms of validation with HITL simulation, we coordinated with another CEV team that earlier conducted a separate low-fidelity HITL RPOD simulation with a CEV vehicle dynamics model to address handling quality issues. Unfortunately, due to the timing of the two events, we were unable to synchronize our human performance model validation with this HITL

simulation. Nonetheless, we were able to review a summary of the HITL results to help us determine the model's confidence level.

The goal of the HITL RPOD simulation study was to determine if the current vehicle design met spacecraft handling quality requirements. A medium-fidelity three-degree-of-freedom CEV simulation on a low-fidelity mockup was the simulation platform of choice. Two display units were provided to the participants; one displayed the docking camera view (including range and range rate) and the other displayed the relative state of the vehicle. Participants only controlled the translation of the vehicle with a Translational Hand Controller. After each scenario, handling quality using the Cooper-Harper Handling Qualities Rating Scale (1969) and workload ratings using NASA Task Load Index (TLX) were collected from the participants. They were then followed by a final debriefing and a survey completion period. Other data collected during the course of the evaluation included videos and still photographs as well as objective data on participant control inputs and vehicle flight dynamics response captured by the simulator.

We reviewed the workload ratings from the handling quality assessment and the videos collected during the evaluation..

MODEL Validation Results

Face/Content/Output Validity

Peer review. Three current NASA astronauts reviewed the face, content, and output validity of the model. Their review was based on their experience and current knowledge of the Orion systems design. They found the inputs, outputs, and the model itself acceptable and consistent with the latest version of the CEV RPOD Concept of Operations. The majority of their comments focused on our assumptions, task names and task descriptions. There were no substantial changes to the model based on these comments.

The SMEs provided us valuable comments that allowed us to fine tune the validation process to enable greater confidence on model. There were two tasks that the review team had concerns with initially. The first had to do with the occurrence of manual flight control. We originally had this task occurring once at a specific time in the model. However, we received feedback that this could occur at any time between the Terminal Phase Initiation maneuver and docking. The second concern had to do with the Transfer-to-Docking Axis maneuver. We originally had this task initiated by the crew. Feedback from the peer review team stated that this would be an automated task. Both suggestions were documented and incorporated into the model.

Mission time. In order to generate a meaningful distribution of results, the baseline model was executed 354 times. This number was taken from a power analysis table by Keppel (1991). The minimum time for performing RPOD Operations was 01: 46:32; the maximum time was 01:47:12; while the average time was 01:46:49. (Task times are in the format HH:MM:SS.) These results imply that when the crew performs the tasks required for a nominal mission, there is not much variability in the amount of time required to perform the various tasks that make up this function. These results were reasonable because there are relatively few decisions and physical actions that the crewmembers must make during RPOD and the majority of the actions they perform require little time. Much of the RPOD phase duration is controlled by the vehicle's orbital dynamics and the selected concept of operations, which involves a slow approach to the ISS.

Mission success. All tasks modeled were completed without failure and the mission was never aborted during the 354 executions of the model. This result was again reasonable due to the low probability of failure for the human tasks, assuming proficient operators and under the assumed environmental conditions. For example, the lowest mean accuracy for a task in the model was 99.94%.

Workload. Figure 1 shows the overall average workload for the one crewmember predicted by the model. The highest period of workload is shown to be when the crewmember is manually controlling the CEV/ISS contact conditions while simultaneously monitoring rates, sensors, etc. To further understand the workload during the peak workload period, we broke down the overall workload during that period further into their respective VACP components (Figure 2). Figure 2 shows that the cognitive, visual, and fine motor channels have the highest workload while there is little workload in the speech and auditory channel. This assessment concluded that the model behaved reasonably in terms of workload.

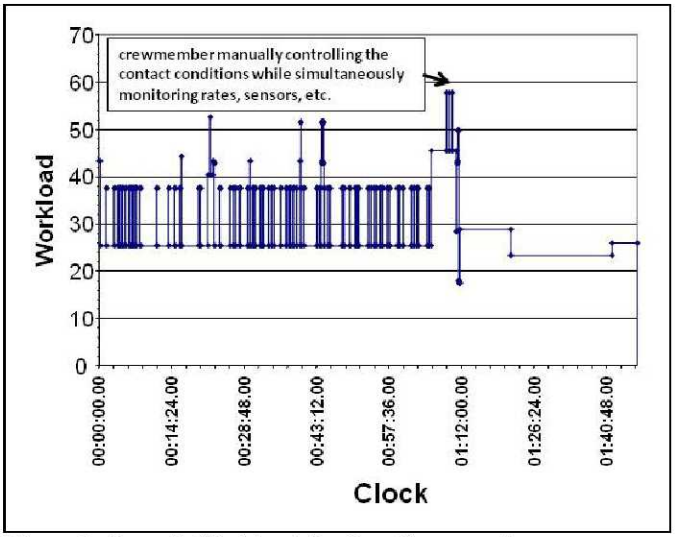

Figure 1. Overall Workload for One Crewmember

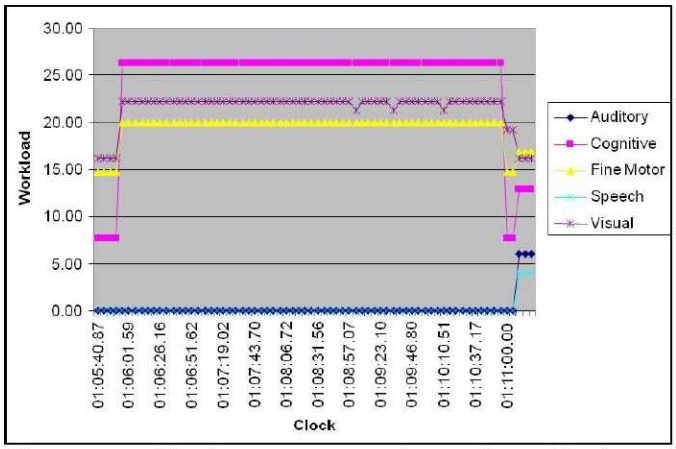

Figure 2. Workload Components During Peak Workload Period

Sensitivity. The sensitivity analysis results are summarized in Table 1. The RPOD sensitivity analysis, which doubled and tripled physical operator task times, predicted the RPOD completion at 02:22:04 and 02:57:18 respectively. (Compared to the nominal model, which predicted an RPOD mission time of 01:46:49.) Since the mission completion performance requirement for RPOD activities was arbitrarily defined to be 2 hours in all scenarios, the models with doubled and tripled task times thus did not meet the requirement. As with the overall mission, we also defined arbitrary performance requirements for each of the individual tasks in the RPOD model. The percentage of model runs that a given task did meet the requirements dropped from 100% in the nominal model, to 70% in the doubled task time model, and to 40% in the tripled task time model. We also noticed that both the predicted total mission performance times and the percentage change in meeting the requirements were linear. These results made sense and were what we expected.

Table 1. Mission-level Results from the Sensitivity Analysis

	RPOD Model with Nominal Operator Task Times	RPOD Model with Doubled Operator Task Times	RPOD Model with Tripled Operator Task Times
Average Mission Time	01:46:49	02:22:04	02:57:18
Minimum Mission Time	01:46:32	02:21:48	02:57:04
Maximum Mission Time	01:47:12	02:22:21	02:57:35
Mission Time Std Dev	7.29	7.17	6.46
# Mission Failures	0	354	354

HITL Handling Qualities Simulation

Because workload measures within our IMPRINT RPOD model are different from the TLX used in the HITL handling qualities simulation, it was difficult to conduct a direct comparison of the results. Instead, we decided to observe the videos from one participant to look for clues regarding the task/time relationship. In particular, we paid attention to the total scenario/run time, the amount of time the participant spent using the hand controller, and the number of times the participant selected an option on the display.

For the IMPRINT RPOD model, while the simulated crewmember is manually controlling the vehicle, he/she is simultaneously monitoring navigation and propulsion systems, sensor data, computed and propagated trajectories, etc. At times, the model simulates the crewmember configuring tools, sensors, etc., via a display. In other words, the crewmember could be manually controlling the vehicle and configuring the sensors and displays at the same time. This was however not what we observed in the videos. The discrepancy could imply that the RPOD model might need additional tweaking. Nonetheless, due to the less-than-ideal quality of the HITL data, we decided that using the HITL data to provide a rough estimate to the uncertainty of the RPOD model was a better approach.

MODEL CONFIDENCE LEVEL ANALYSIS

At the end of this validation cycle, we estimated the IMPRINT RPOD model confidence. It is important to note that we did not aim at creating a highly accurate model but instead a model that matches the then current maturity level of the CEV designs. To that end, we computed the Root Mean Square (RMS) error by estimating the errors from both the face/content/output validation and the HITL handling qualities simulation. For the HITL simulation, we determined that there were two factors we needed to consider. These factors had to do with the confidence level of the CEV vehicle model used in the simulation and the quality of the data available from the simulation. The estimations are summarized in Table 3.

Table 3. Model Confidence Level Estimation

Factors	Estimated Error (%)
Accuracy of the current model based on face and content validation	5
Degrees to which the current HITL simulation matches the current CEV design maturity	5
Type of data available from the HITL simulation to support the validation	10
RMS =	12

Since the type of data available from the HITL simulation was less than ideal, we estimated that the error this factor contributed to be the highest among the three factors. Nonetheless, we believe a 12 percent RMS error at this stage of the design is acceptable.

CONCLUSION

The results of this initial comparative analysis were promising. The IMPRINT RPOD model was considered to be accurate reasonable representation of the current design of the CEV with an estimated RMS uncertainty error of 12 percent. In both the TLX workload ratings from the handling quality simulation and the IMPRINT model's VACP workload output, mental workload was rated higher than the other workload categories.

We believe we have developed a workable and low-cost method to validate human performance models while we were developing our first human performance model for the CEV RPOD scenario. Our approach to validation was to attack the problem from several fronts. First, we began the development of the model early in the CEV design stage to allow the model the mature together with the CEV development. This approach allowed us to keep the cost down by reducing model complexity and therefore making the validation process easier to carry out. Second, we adhered strictly to M&S development standards created by NASA, LM, and DOD. We started by preparing an Analysis Process, Validation, and Verification plan, which eventually led to establishing the model's accuracy and credibility. Third, we involved the stakeholders, NASA astronauts, SMEs and NASA's M&S development community throughout to ensure that the model we developed was built right. Fourth, we applied standard methods to ensure the model's face/content/output validity. These methods were easily conducted. When combined with the involvement of SMEs, they provide high level of confidence to the quality of the model. Lastly, we believe the use of the data from a recent HITL simulation of the same task, though with different objectives, enabled us to get a reasonable estimation of the model's uncertainty error. The CEV dynamics model used in the HITL study have been validated for use. With the participation of subjects, the study provided a slightly more realistic look at the human performance during the task. Unfortunately, due to the lack of coordination with this separate study, we were only able to use the video data for the validation. Nonetheless, the data was found to be valuable and provided us a means to estimate the model's error.

We believe we have developed an acceptable model for this stage of the CEV designs. In the future, we can future improve the model quality by proactively engaging related human-in-the-loop computer simulation activities.

REFERENCES

Hermann, C. F. (1967). Validation problems in games and simulations with special reference to models of international politics. *Behavioral Science, 12*(3).

Improved Performance Research Integration Tool (IMPRINT). (2009). [Computer software]. Available from the U.S. Army Research Laboratory web site, http://www.arl.army.mil/www/default.cfm?Action=445

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed., pp. 72). New Jersey: Prentice Hall.

Law, A. M., & Kelton, W. D. (1991). *Simulation modeling and analysis* (2nd ed.). New York: McGraw-Hill.

McCracken, & J. H., Aldrich, T. B. (1984). *Analysis of selected LHX mission functions: Implications for operator workload and system automation goals* (Technical Note ASI 479-024-84). Fort Rucker, AL: Anacapa Sciences, Inc.

Miller, D. R. (1974a). Model validation through sensitivity analysis. *Proceedings of the 1974 Summer Computer Simulation Conference.* Fort Rucker, AL: Anacapa Sciences, Inc.

Miller, D. R. (1974b). Sensitivity analysis and validation of simulation models. *Journal of Theoretical Biology, 48*(2).

NASA. (1995). NASA Standard 3000: *Man-Systems Integration Standards*.

NASA. (2008). Technical Standard 7009: *Standard for Modeling and Simulations*.

Naylor, T. H., & Finger, J. M. (1967). Verification of computer simulation models. *Management Sci.,14*, 92-101.

Shannon, R. E., (1975). *Systems simulation: The art and science*. New Jersey: Prentice Hall.

Szabo, S. M., & Bierbaum C. R., (1986). *A comprehensive task analysis of the AH-64 mission with crew workload estimates and preliminary decision rules for developing and AH-64 workload prediction model* (Technical Report ASI678-204-86{B}, Vol I, II, III, IV). Fort Rucker, AL: Anacapa Sciences, Inc.

United States Department of Defense Modeling and Simulation Office. (August 2001). *Verification, Validation and Accreditation Recommended Practices Guide*.

Van Horn, R. L. (1971). Validation of simulation results. *Management Science, 17* (5).