Burian, B. K., Pruchnicki, S., & Christopher, B. (2011). Designing realistic full mission human-in-the-loop aviation simulation studies: Lessons learned. In *Proceedings of the Human Factors and Ergonomic Society 55th Annual Meeting*. Las Vegas, NV: HFES.

Designing Realistic, Full-Mission, Human-in-the-Loop Aviation Simulation Studies: Lessons Learned

Barbara K. Burian
NASA Ames Research Center

Shawn Pruchnicki and Bonny Christopher
San Jose State University

The design of simulator based research can be an intimidating task for even an experienced researcher. We present some of the lessons learned from a complex simulator-based research study requiring sophisticated flight scenarios with realistic environmental conditions that was designed and conducted by a large group of researchers distributed across multiple sites. We offer readers both suggestions for approach to design as well as solutions to some of the problems we encountered.

INTRODUCTION

The development and design process for human-in-the-loop (HITL) simulation research can be daunting for even the most experienced researchers. Often important steps or considerations are overlooked requiring replanning and re-designing, sometimes multiple times. Special challenges are posed when the study team is large and distributed across several geographical locations. During the lifecycle of our most recent HITL simulation study, we realized that others could benefit from both our positive and negative design experiences. Although some of our suggestions and observations also apply to the design of part-task studies (or many other types of research studies, as well) they are particularly targeted to the unique difficulties encountered during the development of full-mission aviation studies in which participants are required to complete all the tasks that would exist in the real-world operational environment being simulated. Although our examples and experiences are in the field of aviation, researchers designing HITL simulation studies in other domains, such as medicine, should also find many of our suggestions to be pertinent. For simplicity, we will refer to all simulation data collection devices as "simulators" even if technically, some are "flight training devices."

PRACTICE INNOVATION

Full-mission, HITL simulation studies have been used in aviation research for some time (e.g., Chou, Madhavan, & Funk, 1996; Mumaw, Sarter, & Wickens, 2001). Unlike statistical simulations (Burton, Altman, Royston, & Holder., 2006; Maldonado & Greenland, 1997) or modeling-based simulations (Balci, 1994; Norman & Banks, 1998), there is little guidance available regarding the design and development of HITL simulation studies. HITL studies are being used more and more frequently in aviation, particularly with regard to assessing the functionality of proposed aspects of the Next Generation air traffic management system (NextGen). Thus, we believe that researchers engaged in the conduct of such research could benefit from our experiences in designing and conducting just such a full-mission, HITL simulation study. Every HITL simulation research project is unique, as are the agencies, facilities and researchers involved. Nonetheless, the issues and concepts discussed here transcend these differences and offer some practical insight and suggestions for all.

FINDINGS

Our observations and suggestions for designing full-mission HITL studies are presented in the 10 steps below. Although the steps are presented sequentially (and steps appearing later should be accomplished later in the process) some steps describe actions that should occur throughout the entire project such as Step 1: Ensuring Effective Communication. Additionally, researchers will find that they must accomplish the activities associated with some steps several times in an iterative fashion throughout the scenario design and development lifecycle.

Before discussing our findings, some assumptions must first be made clear. We assume that first, careful consideration has been undertaken and that a full-mission, HITL simulation study is required to gather the data of interest (Norman & Banks, 1998.) As any researcher who has conducted one will attest, these studies are typically expensive and labor intensive. Their development, data collection, and data analysis processes are often lengthy and complicated (Law, 2003). Therefore, researchers should be certain that a HITL simulation study is truly necessary prior to beginning any design activities. It is also assumed that the researchers and any subject matter experts (SMEs) who may be part of the design team have adequate domain expertise to design valid and realistic scenarios and tasks. Similarly, it is assumed that all data collection devices (e.g., the simulator, eye trackers, physiological and subjective measures) will yield valid data of interest and that members of the team, including consultants or SMEs, have the expertise necessary to construct data collection apparatus needed and to analyze the data.

Step 1: Ensuring Effective Communication

Clear communication and policies to support it, should be established from the very beginning of the design process and must be a high priority throughout the life of the study. The roles and responsibilities of all members of the team should be clearly delineated as well as the process that will be followed for design, data gathering and analysis. This is especially important when a large and/or distributed team will be working in collaboration across multiple sites. Clear and constant communication ensures that the entire group maintains focus on meeting the agreed upon objectives. Many groups might find the practical implementation of this goal rather challenging.

For collaborative teams at different physical locations and possibly across different agencies, email can offer a huge advantage in achieving the goal of effective communication, in addition to regular teleconferences and web-based and in-person meetings. Despite the advantage and the amazing timeliness that email offers, a few words of caution are warranted. The amount of professional email we all receive can be overwhelming and at times almost paralyzing. Too many emails discussing little tidbits of information regarding the study can be very difficult to manage in not only capturing this information but also in later retrieval. One solution is to capture the data offered by creating a living word processing document with all like information shared on a particular topic. This can be simply a matter of cutting and pasting from emails into the document. This offers a potential solution to information tracking and later retrieval in addition to providing a historical perspective of ideas and changes.

Another communication issue for consideration is who from the team will be included in the various email based discussions. Occasionally someone is unintentionally omitted from a discussion or email exchange. The creation of distribution lists for certain email discussions on certain topics can be very helpful in preventing an inadvertent omission of relevant team members and lost time regarding problem solving. However, we recognize that there is a difficult balance between appropriate email inclusion and extraneous email for some team members.

Step 2: Document Control

Because the very nature of research design is an iterative process, documents are constantly being reviewed and revised. This poses a significant barrier to ensuring that the team has access to and is working from the most current version of a given document. Specific solutions will vary, but we found that having an internet-based location to manage document control, such as SharePoint, is an effective tool. This allowed us to provide a central point of access for all team members, especially those in different geographical locations. In an effort to ensure that the most recent version of a particular document is posted and is being utilized, it is helpful if the date of any revision is always included in the file path (document name) before posting to the central internet site (e.g., “Simulator weather conditions leg one 12 20 10.doc”). For documents experiencing frequent changes, maybe even daily, it might even be necessary to include the time of day for the revision as well (in an agreed to time zone if your team is spread out among several). Common word processing software can also be set to automatically add date and time in the header or footer when the document is opened. To eliminate the possibility that a team member might access an out-of-date document off the website, we found it helpful to move earlier versions to an archive library as soon as a new version was posted so that only the most recent version of a document was visible in the active document library. Saving earlier versions of a document in an archive, rather than just deleting them, ensures that information which has been revised is still accessible if needed at some point in the future.

When numerous team members are involved in the review and revision of a document, we suggest that a single individual be responsible for managing the review process. This individual distributes the version for review, facilitates the desired review process (i.e., review of the document occurs sequentially, passing from one team member to the next, or in parallel, by all team members at the same time), integrates edits, addresses conflicting comments or suggested changes, distributes or posts the final revised version, and informs the team of the revised document’s location and status.

These first two steps (ensuring effective communication and document control) are foundational and pertain to the day-to-day way that the research team will function. The suggestions made may seem obvious, but we have found that making the process by which the team will work together explicit and reinforcing and following it throughout the lifecycle of the study greatly reduces the likelihood of a wide variety of problems. It is extremely frustrating to find that extensive revisions have been made to a document that was not the most recent version or that, after much work on scenario design, members of the team do not all share the same understanding of the goals of the study or of a particular task.

Step 3: Early Discussions of Scenario Design and Participant Tasks

Once the rules for team process and functioning have been established and agreed to by all, it is time start considering in more detail the goals of the simulation study and the experimental design. During this step we suggest that researchers develop a relatively high level outline of the scenarios and tasks to be accomplished, degree of realism necessary, the types of participants needed, simulator capabilities (see step 4 for more information on this), and similar considerations associated with obtaining the data of interest. It is not too soon in this step to begin a conversation about data and statistical analysis techniques to be employed (Burton, et al., 2006). Other logistics that might be taken for granted that may potentially cause significant roadblocks later in the study include approaches to participant recruitment and the amount of their time required during data collection. Financial resources must also be considered since programming is frequently required, new equipment may need to purchased, and participants may be paid for their involvement.

We found it wise to include the entire team in these conversations, regardless of the role the various members might end up having (e.g., principal investigators, research associates, interns, programmers, simulator technicians, data analyst, etc.). This helps to ensure that all share a common understanding of the overall design of the study and its relationship to achieving the research goals.

Once a team begins discussion of study design, it can be easy to migrate toward in-depth design activities too quickly. In fact, activities in this step will be revisited several times throughout the scenario design process. Therefore, we caution teams that the first time or two through this step, these discussions should be at a very high level, outlining goals and objectives and general thoughts with regard to type of participants and overall tasks. Multiple steps which follow should be accomplished before detailed, in-depth scenario and task design is undertaken.

Step 4: Know Your Simulator

Aircraft and Air Traffic Control Workstation simulators are extremely complex devices that offer a tremendous ability to not only train pilots and controllers but answer both simple and difficult research questions. This complexity can offer both design flexibility as well as obstructions. Additionally, simulators which have been constructed for use in training may need alterations or additional programming to increase their utility as research tools and ensure that needed data can be captured.

It is essential that researchers planning to conduct HITL simulation studies know everything possible about the simulator that will be used for data collection. For example, what are the simulator's capabilities and limitations? What type(s) of aircraft or workstation can it emulate and with what degree of realism? What type of additional components will need to be constructed or programmed to assess the variables of interest? There are a wide range of other aspects of simulator functionality that will need to be considered which have direct relevance for study design, research questions that can be answered, and data that can be gathered. For example, some of the considerations with regard specifically to aircraft simulators include, but are by no means limited to, the following:

- Constraints on how other traffic can be displayed on cockpit displays and out the window
- Types and levels of visibility and lighting conditions that can be simulated
- Types of weather and environmental conditions such as number cloud layers that can be shown, types of winds, icing, turbulence, lightning, convective activity, etc. that can be realistically simulated
- Types of terrain, geographical features and areas, airports and airport features (e.g., taxiways, buildings), urban areas and the like, that can be displayed and with what degree of realism or resolution
- Types of avionics, displays, information, and other technologies (e.g. XM Weather, datalink, etc) that are available and the constraints or possibilities that their design and functioning impose or offer
- Types of off-nominal, emergency, or abnormal situations or conditions that can be realistically simulated
- Motion vs. non-motion capabilities
- Ability to realistically emulate aircraft performance and handling characteristics related to various environmental or emergency/abnormal conditions such as in-flight icing or a blown tire during the landing roll-out
- Ability to add or integrate new technologies or displays or perform various procedures which are to be evaluated in the study.

This list of course is by no means exhaustive. With the expense of modern day simulators, we are all faced with the reality of having little to no choice in which test bed we use when conducting HITL simulation research. Access to simulation facilities and the capabilities and limitations of these devices will force researchers to craft their studies around these considerations. Obviously, detailed knowledge of the constraints imposed by the simulator to be used should be identified early in the scenario development phase as these considerations will shape the very nature of study hypotheses and objectives.

All of the steps discussed in this paper are important, but this is probably one of the most important as it has direct impact upon several considerations such as what research questions can be answered, necessary participant experience, and what kind of data collection and analysis will be possible. Everyone on the team must have a clear understanding of all simulator constraints, capabilities and limitations and be mindful to maintain an open dialogue of these issues during study design.

Step 5: Data Sources and Management

Once team members have a solid understanding of the simulator's capabilities, limitations, and functionality, they must ensure that it can capture and record the required data streams and that any technological difficulties in doing so are resolved early in the design phase. Non-simulator based data acquisition devices such as eye trackers, heart monitors, videotape recorders and others must also be evaluated to ensure that all needed information will be adequately, and possibly redundantly, captured. For example, will keystroke data and response time be captured digitally off the simulator itself or must this be gathered later from post simulation observations of video recordings? If video and audio recordings are possible and required for the study, do they possess the fidelity required for analysis? Researchers do not want to discover during data collection that video fidelity and/or sound quality is insufficient to meet their research objectives.

Once digital or analog data streams have been determined, researchers should be mindful of how they will be integrated when ready for analysis. Will all data streams be time-stamped so they can be aligned manually during analysis or can they be automatically synchronized? It is possible that these will not be important considerations and yet some researchers will find that automatic data stream synchronization is a mandatory requirement of data collection.

Step 6: What Are We Doing Again?

Now that the team has a thorough understanding of the possibilities and requirements of the simulation environment, data collection devices and how these considerations affect the structure of the study, it is time to move forward by going back to revisit Step 3: discussions regarding scenario design and participant tasks. Revision of the study plan, research objectives, variables to be assessed, and even the population of interest will often need to occur as what has been learned in steps 4 and 5 is evaluated in addition to consideration of budget, time line, and programming and equipment construction needs.

Do not lose heart. By delving into the possibilities and constraints of the simulator and other data collection measures and devices, and by considering programming and equipment needs, additional research questions and data collection possibilities may be discovered. Although not unique to simulator based research, this critique, evaluation, and re-planning is the essence of rigorous research and should be viewed as the study enhancing endeavor that it is. However, as with the first time through step 3, discussions about the simulation and tasks to be accomplished by participants should still be at a moderately high level. The time for actual detailed scripting of scenarios and tasks must wait until a few other steps are accomplished.

Step 7: A Room Full of Experts

For studies such as these, a broad mix of SMEs representing a variety of domains will typically be needed. Individuals with expertise in the development of scenarios and realistic tasks may not be the same SMEs as those who are intimately familiar with the simulator's capabilities and performance, avionics functioning and usage, and other aspects of the aviation operational environment. Similarly, other experts with knowledge in experimental methodology and research design, aviation meteorology, computer programming and the ability to assess not only participant behavior but perform other types of analyses as well (e.g., cognitive workload, situation awareness, voice and speech analysis, analysis of eye tracker and physiological data, etc.), will often be necessary members of the HITL simulation team. From a staffing and financial perspective, hopefully many on the team have the ability to fill several roles. Ensure early on that the researchers involved can fully manage the various technical aspects of the study you are about to conduct. You do not want to discover that your team is missing an area of expertise or have overestimated a member's abilities late in the design process. Honest evaluations of the subject matter expertise needed, the availability of SMEs with that expertise, and related budgetary considerations must be undertaken. It is possible that the unavailability of a particular skill set or area of expertise may affect your research objectives, variables of interest, design, or methodology.

Step 8: Scenario Design Part 1 – Defining the Operational Environment

Scenario design goes hand-in-hand with the definition of the operational environment in which the scenario will occur. In some studies, the effect that variations in the operational environment, such as how the amount of traffic on a radar display affects task completion, is the variable of interest. In other studies, the operational environment serves mostly as a backdrop to the scenario and the tasks to be accomplished by study participants. After completing earlier steps and, in particular, after gaining a solid understanding of the simulator's ability to realistically portray aspects of the operational environment, it is time to begin scenario design by defining, in detail, the operational environment.

For example, when conducting a HITL study using an aircraft simulator, by this point the team will probably have decided upon the general weather conditions desired for the scenario, geographic location and airports to be included, amount of other aircraft traffic, and time of year and time of day for the scenario flight. Now is the time to turn attention to specifying weather and other environmental conditions, both in studies when these conditions only serve as the backdrop to the scenario, as well as in studies when they are experimental variables to be manipulated.

We discovered that once armed with the general weather needs of a scenario, the best way to define specific weather conditions was to begin by taking a map of the route of flight and sketching out the general location and movement of any low or high pressure systems and frontal locations which, in turn, normally determine wind directions and speeds. Make sure that the large area weather conditions chosen will maintain realism for not only the departure and arrival weather required, but for the enroute portion as well. Consider the specific runways to be used for departures and arrivals as the wind around these areas is determined. Winds aloft for the area will also need to be programmed with realistic changes in wind direction and speed in addition to temperature changes relative to altitude. Once a "global" sense of what the weather is doing and how it relates to the flight, weather reporting stations along the route of flight need to be identified. After those are known, surface conditions, area forecast descriptions, and winds aloft related to these reporting stations can be defined. We found it helpful to obtain a real-life briefing for our scenario route of flight and then inserted the weather we had crafted for our simulator flight into this briefing. By doing this the weather briefing packet we handed our pilot subjects and air traffic controllers not only looked real but described for everyone the weather that we designed and had programmed into the simulator.

Changes to various environmental conditions will also need to be considered based on the length of time it will take participants to fly the scenario. For example, "recorded" automatic terminal information service (ATIS) identifiers (e.g., "XYZ Airport Information *Sierra*") will need to change every hour at a minimum. Similarly, reasonable changes in weather with the passage of time, as well as ambient lightning changes (e.g., dawn-to-day, dusk-to-night), may be needed. After having defined weather and other environmental conditions in great detail, these parameters must be recorded in the format needed by those who will program them into the simulator.

Step 9: Scenario Design Part 2 – Specifying the Scenario and Tasks

Now is finally the time to design, in detail, the rest of the scenario and the tasks to be accomplished by study participants. The team should have a clear understanding by now of what tasks the participants will be asked to perform, how their performance will be monitored, and the types of data collection that will be employed. There may be a wide number of other aspects of the scenario to be considered during this detailed design phase. For example, as we completed this step for our HITL study using an aircraft simulator, we had to pay attention to types of airspace to be crossed, departure, arrival and approach procedures, frequency changes across ATC sectors and areas of control, and the range of speeds with which participants might choose to fly scenario flight.

Depending on the research questions to be answered and the number of tasks to be examined, more than one scenario may need to be developed. In these cases, researchers must also keep in mind the effects of fatigue and participants' physiological needs when developing the scenarios and the overall study design.

Step 10: As Time Goes By

Designing a full-mission simulation study takes time. Researchers who are truly interested in replicating a realistic flight environment must be aware that the real world operational environment does not stop changing to give them time to finish designing the study and conduct data collection. Instrument arrivals, approaches, and departures can change as can frequencies and airspace/sector boundaries. With the introduction of more and more procedures related to NextGen, other aspects of the operational environment, required technology, and flight tasks will change as well. Many of these changes, or potential changes, will be the variables of interest that will be examined in the study. However others, such as a particular departure available at an airport, may not be. The research team must keep up with changes to the aviation operational environment and decide if any changes in the real world require changes to the simulated one.

Eventually, most researchers will likely need to choose a point in time where the charts, plates, Global Position Satellite (GPS) databases and other operational information that change over time in the real world are essentially "frozen" and are no longer updated. Most methodologically sound study designs cannot accommodate some of the participants flying one procedure and the rest flying its recently updated cousin. One departure procedure being used in our study was significantly altered a few short weeks before the start of data collection. We decided to incorporate the changed procedure since we had not yet begun data collection but it necessitated re-doing a portion of our task analysis and concurrent task maps and actually changed the workload for the pilots flying it.

DISCUSSION

Although we have presented the tasks involved in designing a HITL simulation study as a series of sequential steps, we must remind readers that some of these steps (i.e., 1 and 2) are not steps at all but are foundational to the entire process of scenario design. Furthermore, research teams will cycle through a few of these steps on multiple occasions as more information is gathered and the research objectives are sharpened.

A validation of the scenario design should, of course, be undertaken following the completion of scenario design. However, we suggest that validation runs in the simulator occur multiple times throughout the scenario design process, rather than just prior to the planned start of data collection, to ensure that the scenario works as intended and to allow time for any needed changes to programming, scenario design, data collection procedures or even possibly changes to research objectives.

Scenario design for HITL simulation studies does not have to be an overwhelming task. By proceeding in a systematic fashion, ensuring effective communication, following a disciplined system for managing documentation, and gathering all necessary information and needed personnel at the beginning of the process, researchers will be well able to design a realistic scenario that accomplishes the research objectives with the least amount of difficulty.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues at the FAA CAMI Human Factors Research Laboratory and Samuel Cotton for their contributions that led to this paper. We would also like to thank Walter Johnson, Vernol Battiste, Mary Connors and anonymous reviewers for their thoughtful comments and helpful feedback on an earlier draft of this paper.

REFERENCES

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research, 53*, 121-173.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*, 4279-4292.

Chou, C-D., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. *International Journal of Aviation Psychology, 6*(4), 307-320.

Law, A. M. (2003). How to conduct a successful simulation study. In S. Chick, P.J. Sanchez, D. Ferrin, & D.J. Morrice, (Eds.) *Proceedings of the 2003 Simulation Conference, 1*, 66-70.

Maldonado, G., & Greenland, S. (1997). The importance of critically interpreting simulation studies. *Epidemiology, 8*(4), 453-456.

Mumaw, R. J., Sarter, N. B., & Wickens, C. D. (2001). Analysis of Pilots' monitoring & performance on an automated flight deck. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Norman, V., & Banks, J. (1998). Managing the simulation project. In J. Banks (Ed.) *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. New York: John Wiley & Sons, pp 745-763.