Title

Enabling Autonomous Crew Task Performance with Multimodal Electronic Procedure Countermeasures

Author

Homer, Katie

Publication Date

2024

Peer reviewed|Thesis/dissertation

Enabling Autonomous Crew Task Performance with Multimodal Electronic Procedure Countermeasures

By

KATHERINE HOMER

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Mechanical and Aerospace Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Stephen Robinson, Chair

Rich Whittle

Sanjay Joshi

Committee in Charge

2024

Contents

List of Figures

List of Tables

Abstract

Enabling Autonomous Crew Task Performance with Multimodal Electronic Procedure Countermeasures

Future long duration exploration missions (LDEMs) conducted by NASA will have an increased need for crew autonomy during routine and emergency procedures due to the increased distance from Earth causing time delays in communications. Presently, ISS in-space tasks are completed by astronauts using simple text-based procedures supplemented with real-time communication between the crewmembers and mission control personnel. As LDEMs require increased crew autonomy, more information must be stored on-board such that it can be accessed by crewmembers in a timely and context-appropriate manner during procedure execution. Emergent technologies in multimodal interactions such as Internet-Of-Things (IoT) sensors and enhanced visual displays are likely to play essential roles in safe crew-autonomous procedure execution. With this in mind, two studies were conducted: a study in NASA HERA to test enhanced multimodal procedures in a spacecraft analog environment, and a Davis study to determine how individual multimodal enhancements affect task performance. The goal was to determine how subjects' task performance on a complicated manual repair task differed between enhanced procedures and traditional unimodal PDF procedures. An Enhanced Procedure Viewer system was developed that provided a variety of procedural enhancements: step navigation, enhanced visuals, real-time sensor feedback, and laser guidance. Results concluded that different enhancements in the multimodal procedure could decrease task completion time, increase task completion accuracy, decrease subjects' perceived workload, and increase the level of trust subjects had in the procedure system.

Acknowledgments

To everyone in the Human/Robotic/Vehicle Integration and Performance Lab, thank you for being such an amazing group to work with. I find myself inspired by every person and project in the lab and am honored to have had the opportunity to work amongst you all.

I would like to thank Prof. Stephen Robinson for all of his support and guidance throughout this project. Your insights and recommendations have been invaluable. I especially appreciate the opportunities you have provided to me; from zero-g parabolic flights to presenting at NASA conferences, this has been the best Masters's project experience I could have hoped for.

A big thank you to the rest of my committee, Prof. Rich Whittle and Prof. Sanjay Joshi.

Thank you to my family for all the love and unwavering support. From encouraging me to pursue my Master's degree to helping me edit this thesis, you have been the perfect cheerleaders.

Thank you to all of the subjects who volunteered both at NASA HERA and UC Davis; there were a lot of you and I could not have done this without you.

Special thank you to Dr. John Karasinski for helping originally develop this project and get it funded. I have appreciated all the feedback you have provided and look forward to publishing research papers with you.

And lastly, thank you to Kelden Ben-Ora, the best team member a team lead could have. We joined the HRVIP Lab and this project at the same time as undergraduates, and while I might have taken over as team lead the last three years, you have been just as important in ensuring this projects' success. I am honored to be handing over the torch to you as you continue the HCAAM legacy.

Chapter 1

Introduction

1.1 Overview

Current human spaceflight operations in low-Earth orbit depend on the near-continuous and reliable communication between the spacecraft members and Earth-based personnel at a Mission Control Center (MCC). NASA has spent over 60 years conducting manned spaceflight operations and throughout this time, all missions have maintained a near real-time communication loop between the astronauts and the ground crew (1). This space-to-ground communication is essential as it grants astronauts access to the vast knowledge and resources available on Earth as well as allows for flight controllers to supervise and collaborate with astronauts.

Flight controllers spend hundreds of hours per day monitoring the status of spacecraft and working on anomaly resolution, freeing astronauts to focus on other tasks (2). Crew time is usually "overbooked" so astronauts perform primarily complex tasks which require a physical presence on site including scientific experiments and maintaining the spacecraft (3)(4). In order to facilitate task execution, astronauts receive step-by-step instructions and are often guided in real-time by subject matter experts on the ground while most of the "monitoring and commanding" tasks are performed from ground personnel (4). The current near real-time communication is especially important for FDIR (fault detection, isolation and recovery) and emergency scenarios as it provides the astronauts with a team to collaborate with as they diagnose, repair, and recover from issues. ISS missions are supported by 80+ experts on the ground at any given time, with a combined 600+ years of system-specific experience across 22 unique console disciplines (2).

With the ISS experiencing approximately 1.7 anomalies per year requiring immediate response since 2001, the knowledge and manpower of ground teams has been vital for the continued success of manned spaceflight missions (5). Even during the Apollo program, the landing sites were selected so that line-of-sight communication with Earth was maintained (1). This was important considering the crewed Apollo missions experienced a total of 362 anomalies across 11 missions, 35 of which were considered urgent and significant (5)(6).

In the coming years as long-duration exploration missions (LDEMs) begin, the increased distance from Earth will cause significant delays in communication. Transmission latencies begin at a few seconds on the moon to a few minutes during transit to Mars, increasing as the distance from Earth increases. Once astronauts reach Mars, communication delays can reach up to 22 minutes one-way, leading to a 44 minute delay in round trip communication (7). In addition, it is likely that communication between the astronauts and MCC will be inconsistent with exchanges only possible during certain times of the day (1).

This new challenge leads to a need for increased crew autonomy which can be achieved by storing more information on-board in such a way that it may be accessed by crewmembers in a timely and context-appropriate manner during routine and emergency procedures. The 2020 NASA Technology Taxonomy reports that autonomous systems will increase crew autonomy by increasing knowledge of systems through state estimation and monitoring as well as assisting with fault diagnosis and prognosis through anomaly detection (8). Since astronaut crews on deep-space missions will not have the currently relied upon real-time support from ground experts, they must have access to procedures that are enhanced beyond the traditional text-on-page to reduce procedure deviations or execution errors. One way this could be accomplished is to develop procedure countermeasures to help mitigate issues caused by time delays. Countermeasures are defined as actions, devices, procedures, techniques, or other measures that reduce the vulnerability of an information system (9). In this case, procedure countermeasures would help to prevent astronauts from making procedural mistakes in the absence of MCC oversight.

Two categories of activities that currently rely upon substantial ground involvement are executing complex procedures and troubleshooting unanticipated safety-critical anomalies

(2). "Providing the crew with the right data at the right time to make the right decision is a fundamental challenge of crew autonomy" (2). With NASA predicting the chance of an anomaly occurring that requires an immediate response during Mars transit to be greater than 50%, enhanced multimodal procedure countermeasures will be especially vital for critical operations and malfunction recovery (2). While unimodal procedures typically only contain written text, multimodal procedures can include multiple "modes" of communication such as text, images, videos, or audio. Additional emerging technologies in multimodal interactions such as augmented reality (AR) visual displays, "Internet of Things" (IoT) sensor feedback, spatial audio, and tactile feedback are also likely to play a role in increasing crew autonomy and will be of use in what we define as enhanced procedures.

The purpose of this thesis is to develop recommendations for updated standards and guidelines for multimodal interactions with electronic procedures. Two experiments were performed to test a multimodal Enhanced Procedure system which utilizes real-time feedback from IoT embedded sensors. The procedures led subjects through a complex manual repair task: the disassembly of a generator to check the status of an internal piece. The studies investigated the effects of crew performance (efficiency and accuracy), workload, and trust with the use of multimodal enhanced electronic procedures, as compared to traditional PDF procedures.

1.2 Methods

1.2.1 Task Performance Metrics

Two metrics were considered when analyzing task performance of human subjects: efficiency and accuracy. For this study, efficiency is defined as the speed in which subjects complete the task and accuracy is determined by how correctly subjects follow steps in a procedure. Errors, or deviations from the written procedure, usually result in a decrease in task completion accuracy. Typically, well-trained and experienced workers complete tasks more efficiently and with greater accuracy; for example, studies show more experienced surgeons worked faster and had less hand distance movement than less trained individuals (10). Mechanical repair tasks are a similarly complex manual task which requires workers to efficiently use tools to accurately complete a procedure.

1.2.2 Workload

Workload is defined as "the perceived relationship between the amount of mental processing capability of resources and the amount required by the task" (11). In simpler terms, workload is the relative amount of mental effort required to complete a task. To determine subjects' perceived workload, they were asked to complete the NASA-TLX survey (12) after finishing the manual repair task. The NASA-TLX survey asks subjects to rate their performance on a scale of 1 (low) - 20 (high) across 6 dimensions: mental demand, physical demand, temporal demand, effort, performance, and frustration. All 6 metrics are then compared in pairs and the survey results are used to determine an overall workload rating.

In practice, better procedures require a lower workload as workers follow steps and find parts with less effort. Multiple resource theory stresses the importance of distributing tasks and information across various sensory channels to reduce workload (13). In procedures, this could be completed by incorporating multimodal technologies including enhanced visual, audio, and tactile feedback to reduce the workload from the oversaturated visual channel. Large amounts of text can be overwhelming so by conveying information through other methods, workers may find it easier to understand and follow complicated procedures.

1.2.3 Trust

As AI becomes more developed and machines become more sensorized, it is important to ensure that humans continue to have appropriate trust in the technology they work with (14). To assess this level of human trust in our Enhanced Procedure system, subjects were asked to complete a Trust in Automated Systems Survey developed by Jian et al (15). The survey is 12 questions long and asks subjects to rate on a scale of 1 (not at all) - 7 (extremely) their impressions when using the technology. Both negative and positive feelings are rated, such as feelings of deception, confidence, and reliability. Once complete, the individual scaled ratings are combined to generate an overall score of the workers' perceived trust in the system.

1.3 Literature Review

During the Apollo and Space Shuttle missions, procedures were printed in large paper volumes for astronauts (see Figure 1.1) (1)(3). In the 1990's, the Shuttle Program began migrating paper procedures to PDF-based procedures which were viewable on computers (16).

This progressed into the development of a PDF procedure viewing system called the Manual Procedures Viewer (MPV). Since then, the MPV has been replaced with the electronic XML based International Procedure Viewer (IPV) (17). To continue enhancing procedures and increase their effectiveness, a combination of multimodal communication modes can further be utilized to translate information through images, videos, audio interactions, augmented reality, animations, IoT (Internet-of-things) sensor feedback, or haptic feedback. These features could transform how procedures are used with sensors to highlight the locations of required tools, animations to illustrate complex steps and identify relevant parts, timers and data entry areas so that all required information is in the same location, and real-time interactive feedback from sensors to confirm step completion. Multimodal interactions have the potential to increase the communicative power of procedures and enhance procedures far beyond traditional text. As LDEM communication time delays increase and MCC has less capability to assist the crew, incorporating more multimodal countermeasures into procedures will allow astronauts to communicate with procedures and the technology they work with on a deeper level, granting workers a higher level of situational awareness and making procedures a conversation with two-way interactions.

Figure 1.1: Astronaut Lee Archambault, STS-119 commander, looks over checklists from the commander's station on the flight deck of Space Shuttle Discovery. Courtesy NASA (1).

Today (2024), the International Procedure Viewer is jointly used by flight crew and MCC to facilitate onboard tasks (18). Current U.S. procedures are written by NASA Operation Support Officers (OSOs) based on technical specifications provided by the experiment's vendor or technical engineer. The OSOs write the procedures using standard symbols and formatting defined by the Operations Data File (ODF) standard which the astronauts have been trained to interpret (19). An example procedure is shown in Figure 1.2 (20).

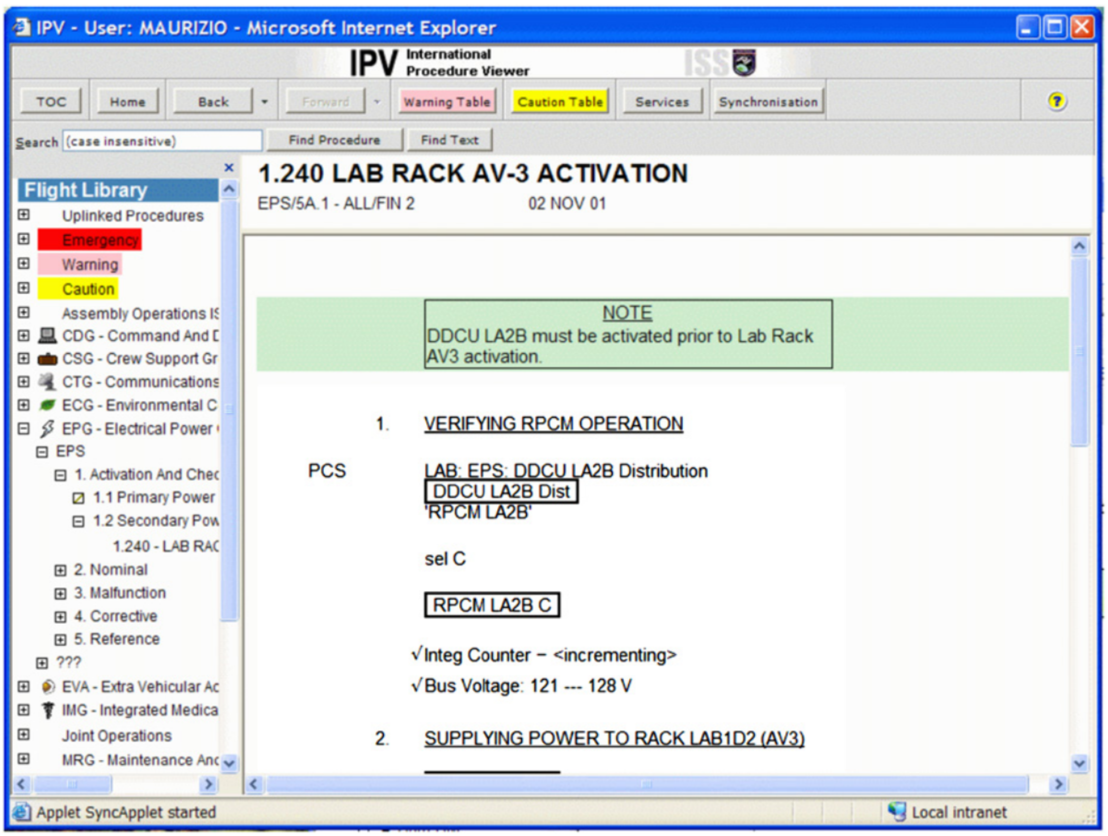


Figure 1.2: International Procedure Viewer, the web based browser used to display all electronic procedures aboard the ISS. (20).

Crew procedures have strict standards and specifications to ensure all data presentation is consistent to maximize crew efficiency and safety (21). Almost all procedures are electronic with the exception of a few safety critical procedures deemed important enough to be printed and stored on board. Once the OSO writes the procedure, the Word document is converted into XML using the software system Procedures Authoring Tool (PAT) (17). This required conversion from Word to XML is an inefficient process which currently requires a special team; future procedure systems should be a "one stop shop" to streamline the new generation of procedures.

The IPV, as seen in Figure 1.2, is a web based browser which transforms the XML procedures into HTML and displays all of the procedures for astronauts and MCC (20). These procedures provide step-by-step instructions for everything astronauts do onboard the ISS, including science experiments and spacecraft maintenance (3). The IPV system boasts numerous features including a step highlighter, images, data entry capabilities for organization (ex: report serial number), step searching, and links to videos and other related procedures.

All astronauts are taught the basics of how to use the IPV system (standard symbols, buttons, navigation, etc.) and then allowed to choose how they want to utilize the technology. This freedom allows for astronauts to use the features they find helpful and ignore those they do not. While the IPV system is not currently "smart" or connected to onboard telemetry, the ISS does have Timeliner, an automation system used to automate procedural tasks that would typically have been completed by a crewmember (22). Making more procedures entirely automatic is one way to reduce astronaut workload and ensure proper completion, but not everything can be automated, especially when it comes to onboard mechanical repair or unplanned tasks. Instead of replacing the worker with automation, this thesis looks to utilize multimodal enhancements such as IoT to assist the crew with manual repair tasks. By providing astronauts with additional tools and relevant information, their situational awareness of the spacecraft and technology they are working with can increase, thereby reducing the required workload and increasing accuracy (and therefore safety) when completing complex tasks.

Ideally, providing more multimodal countermeasures to astronauts will allow them to more accurately complete complex tasks which will be vital as available oversight from MCC decreases with communication delays. Currently, MCC typically monitors the state of the spacecraft, searching for anomalies by plotting telemetry plot trends over the span of weeks or months (23). While onboard software can easily identify rapid changes in the spacecraft, slower anomalies such as small leaks can be more challenging to detect without analyzing long term trends. Often, MCC is responsible for monitoring these trends and informs the crew when telemetry limits are violated (23). Due to limited bandwidth during LDEMs, MCC will no longer be able to downlink thousands of streams of real-time sensor data and multiple high-definition video feeds from the vehicle, so this responsibility will largely transfer to the onboard software and crew (24). MCC also currently monitors astronauts' progression of tasks via video to ensure satisfactory step completion. When video is unavailable, MCC must rely on available telemetry and communication with the crew. MCC typically identifies procedure deviations by either seeing issues arise real time in the camera feed, or by having the crew call down to notify them. While MCC may typically be able to identify if astronauts are struggling, missing steps, or interacting with the wrong hardware through visual observation, when there is no video or ground telemetry, MCC may not know there is a mistake until the item in question fails to work (23). This can waste precious crew time and may lead to unnecessary risks.

Not only will ground based assistance be delayed for near-future LDEMs, but the tasks the crew complete will likely not all be preplanned or trained for. Astronauts spend a significant amount of time training prior to missions. Training is an effective tool to improve task performance; it has successfully reduced errors in high-risk settings such as emergency rooms, aviation, and the military–but the way training is designed, delivered and implemented matters (25). The literature exploring the retention of complex skills is limited but shows that competence decreases over time with factors including: quality of initial training, practice or refreshers, personal factors, and task complexity (26)(27). Current experiments aboard the ISS range from a few hours to several weeks and often require significant preflight crew training (28). As mission length increases from months to years, preplanning all tasks and experiments prior to mission launch will become increasingly difficult since project objectives will likely evolve as the mission progresses and experiments begin to deliver results. As such, instead of training the crew on specific procedures, the crew will likely be trained more generically to use different lab equipment and tools in a variety of ways (29).

Another reason to train the crew on what all onboard equipment is made of and capable of is because there will likely be a scenario in a LDEM where an unforeseen tool or piece of hardware is needed and simply unavailable (29). Today aboard the ISS, if vital equipment must be replaced it can be sent up relatively easily or, should a life threatening emergency occur, the crew could use the Soyuz or Dragon capsule to evacuate back to Earth. These options will not be available in LDEMs because the spacecraft will be too far away from Earth for prompt assistance, requiring the crew to rely on what they have on board. Procedures must be sufficient for these emergency scenarios when critical, unplanned tasks must be completed under temporal pressure. Just-In-Time-Training (JITT) refers to teaching the crew a complex and critical task, usually hours or minutes before the task must be completed (28). For these scenarios, crews may need flexibility in pulling from and combining multiple procedures, as well as quick access to resources like system documentation and schematics (2). When using or fixing systems in unplanned ways, crewmembers must be provided with clear procedures and all relevant state and system information to make up for their lack of training (29).

One way to improve training and procedures is to make them more interactive by adding feedback or checklists. Checklists are now common in the aviation industry, and real time concurrent feedback during training and complex task completion is becoming more common in various industries including medicine and automotive maintenance (10)(30)(31). Numerous medical studies indicate that subjects trained to complete motor tasks such as CPR or cervical spine mobilisation (CSM) apply more accurate forces and produce the required coordination patterns more successfully when they are trained with automated real-time feedback devices (32)(33)(34). When students were CPR trained with concurrent quantitative feedback, not only did they have an immediate improvement in performance accuracy and consistency, but the improved performance still remained when tested a week later (32). Similar real time performance feedback could be incorporated into mechanical repair procedures by displaying telemetry information collected from system embedded sensors.

As previously mentioned, the Internet of Things is a "network of physical objects embedded with sensors, software, and other technologies to enable the collection and exchange of data" (35). IoT is commonly used on Earth to enable smart home devices such as virtual assistants (Amazon Alexa), to control lights (Philips Hue), temperature (Nest Thermostat), and to provide home security monitoring (Ring Doorbell) (3). IoT sensors are also being increasingly used in mechanical systems to assist in the monitoring and maintenance of smart factories (36), smart medical equipment (37), and smart vehicles (38). IoT has proven useful as a tool to enable predictive maintenance, automated robotic systems (35), environmental monitoring (39), automotive repair (30), habitat monitoring (40), precision farming (41), and fault diagnosis of technology such as industrial electrical panels (42). Sensors can provide basic information on location, orientation, movement, or any other measurable feature such as temperature, pressure, or motor rpm (43)(44). IoT allows each object to broadcast information about itself (ex: Door A is open) or where it is (ex: Module B is installed in Rack 6) (44). Numerous studies have shown the capability of IoT sensors and RFID tagged items to allow for tool location and identification (44)(45).

By placing RFID tags on tools and the crew, WiFi transceivers can provide a rough estimate of proximity between objects through RSSI (Relative Signal Strength Indication) analysis to guide the crew to desired locations. Further, accelerometers can confirm when the correct tool has been picked up and the crewmember can be notified of success via haptic feedback. This relevant information regarding the state of the system can be collected and displayed directly into procedures to increase the situational awareness of workers as they complete manual repair tasks. This telemetry could simultaneously be used to monitor procedure progression, ensuring that expected values are present and alerting the crew to unexpected anomalies (ex: power switch should be OFF).

Future spacecraft will be heavily sensorized to allow for continuous monitoring. As novel approaches to structural health monitoring move in favor of using sensors embedded directly into materials, technology including habitats themselves are becoming more self aware, capable of identifying faults or degradation, and predicting future failure points. For example, fiber-optic cables run through concrete were able to measure very slight fluctuations caused by changes in the concrete they were embedded in and inflatable habitat structures have used sensors embedded in the flexible structural restraint webbing layers to monitor the equipment (43)(46).

As the habitats astronauts live in become “smarter”, astronauts will learn to “talk” to the technology they work with to predict, diagnose and address problems. Using sensor information, robotic systems can be programmed to react accordingly and perform various tasks autonomously, however there will continue to be a need for human involvement for critical decision making and unplanned maintenance and repair. The embedded IoT sensor state information can be helpful to astronauts when completing complex tasks. However, little research has been completed to explore how best to transfer relevant information to the onboard crew. One such way is to incorporate the relevant information directly into crew procedures as demonstrated in studies conducted at NASA Ames. Marquez et al. successfully fed sensor data directly into a prototype procedure execution tool to provide feedback to users when they correctly (and incorrectly) followed procedure steps (3). The sensors successfully automatically tracked workers' progress throughout the procedure and the study demonstrated how the enhancements allowed "naive users to quickly and correctly complete procedures which would otherwise require additional training" (3). This thesis aims to continue this research by further identifying how different multimodal enhancements affect task performance.

An example of an enhanced procedure which utilizes real time feedback in use today is Boeing's electronic checklist (ECL) for commercial airplane cockpits which was certified in 1996. The electronic checklist is integrated with the airplane and has closed loop checklist processing where certain steps tied to aircraft sensors are verified complete by the checklist. The system is "essentially an intelligent aircraft checking the human" (47). Boeing's ECL requires pilots to accomplish steps in sequence to reduce errors and certain steps cannot be bypassed until the sensor feedback confirms completion. The ECL system is smart enough to display the correct checklist at the appropriate time and can even alert pilots when checklists have not been completed in critical phases of flight. Early prototype development began in the 1980s when accident research by Boeing revealed that crew procedural errors, specifically errors in accomplishing checklists, were causal or contributing factors in a substantial number of incidents and accidents (31). While Boeing originally designed ECL to prevent crew errors associated with paper checklists, it was determined that the ECL had many other benefits as well including shorter checklist accomplishment times, lower cognitive workload, and decreased training time (47). In an unpublished simulator study, Boeing found a 46% decrease in errors compared to paper checklists (48). Implementing a similar system into future spacecraft could allow for a significant increase in astronauts' abilities to monitor and manage increasingly complicated systems with progressively less MCC oversight.

While significant research and training has been implemented to reduce pilot errors during flight, it is estimated that for every 1 hr of helicopter flight, 12 man-hours of maintenance occurs. A similar ratio can be expected for other similarly complex aviation services including airplanes and spacecraft so the importance of effective aircraft maintenance cannot be overstated. Notably, between 12 and 15% of the global aviation accidents are caused by aviation maintenance errors and this increases to 23% when serious incidents are included (49). In a study performing root cause analysis on 58 maintenance-related helicopter safety occurrences, H. Rashid et. al. determined that 25.88% of the occurrences were caused by "Incorrect installation/assembly" or "Part(s)/material omitted at installation/assembly". Further, procedural mistakes, loss of situational awareness, and improper procedures were specifically identified as conditions that contributed to several incidents (49). This is a substantial percentage of aviation accidents that could be preventable with improved maintenance procedures.

1.4 Research Objectives

This thesis is designed to test if multimodal enhancements such as IoT sensor feedback will improve task execution, and therefore enhance the safety of the long-duration crew. Multimodal procedural enhancements that will be tested include: IoT sensors to provide real-time feedback on step completion, laser guidance to point out key areas of interest, interactive step navigation, and enhanced visuals which include videos for complicated steps. This study will determine the effects of participant task performance (efficiency and accuracy), situational awareness, workload, and trust with the use of multimodal procedures compared to traditional PDF procedures. We hypothesize that enhanced multimodal procedures will increase accuracy, decrease workload, and not change the efficiency or trust subjects have in the system. From the results, recommendations for updated standards and guidelines for multimodal interactions with procedures will be developed.

Our research has four main aims:

1. Develop multimodal procedure tools for autonomous crew task performance.

2. Evaluate crew task performance (efficiency, accuracy and workload) when using the procedure tool.

3. Validate effects of multimodal countermeasures in the NASA HERA analog environment.

4. Report findings and recommend standards and guidelines for multimodal interactions with procedures.

Chapter 2

Generator Task and Procedure Development

2.1 Generator Repair Task and Enhanced Procedure Viewer

To test different multimodal procedural enhancements, a manual repair task was developed for a Honda EU2000i Inverter Generator, serving as an analog to a complex spacecraft subsystem. This repair task has the advantage of utilizing entirely commercial off-the-shelf equipment (COTS), which is widely available, and is easily relocated. The generator is an excellent spacecraft subsystem analog as it has many similarities with spacecraft hardware components – precision manufacturing, integration of mechanical, electrical, and fluid-handling components, and requirements for a variety of manual tools for inspection and repair.

Subjects in the study were tasked to complete a complicated and challenging manual repair task on the generator to check the status of the float valve tip (see Figure 2.1). This task requires subjects to partially disassemble the generator, removing numerous parts (screws, nuts, tubes, gaskets, etc.), to access the carburetor float. It initially takes around 45 minutes to complete the tasks' 69 steps and requires numerous manual tools including a screwdriver, ratcheting socket wrench, clamps, and pry tools.

To complete the repair task, subjects are provided with a procedure displayed on a Surface Pro tablet. For this study, a novel, browser-based Enhanced Procedure Viewer (EPV) was developed as shown in Figure 2.2. The EPV combines multiple multimodal procedural

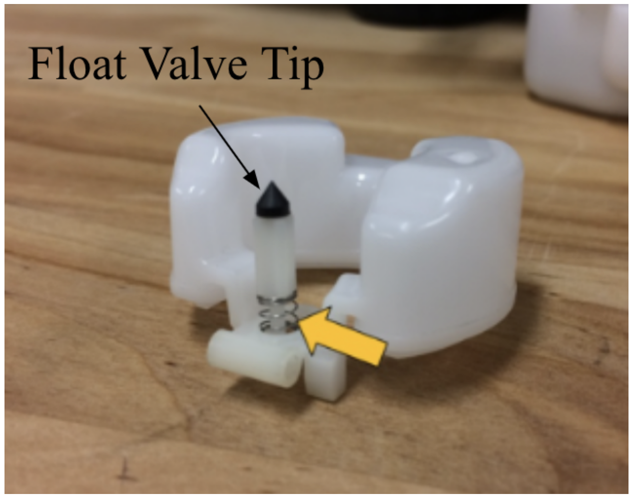


Figure 2.1: The procedure guides subjects through the disassembly of a generator to check the status of the float valve tip.

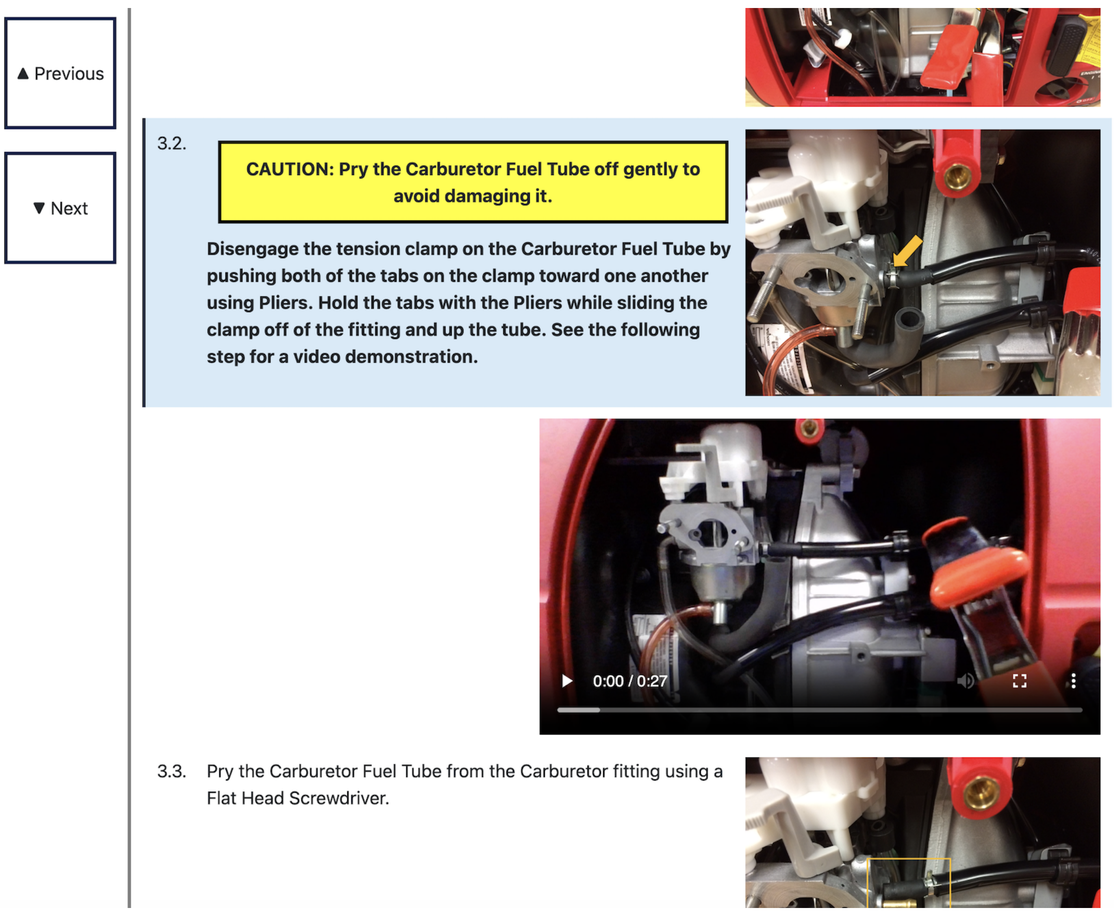


Figure 2.2: The procedure interface is a browser-based software that integrates sensor information, videos, and laser guidance directly into the procedure steps.

enhancements in addition to the traditional written text to aid the subject as they complete the manual repair task. The enhancements included in the EPV are outlined in Figure 2.3 and include: step navigation, enhanced images and videos, interactive sensor feedback, and laser guidance.

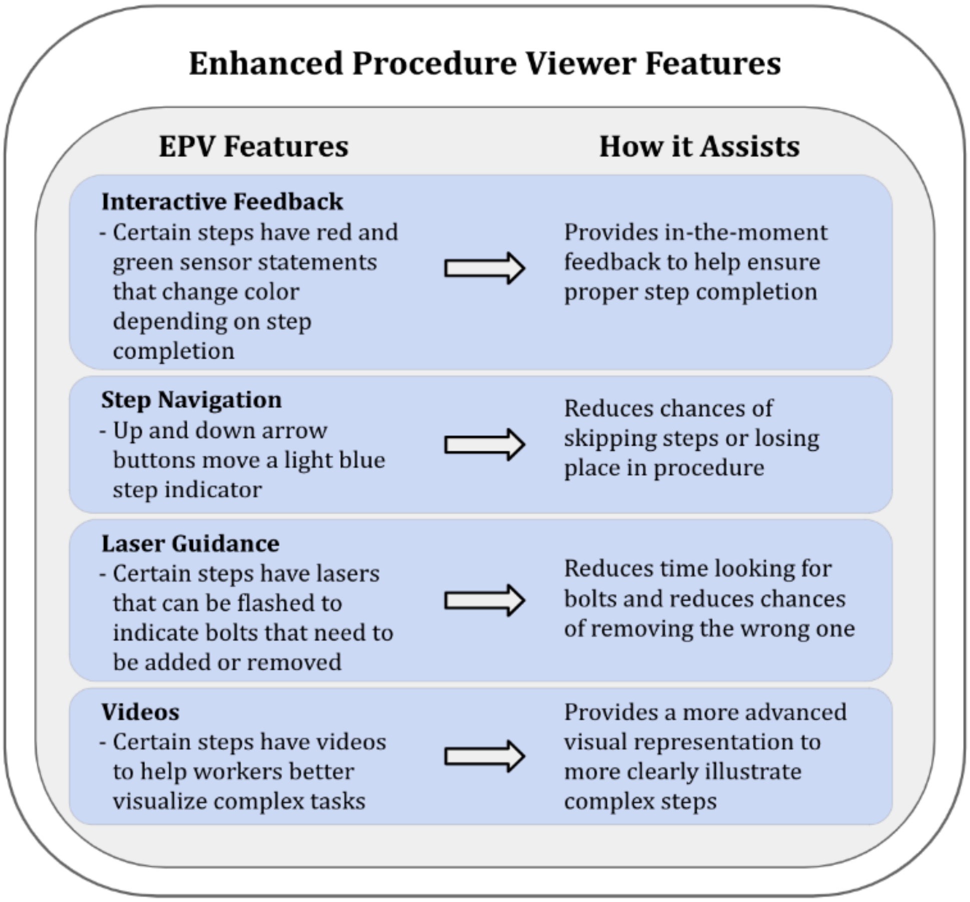
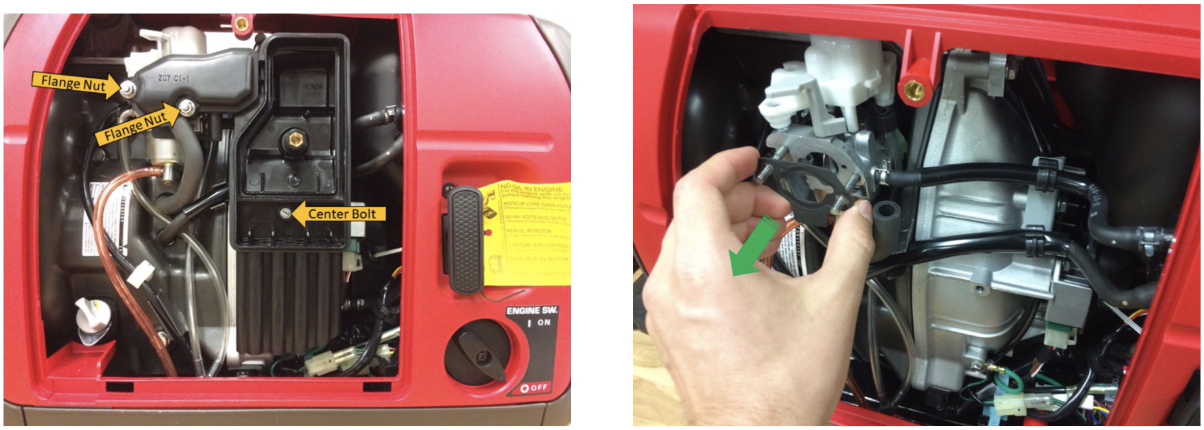


Figure 2.3: Features of the Enhanced Procedure Viewer.

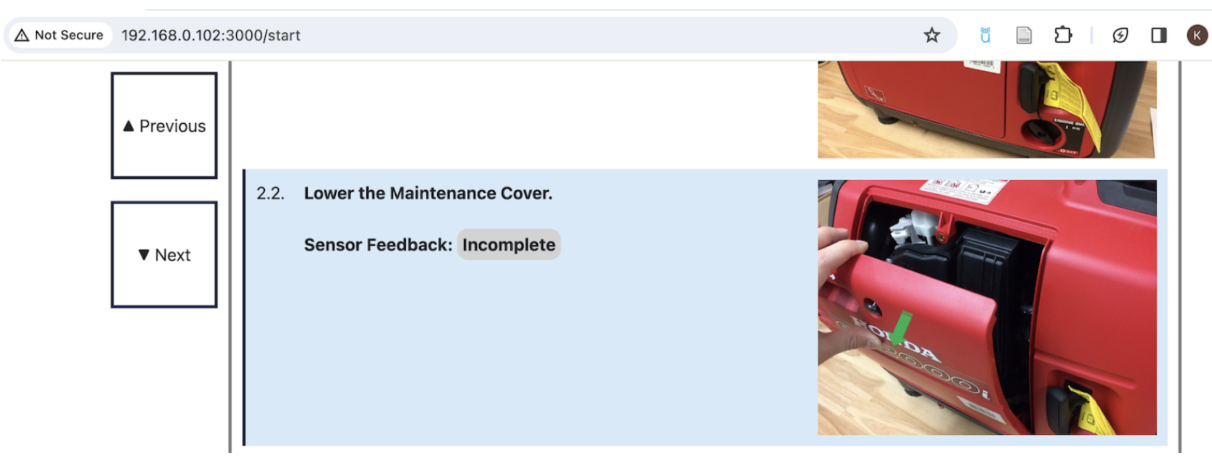
The EPVs' step navigation is controlled via up and down arrow buttons on the left side of the screen which move a light blue step indicator as shown in Figure 2.2. This navigation aid can help subjects keep track of their progress and reduce errors from missed steps. Enhanced images were developed for each step with color coded overlays to help subjects identify relevant parts (yellow) and indicate how they should be manipulated (green) as shown in Figure 2.4. The enhanced visuals also include videos for complex steps which provide additional context and support for the subject.

Interactive sensor feedback is provided for certain steps via gray and green step com-

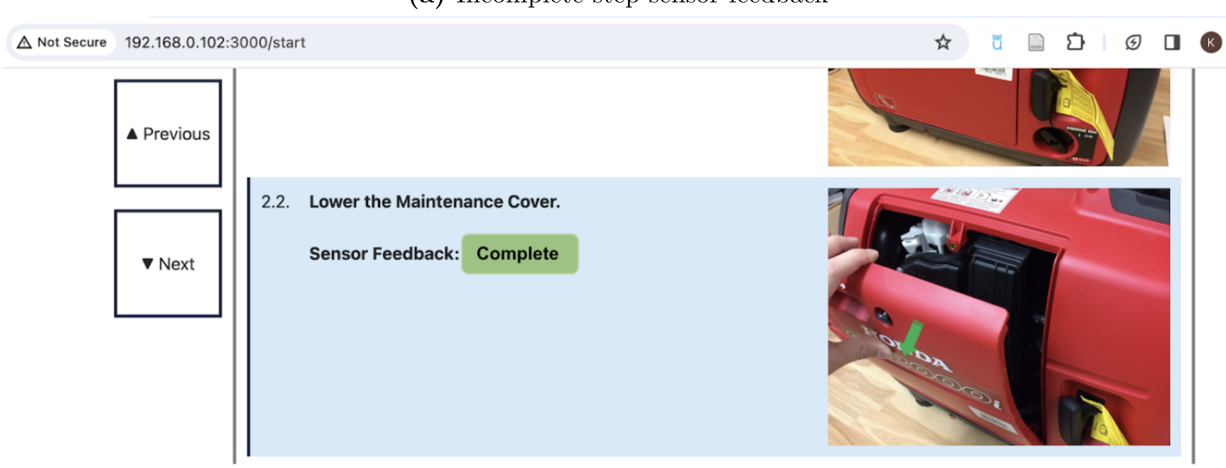


(a) Labeling object location **(b)** Labeling object manipulation

Figure 2.4: Enhanced images have color coded overlays which identify objects of interest in yellow (a) and how items should be manipulated in green (b).



(a) Incomplete step sensor feedback



(b) Complete step sensor feedback

Figure 2.5: Gray sensor feedback of incomplete steps (a) and green sensor feedback of completed steps (b) are provided to subjects as real-time step completion confirmation.

pletion statements as shown in Figure 2.5. This step completion information is generated from sensors embedded into the generator which monitor the state of the system and provide real-time feedback on whether the step has been completed correctly. The embedded sensors include: photoresistors, hall effect sensors, and an accelerometer with each sensor feeding step-specific data back to the subject through the EPV.

Before step completion, the feedback is gray; once the sensor identifies the step has satisfactorily been completed, the feedback turns green to indicate the subject can confidently move on. Lastly, on-when-needed laser indicators have been built into the generator to provide aid in locating objects of interest. These lasers are activated by clicking a button embedded in the EPV as shown in Figure 2.6. Specifically for this task, these laser indicators can be flashed to highlight which bolts need to be removed/replaced which can reduce time spent searching and reduce the chances of manipulating incorrect items. An example of the lasers in use is shown in Figure 2.7.

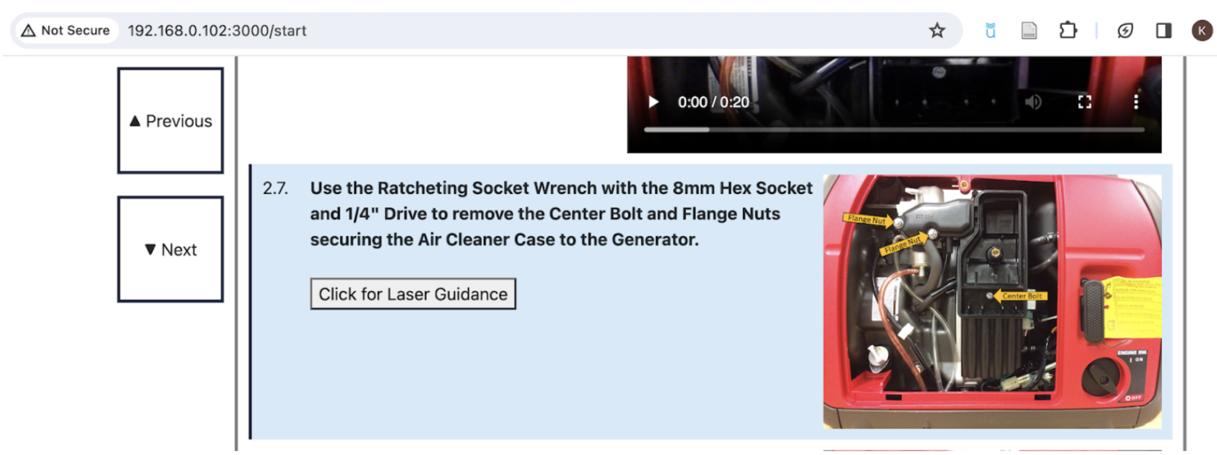


Figure 2.6: Laser activation button in EPV.

Since this EPV integrates the dynamic data from sensors directly into the procedure in real time, the crewmember is provided with enhanced situational awareness, allowing them to work more autonomously. Many of the enhancements, such as the step videos and laser guidance, are optional so subjects can choose when they want additional assistance.

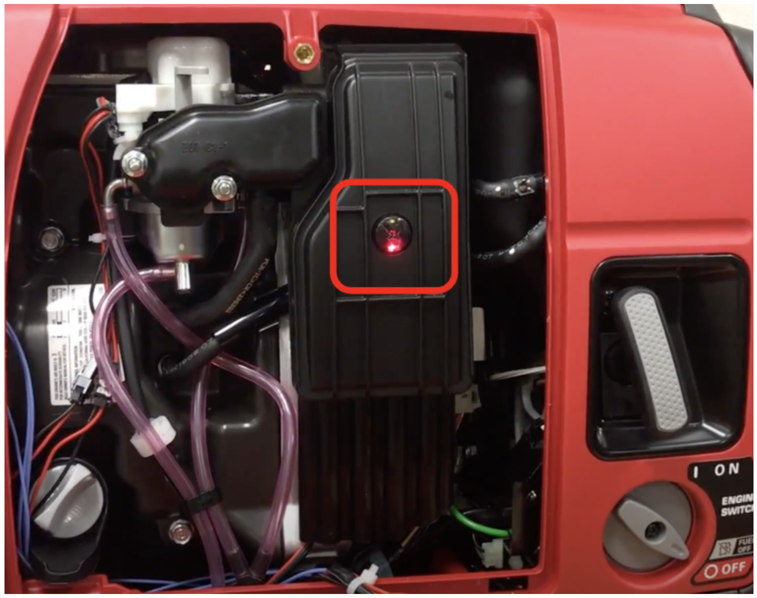


Figure 2.7: Laser indicators (boxed in red) can be activated to identify the screws/nuts that need to be manipulated.

2.1.1 System Architecture

To support the EPV's electronic features, a Raspberry Pi microcomputer and sensors were built into the Honda generator. Our system architecture is outlined in Figure 2.8 with the two-way communication between sensors and EPV facilitated via a router. During the task, the Raspberry Pi continuously monitors the sensors while the EPV continuously monitors subject requests for laser activation and step navigation.

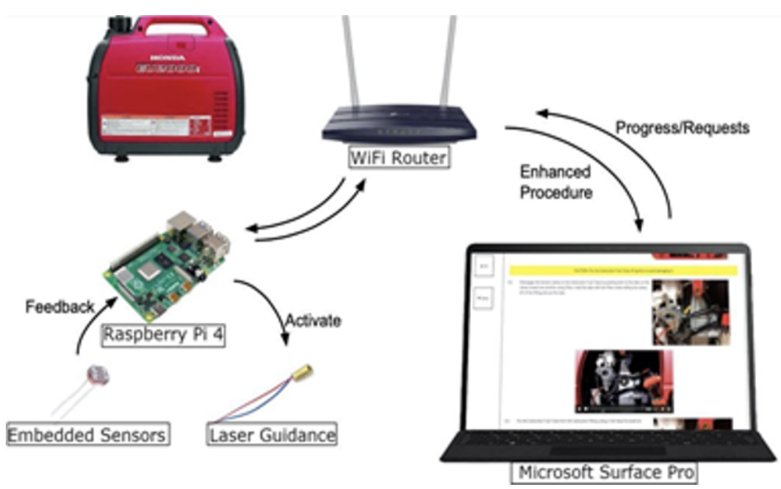


Figure 2.8: Spacecraft subsystem analog and its supporting system elements.

The Raspberry Pi is hardwired to the embedded sensors and attached to the outside of the generator as shown in Figure 2.9 with the sensor wiring exiting out of the left generator panel. The Raspberry Pi is responsible for monitoring the sensors (thereby monitoring the state of the generator) and relaying this real-time information to the EPV via a wireless router.

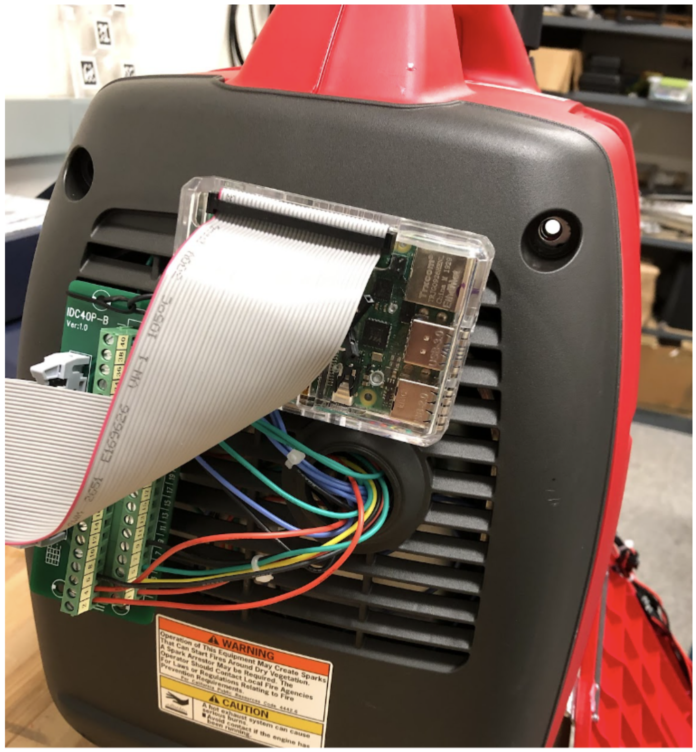
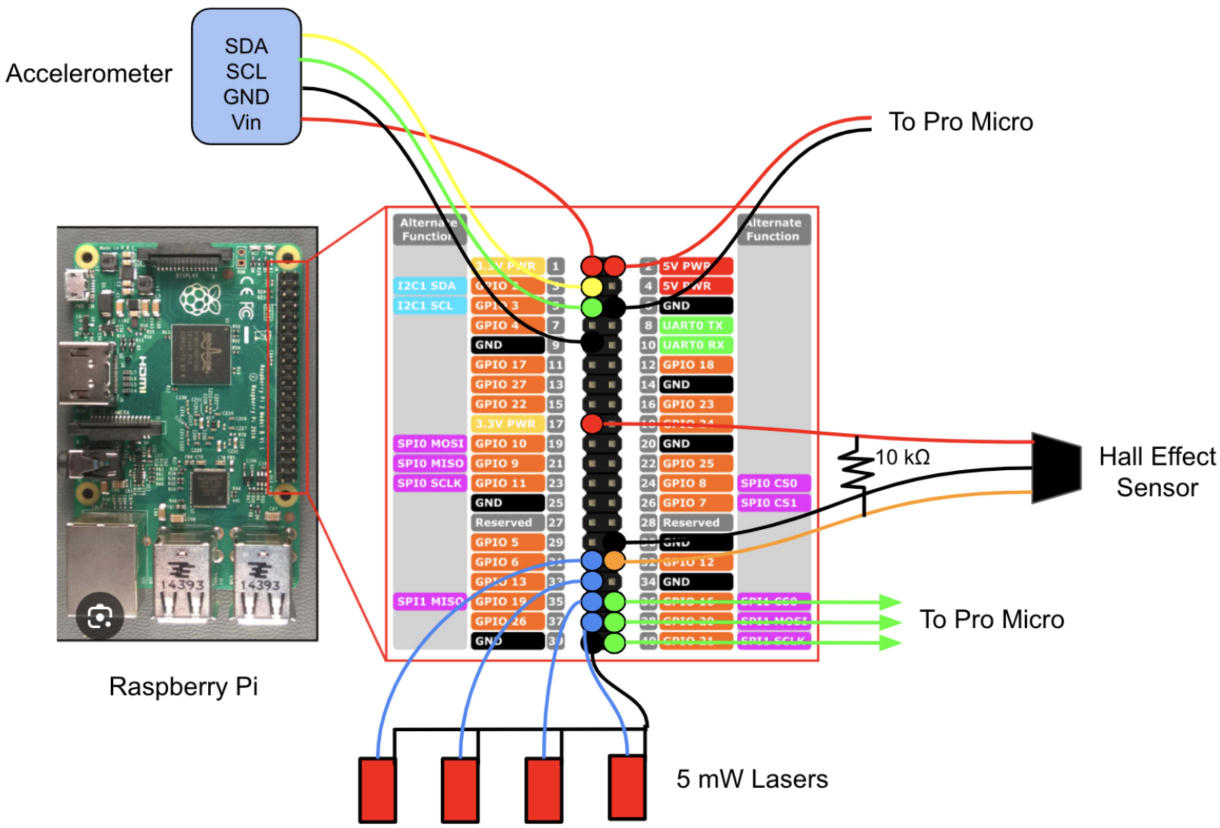
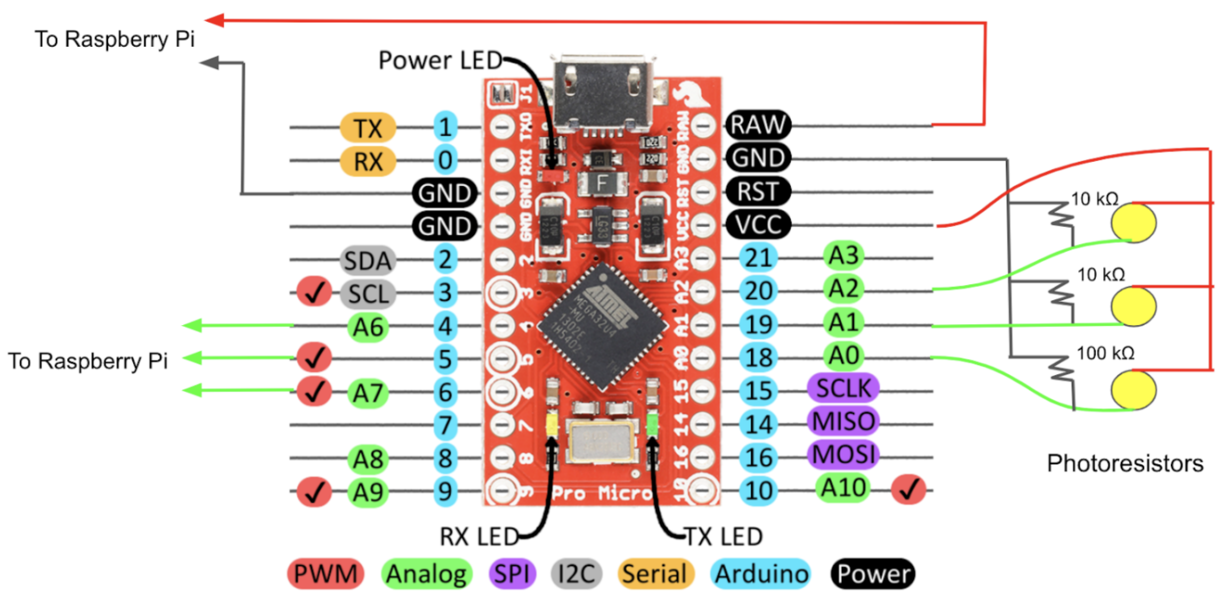


Figure 2.9: The Raspberry Pi microcomputer is hardwired to the embedded sensors and attached to the outside of the Honda generator.

A wiring diagram for the Raspberry Pi and sensor system is illustrated in Figure 2.10. Embedded sensors include: an accelerometer, hall effect sensor, 4 lasers, and 3 photoresistors connected to a Pro Micro. Each of the sensors was built so they could be easily replaced in case of damage. They are secured into the generator with velcro or zip ties and wired to the Raspberry Pi with JST wire connectors for easy removal. There are three photoresistors embedded into the generator, all of which are attached to the Pro Microcontroller for state monitoring. Each photoresistor is placed in a 3D printed housing to keep them correctly oriented as shown in Figure 2.11. The velcro is attached to the underside of the 3D printed housing.



(a) Wiring diagram connecting the Raspberry Pi to the embedded sensors.



(b) Pro Micro wiring diagram connecting the photoresistors to the Raspberry Pi.

Figure 2.10: Wiring diagram for the embedded sensor system.

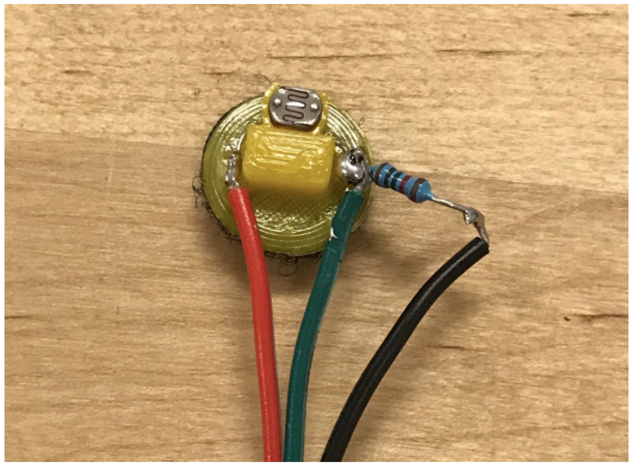


Figure 2.11: Photoresistor setup with 3D printed housing and resistor.

The photoresistors are placed into the generator so they can identify when caps and covers have been opened. An example of this is shown in Figure 2.12 where two photoresistors monitor the removal of the spark plug cover and spark plug cap respectively. All three photoresistors are directly connected to a Pro Micro which is placed inside the left cover of the generator as shown in Figure 2.13. This Pro Micro monitors the photoresistor outputs and relays when the observed light reaches a certain threshold indicating step completion to the Raspberry Pi.

Figure 2.12: Photoresistors velcroed into the spark plug cover and spark plug cap of the generator.

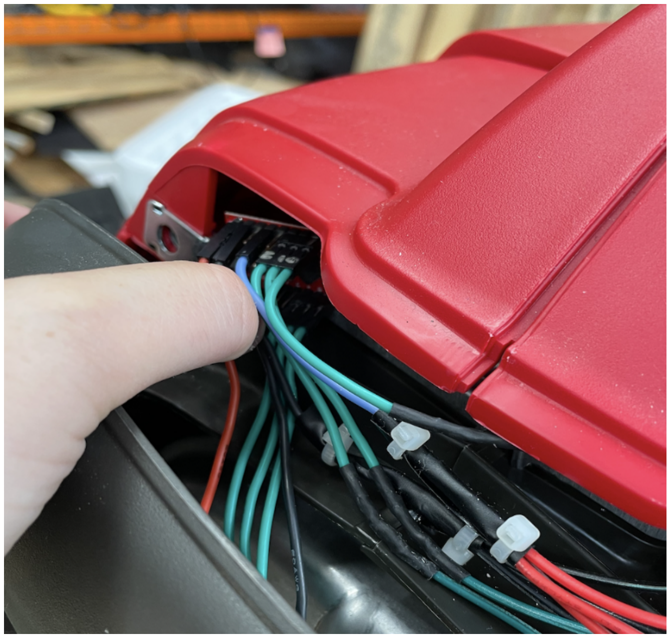


Figure 2.13: The Pro Micro is stored inside the generator's left cover.

A hall effect sensor is embedded into the generator with velcro, as shown in Figure 2.14, such that it may monitor the placement of the generator's air cleaner case. A magnet is attached to the air cleaner case that alters the magnetic field around the hall effect sensor, allowing the sensor to determine when the air cleaner case is in place.

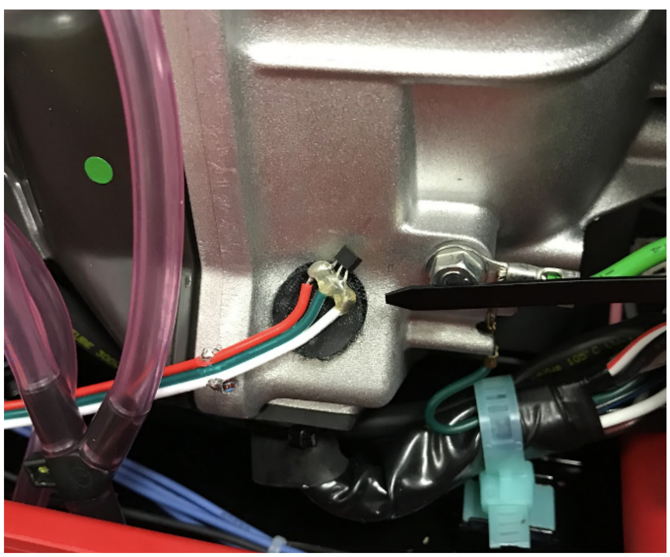


Figure 2.14: A hall effect sensor is embedded in the generator to monitor the removal of the air cleaner case.

An accelerometer is zip tied to the carburetor as shown in Figure 2.15 to determine its vertical orientation. To access the float valve tip, the carburetor must be removed and flipped upside down so the float chamber can be removed. The accelerometer can thereby inform subjects whether they have the carburetor in the correct orientation to move forward with disassembly.

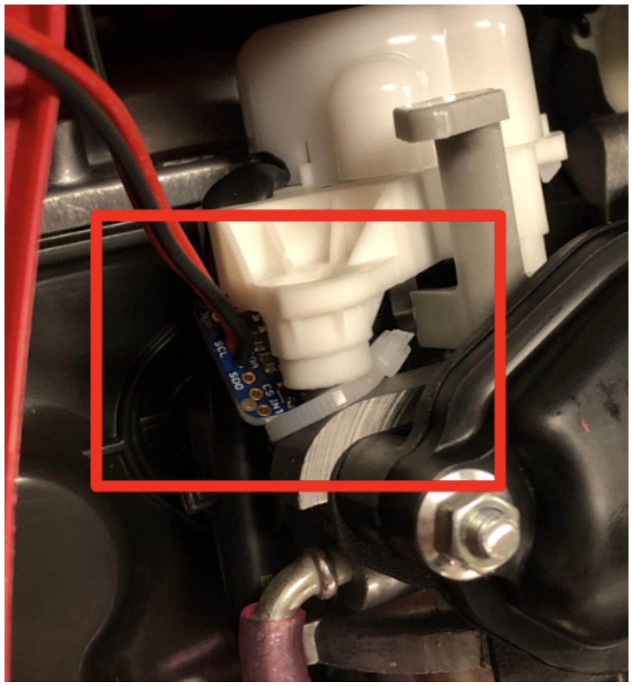


Figure 2.15: An accelerometer is zip tied to the carburetor to monitor the carburetor's vertical orientation.

Lastly, four 5 mW lasers were placed into 3D printed housing and bolted to the front cover of the generator as shown in Figure 2.16. By bolting the lasers directly into the generator, accurate placement can be consistently guaranteed without continuous adjustment of external hardware. The lasers were placed into position by measuring the approximate angles and adjusting the movable 3D printed laser housing until accurate placement was achieved. The 3D printed housing was then bolted to the front cover and further superglued into place to prevent rotation and drift. This process is illustrated in Figure 2.17.

While the sensor status of the photoresistors, hall effect sensor, and accelerometer were monitored by the Raspberry Pi and translated to the subject via the EPV, the lasers were activated by the subject through the EPV. As such, the EPV monitored for laser requests

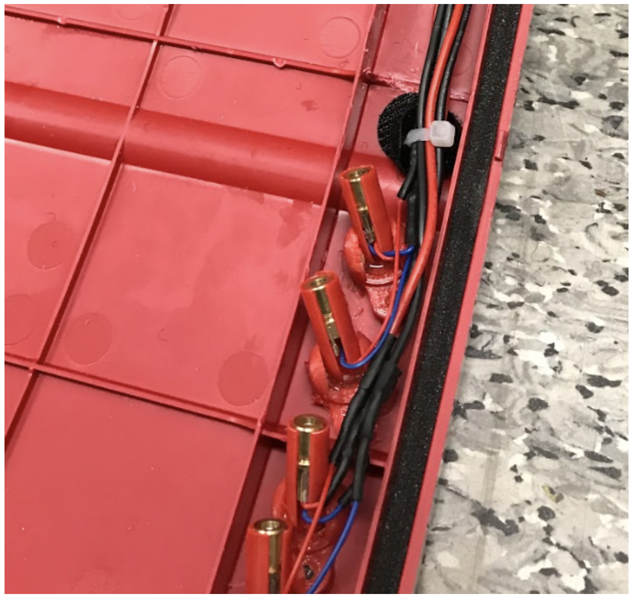


Figure 2.16: Lasers are placed into 3D printed housing and bolted to the front cover of the generator.

Figure 2.17: The lasers were placed into the approximate locations and finely adjusted until optimal placement was achieved.

from the subject and sent the requests to the Raspberry Pi which then activated the indicators. When requested, the lasers flashed on and off to dynamically draw the subject's eye to the locations in question.

Chapter 3

UC Davis Study

3.1 Experimental Design

A study was conducted at UC Davis to identify how different procedural enhancements affect task completion. The goal of the study was to determine each of the enhancements' relative sensitivity and how they uniquely affect procedure execution. To do this, 5 procedure variations were developed for the generator repair task, each of which has a different combination of enhancements. These procedures include the: PDF Procedure, Base Procedure, Visuals Procedure, Feedback Procedure, and Enhanced Procedure; a breakdown of their enhancements is illustrated in Table 3.1.

Table 3.1: UC Davis study breakdown of procedural enhancements for each procedure type.

	Step Navigation	**Enhanced Visuals**	**Sensor Feedback**	**Laser Guidance**
PDF				
Base	X			
Visuals	X	X		
Feedback	X		X	X
Enhanced	X	X	X	X

Each of the 5 procedure types has identical wording for each step, what changes between them are the provided multimodal enhancements. First is the PDF Procedure which is a traditional, scrollable pdf document with no enhancements. The Base Procedure is identical except it is an electronic web-based procedure with added step navigation via a step highlighter which is moved by the subject with the up and down arrow buttons.

Note that while the PDF Procedure and Base Procedure do not have the enhanced visuals feature, the procedures do have basic pictures. Some of these images are from the original generator manual which are not colored and identify numerous parts per picture as shown in Figure 3.1. As such, not every step has an image and not every image includes a colored identification overlay to quickly identify the specific item in question.

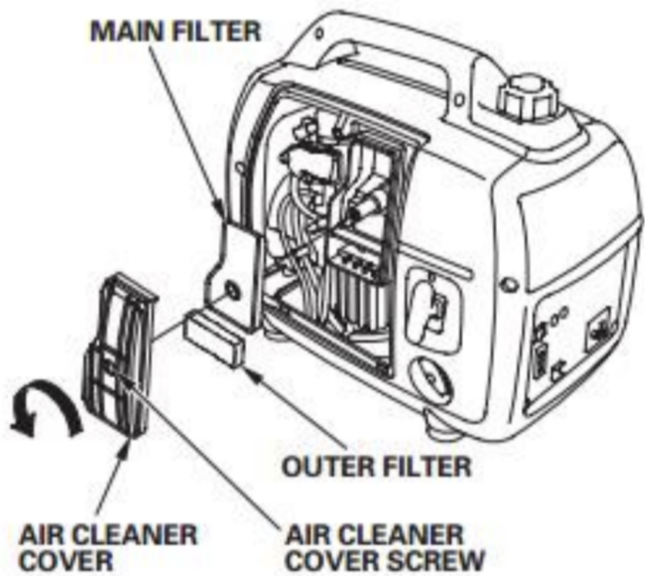


Figure 3.1: Example image from the Base Procedure from the original generator repair manual.

In comparison, the Visuals Procedure does have an image for every step with color coded visual overlays to identify which parts are referenced (yellow) and how they should be manipulated (green). Step navigation is also included, as well as videos to demonstrate complex steps to add additional situational awareness and context to the subject. Next, the interactive Feedback Procedure has step navigation, sensor feedback to provide real-time feedback on step completion, and laser guidance to aid the subject in identifying objects of interest. Lastly, the fully Enhanced Procedure combines all of these enhancements together: step navigation, enhanced visuals, sensor feedback, and laser guidance.

To determine our required sample size, we performed a power calculation using G*Power (50). Our sample size justification is based on the primary dependent variable of accuracy. We do not have a precise estimate on expected effect magnitudes and variability, however we do expect a large effect on deviation rate for each independent variable. Manual assembly tasks in literature have often found large effects on task performance from training subjects (51)(52), and procedural type tasks have found that interactive cognitive aids can provide significantly lower error rates from even well-trained subjects (48).

The "default" probabilities of Type 1 and Type 2 error weightings, $\alpha = 0.05$ and $\beta = 0.20$, are appropriate for this study because this is a foundational study without limitations on the n value. Typical Human Research Performance (HRP) studies have problems associated with Type 2 error caused by unavoidable small n values (53). Because this is not an analog experiment, we do not have this limitation and as such are justified in using the "default" error weightings. Using the power calculator G*Power, Figure 3.2 was developed which shows a range of effect sizes and their correlated total sample sizes. Estimating the effect size to be large (f = 0.8), we estimated that we would need at least 25 total participants (5 per group) to achieve a power of 0.8. In comparison, for the smallest effect we might expect (f = 0.5), the largest total sample size we would require is 55 subjects (11 per group). As we expected all our dependent measures to be normally distributed, we planned to conduct ANOVAs to analyze the independent variable (type of display). This was a between subjects study so subjects were randomly assigned one of the five displays and no participants repeated the experiment under an additional display.

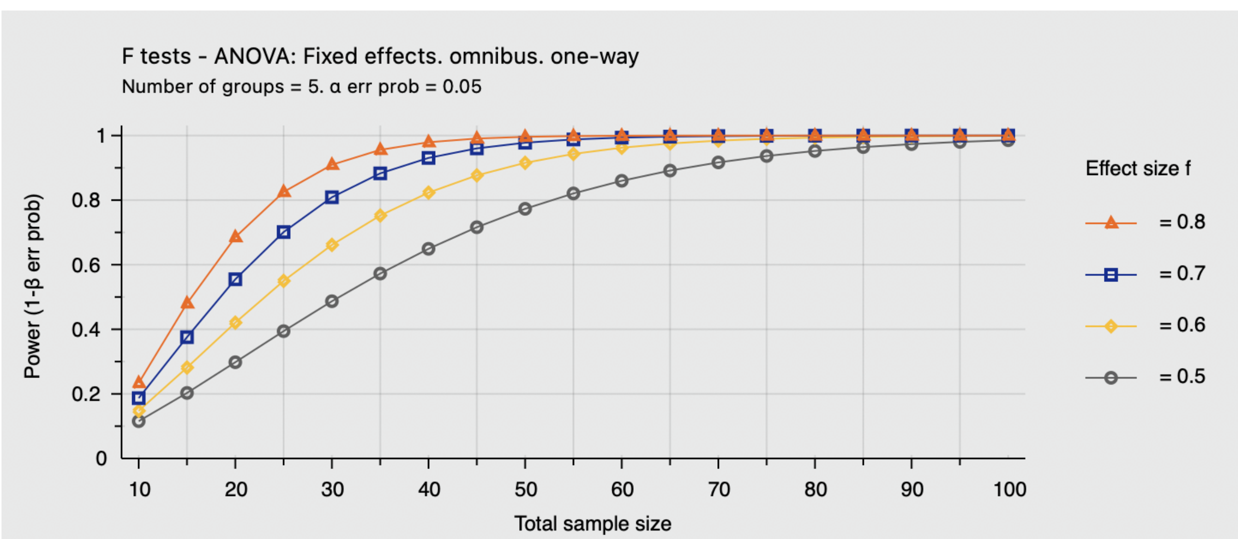


Figure 3.2: Power calculation to determine total required sample size based on expected effect size. (50)

To ensure the study had sufficient power, 12 subjects completed the task with each of the 5 procedure types leading to an n = 12 per procedure and n = 60 for the study. Subjects were recruited from the UC Davis Engineering population and signed a consent form prior to the study. Before the subjects started the task they completed a demographic survey asking questions about their background (gender, age, education), aptitude for using technologies (computer self-efficacy), and aptitude for mechanical repair/maintenance. They also completed a tool identification questionnaire to ensure they had a baseline familiarity with mechanical repair tools. This questionnaire was used as the studies exclusion criteria; should the subject fail to identify 75% of the tools they would be excluded from the study to ensure all subjects had a relatively similar mechanical repair background.

After the demographics and tool identification questionnaires, the subjects watched a training video which introduced them to the tools they would be using for the mechanical repair task, the EPV, and the goal of the generator repair task. Subjects were encouraged to ask questions at this time and each subject was also encouraged to practice with the ratcheting socket wrench prior to the experiment beginning. The ratcheting socket wrench was one of the more complicated mechanical repair tools used during task execution and prior testing indicated that subjects frequently misused the tool. To prevent confusion and ensure that all subjects began the task with the same baseline knowledge, subjects who did not know how to use the ratcheting socket wrench were taught at this time.

After all questions were answered, subjects completed the generator repair task, disassembling the generator to reach the float valve tip. Video was recorded for each of the trials to allow for post video analysis of both timing and accuracy. After task completion, subjects were given three post-task surveys:

1. NASA TLX survey - to assess their cognitive workload
2. Trust survey - to assess their trust in the procedure system
3. Post-Task survey - to collect their thoughts on the procedure system and their task performance

Our primary dependent variables for human performance are efficiency, accuracy, and workload. We will also be considering the subject's level of trust in the system, their confidence when completing the task, and how they perceived their task performance.

3.1.1 Hypotheses

We hypothesized that:

1. When subjects used the enhanced visuals, they would complete the task more efficiently as it would be easier to locate the parts mentioned in each step.

2. When subjects used the real-time sensor feedback and laser guidance they would report a higher subjective confidence during task execution as the enhancements confirmed adequate step completion.

3. When subjects used the step navigation, they would complete the task more accurately as the navigation tool reduced missed steps.

4. Subjects would report a decreased subjective workload when using either the enhanced visuals, sensor feedback, laser guidance, or step navigation.

5. Subjects would not report a change in their level of trust in the system dependent on any of the enhancements.

3.2 Experimental Results

3.2.1 Participants

We recruited 60 subjects (35 males, 25 females) from the UC Davis student engineering community. All subjects gave informed consent in accordance with the UC Davis Institutional Review Board. Their average age was 21.3 ± 4.27 (mean $\pm$ sd) years overall (21.7 ± 5.23 years for males, 20.7 ± 2.32 years for females). Of the subjects, 56 were very familiar with operating a laptop, 32 were mechanics or had experience with manual repair tasks, and 29 used mechanical hand tools often. Subjects also rated their mechanical repair ability on a scale of 1-5 with higher numbers indicating an increased ability to perform mechanical repairs. After the study, a Kruskal-Wallis test was performed to determine if there was a correlation between subjects' reported repair ability and the procedure type they were assigned to. The test determined that subjects were reasonably distributed with subjects' self-reported manual repair ability not being correlated with procedure type $\chi^2(4) = 1.17$, p = 0.884.

3.2.2 Analysis

The main experiment time log for each trial was generated by the Raspberry Pi microcomputer that was running the EPV system. During task execution, the Raspberry Pi recorded all sensor data and EPV navigation (usage of the up and down arrow step navigation buttons) to a downloadable file. Subjects were additionally tasked with manually recording the time at predetermined points throughout the procedure as a backup. Because a proctor was visually observing the subjects during the experiment, any subject that forgot to use the step navigation system after 2-3 steps was reminded to use the feature. Post-experiment video timing analysis was also completed for each trial to update the time log for any of these lapses.

For the PDF Procedure, the subjects used a PDF displayed on a Surface Pro which did not have any step navigation or automatic time recording abilities. As such, for these 12 trials the timing data was entirely analyzed post-hoc by the study proctor using the audio and video recordings. This video analysis was completed by a single individual to ensure consistency in the analysis process. Past studies in the HRVIP Lab indicate that multiple reviewers can be trained to consistently analyze videos identically, however this is a time intensive, iterative process in which the reviewers must continuously review and discuss different task scenarios to establish a consensus on the analysis procedures (54). This process can take a significant amount of time, so for this study a single individual was chosen to complete all video analysis.

To record step time, the reviewer made note of the beginning time for each individual step. The end of one step therefore corresponded to the beginning of the following step which ensured the entirety of the task completion time was recorded. Because the EPV had a navigation system, steps were defined as "starting" when the subject clicked the "Next" button, thereby highlighting the "new" current step in question. For instances when subjects were not consistently using the provided navigation system, steps were defined as starting when subjects finished the previous step and redirected their attention back to the procedure tablet. If a subject completed multiple steps without checking the procedure tablet, the current step was defined by what the subject was actively touching or focusing their attention on. All time spent on the task (time from beginning step 1 to clicking the "End Trial" button) was accounted for.

To assess task accuracy, the accuracy of each step was considered. During the experiment, the proctor noted any observed procedural deviations. After the completion of all trials, the proctor re-watched them again via the recorded video to ensure all deviations were caught and recorded consistently. Rather than considering accuracy in a binary sense (correct or incorrect), steps that were not completed exactly as described in the procedure were defined as deviations and classified based on the technique outlined in "Methodology to Quantify Accuracy for Procedure Execution Analysis" (54). This paper presents a methodology, Procedure Deviation Analysis, or PDA, which provides a quantitative measure of accuracy that provides insights for training efficacy and procedure design due to its broader consideration of task execution errors. The term "error" tends to denote a mistake with a negative connotation while PDA aims to look at procedure execution in a more holistic sense identifying any deviations from the task instructions. This includes deviations that do not necessarily impact the successful completion of the task. A breakdown of the different deviation modes from the paper is outlined in Table 3.2.

Deviation Mode	Description
Omitted	Does not attempt any part of the step
Partial Omit	Completes part, but not all of the step
Fragmented	Completes the step in multiple parts with other steps in between
Sequential	Performs all or part of the step out of order
Execution	Does not complete the step as described in the procedure
Extra Action	Repeats or attempts to repeat all or part of the step

Table 3.2: Brief description of the different deviation modes (54).

PDA identifies 6 deviation modes: omitted, partial omit, fragmented, sequential, execution, and extra action. Omitted deviations occur when a subject completely neglects to perform a step. Partial omit deviations occur when subjects correctly perform part, but not all of a step. Fragmented deviations are when a subject completes a step in multiple parts with other steps in between. Sequential deviations occur when steps are performed correctly but out of order. Execution deviations occur when the subject attempts the step but does not complete it as described in the procedure. And lastly, extra action deviations are when subjects repeat, or attempt to repeat, all or part of the step. For the purposes of our study, we also added two deviation types called "assist", for whenever a subject asked for assistance in order to complete the task, and "interrupt", for whenever a proctor interrupted an inappropriate action. Interrupt deviations typically occurred when the proctor observed subjects manipulating hardware in an inappropriate manner that was likely to cause damage to the hardware. Additional self-reported metrics included workload, trust, perceived performance, perceived confidence, and perceived procedure helpfulness which were recorded in post-task surveys.

Based on the data collected, analysis of variance (ANOVA) and t-tests were completed to analyze the five different procedure types. Because this was a between subjects experiment, one-way ANOVAs were completed to determine how accuracy, efficiency, workload and trust depended on procedure, while chi-square tests were used to analyze the individual procedural deviation types. Additionally, the analysis of how each individual enhancement affected the above metrics was completed using the enhancement breakdown matrix in Table 3.1. This analysis of the individual enhancements had an increased power compared to the procedure type analysis due to the increased n values as shown in Table 3.3.

Table 3.3: n values for the analysis of each enhancement in the UCD study.

Enhancement	**Enhancement Present n**	**Enhancement Absent n**
Step Navigation	48	12
Enhanced Visuals	24	36
Laser Guidance	24	36
Sensor Feedback	24	36

Because the Feedback Procedure had both laser guidance and sensor feedback, in order to identify how these enhancements affect task performance individually, the steps with sensor feedback and laser guidance were isolated and analyzed separately. During the disassembly of the generator, there were two steps that provided laser guidance and five steps that had real-time sensor feedback.

3.2.3 Task Performance: Efficiency

Efficiency, or task completion time, is the time it took for the subject to complete the manual repair task. Using R, a linear model was used to determine that task completion time was not dependent on procedure type for this n value ($F(4, 55) = 2.28$, $p = 0.073$). A plot of task completion time depending on procedure type is shown in Figure 3.3 with the values outlined in Table 3.4.

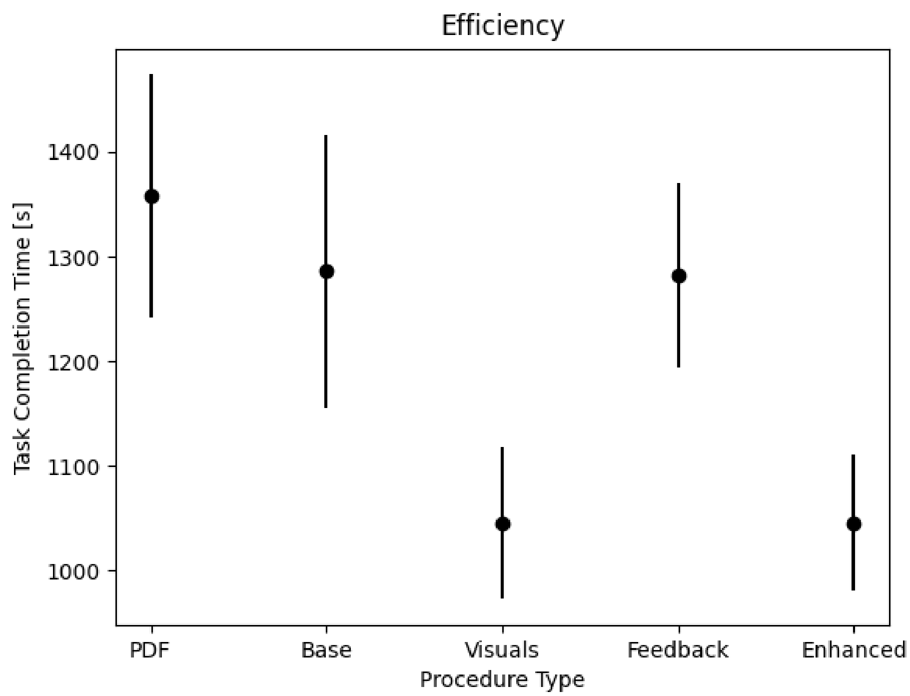


Figure 3.3: Task completion time for the UCD study split up by procedure type.

By using the enhancement breakdown shown in Table 3.1, t-tests were conducted to directly assess how each enhancement affected task efficiency. The task timing depending on each enhancement is outlined in Table 3.5. The t-tests indicate that task completion time is not dependent on the presence of step navigation $t(58) = 1.72$, $p = 0.090$, or sensor feedback $t(58) = 0.705$, p 0.484. Task completion time was however dependent on the presence of enhanced visuals $t(57.7) = 3.31$ $p = 0.002$, with subjects completing the task in 17.4 min (1046 s) and 21.8 min (1309 s) with and without the enhanced visuals respectively.

Table 3.4: Task completion time for the UCD study split up by procedure type.

Procedure	Task Completion Time	Standard Deviation
PDF	22.6 min (1358 s)	6.7 min (401 s)
Base	21.4 min (1286 s)	7.5 min (451 s)
Visuals	17.4 min (1046 s)	4.2 min (252 s)
Feedback	21.4 min (1282 s)	5.1 min (304 s)
Enhanced	17.4 min (1046 s)	3.8 min (226 s)

Table 3.5: Task completion time for the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	19.4 ± 5.5 min	22.6 ± 6.7 min
Enhanced Visuals	17.4 ± 3.9 min	21.8 ± 6.3 min
Sensor Feedback and Laser Guidance	19.4 ± 4.8 min	20.5 ± 6.5 min

Because the Feedback procedure had both the real-time sensor feedback and laser guidance enhancements, further breakdown to individual step timing was necessary to identify the impact of each enhancement individually. There were five steps with real-time sensor feedback. The time spent on each of these steps was identified and coalesced so the step times with and without sensor feedback could be compared without the influence of non-feedback steps. A t-test determined that efficiency was not dependent on the presence of real-time sensor feedback $t(58) = 1.86$, $p = 0.068$. Similarly, there were two steps with laser guidance. A t-test determined that the completion time for these steps was not dependent on the presence of laser guidance $t(58) = 0.541$, $p = 0.591$. This matches the previous results that looked at the combined effects of sensor feedback and laser guidance.

3.2.4 Task Performance: Accuracy

With accuracy defined as the number of correct steps out of the total number of steps, a linear model determined that task execution accuracy was dependent on procedure ($F(4, 55) = 11.8$, $p < 0.001$). The accuracy of each procedure is illustrated in Figure 3.4 and numerically displayed in Table 3.6.

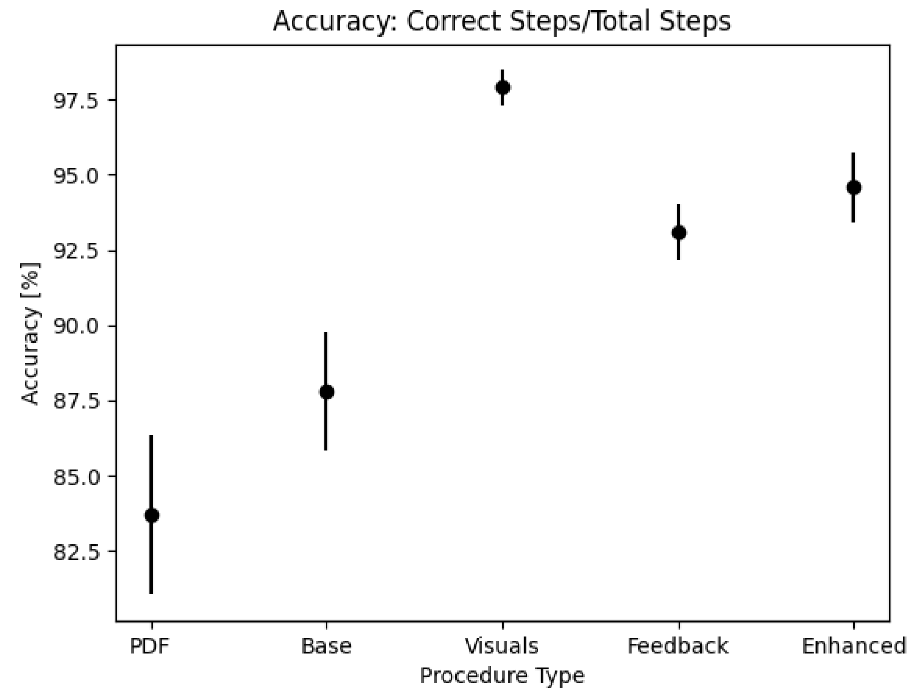


Figure 3.4: Task accuracy for the UCD study split up by procedure type.

Table 3.6: Task accuracy for the UCD study split up by procedure type.

Procedure	Task Accuracy	Standard Deviation
PDF	83.7%	9.21%
Base	87.8%	6.86%
Visuals	97.9%	1.99%
Feedback	93.1%	3.28%
Enhanced	94.6%	4.05%

The post hoc analysis results comparing each procedure using the Tukey correction are displayed in Table 3.7. The results indicate that when subjects completed the task with the Base Procedure they had a lower accuracy than when subjects used the Enhanced Procedure ($p = 0.041$), or the Visuals Procedure ($p < 0.001$). When subjects used the PDF Procedure they had a lower accuracy than when subjects used the Enhanced Procedure ($p < 0.001$), the Feedback Procedure ($p < 0.002$), or the Visuals Procedure ($p < 0.001$). Essentially,

Table 3.7: Post hoc comparisons comparing accuracy for each procedure type in the UCD study.

Post Hoc Comparisons - Procedure

Comparison							
Procedure		Procedure	Difference	SE	t	df	p_{tukey}
Base	-	Enhanced	−6.76	2.33	−2.899	55.0	0.041
Base	-	Feedback	−5.32	2.33	−2.282	55.0	0.166
Base	-	Visuals	−10.15	2.33	−4.354	55.0	<.001
Feedback	-	Enhanced	−1.44	2.33	−0.617	55.0	0.972
PDF	-	Base	−4.10	2.33	−1.760	55.0	0.407
PDF	-	Enhanced	−10.86	2.33	−4.659	55.0	<.001
PDF	-	Feedback	−9.42	2.33	−4.041	55.0	0.002
PDF	-	Visuals	−14.25	2.33	−6.114	55.0	<.001
Visuals	-	Enhanced	3.39	2.33	1.455	55.0	0.595
Visuals	-	Feedback	4.83	2.33	2.073	55.0	0.247

subjects tended to have a lower task accuracy when using either the PDF Procedure or Base Procedure.

To further identify how each enhancement affected accuracy, t-tests were conducted based on the procedures enhancement breakdown in Table 3.1. Task accuracy depending on the presence of each enhancement is outlined in Table 3.8. The effect of step navigation was significant $t(13.1) = -3.47$, $p = 0.004$, as was the effect of enhanced visuals $t(52.7) = -5.42$, $p < 0.001$, and the combined effect of laser guidance and sensor feedback $t(50.2) = -2.41$, $p = 0.019$. These results show that subjects completed the task significantly more accurately when they were provided with either the step navigation or enhanced visuals.

Table 3.8: Task accuracy in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	93.4 ± 5.65 %	83.7 ± 9.21 %
Enhanced Visuals	96.3 ± 3.57 %	88.2 ± 7.75 %
Sensor Feedback and Laser Guidance	93.8 ± 3.68 %	89.8 ± 8.92 %

To identify specifically how laser guidance and sensor feedback affected accuracy independently, the specific steps utilizing each enhancement were isolated. For the five steps with sensor feedback, a t-test determined that sensor feedback did not significantly affect accuracy $t(58) = -0.284$, $p = 0.777$. For the two steps with laser guidance, a t-test similarly determined that laser guidance did not affect accuracy $t(58) = -0.035$, $p = 0.9723$. These results do not match those previously found when the entirety of the task accuracy was analyzed with the two enhancements in combination (Feedback Procedure). This discrepancy is likely due to the fact there are relatively few steps with sensor feedback and laser guidance, so the impact of the enhancements is diluted when looking at the entirety of task accuracy. The more direct analysis involving only relevant steps is a better indication on how the enhancements affect subject performance.

3.2.5 Task Performance: Deviations

To further analyze accuracy, procedural deviations were categorized based off of the PDA methodology previously mentioned (54). Each deviation type was analyzed with chi-square tests to determine if they were dependent on procedure type or any of the enhancements (step navigation, laser guidance, sensor feedback, or enhanced visuals).

Sequential

A breakdown of the average number of sequential deviations depending on procedure type is shown in Figure 3.5 with the values in Table 3.9. A chi-square test was conducted and determined that the number of sequential deviations was dependent on procedure, $\chi^2(4) = 35.2$, $p < 0.001$. Post-hoc comparisons using the Bonferroni correction identified that when subjects completed the task with the PDF Procedure they made significantly more sequential deviations than when subjects used the Base procedure ($p = 0.031$), the Visuals Procedure ($p < 0.001$), the Feedback Procedure ($p = 0.002$), or the Enhanced Procedure ($p = 0.002$).

To further identify how each enhancement individually affected the number of sequential deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of sequential deviations depending on the presence of each enhancement is outlined in Table 3.10. The effect of step navigation was significant, $\chi^2(1) = 32.0$, $p < 0.001$, as was the effect of enhanced visuals, $\chi^2(1) = 12.0$, $p = 0.001$, and the combined effect of laser guidance and sensor feedback, $\chi^2(1) = 5.33$, $p =$

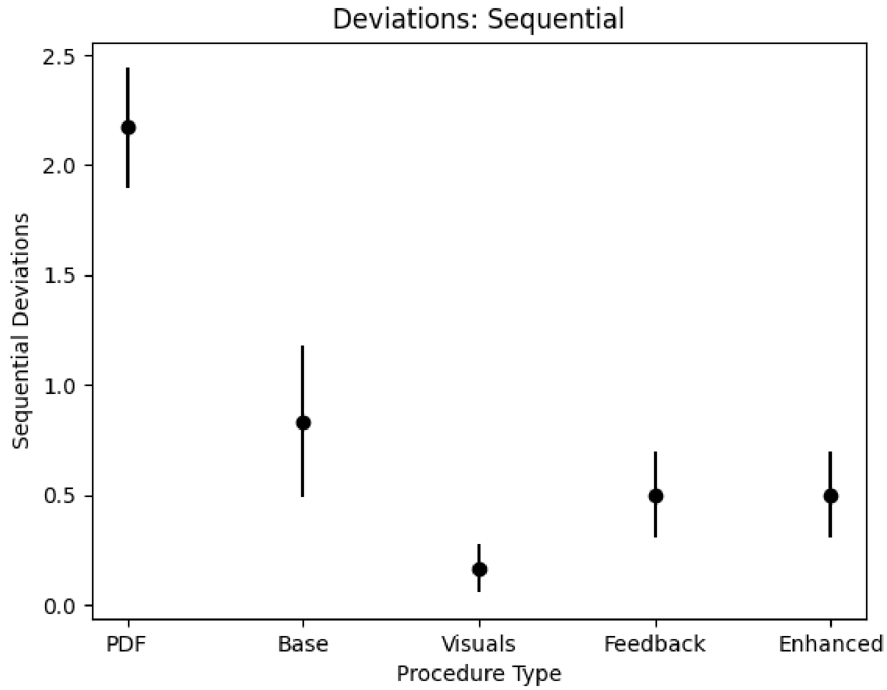


Figure 3.5: Average sequential deviation count per subject in the UCD study split up by procedure type.

Table 3.9: Number of sequential deviations per subject in the UCD study divided by procedure type.

Procedure	Sequential Deviations	Standard Deviation
PDF	2.17	0.937
Base	0.833	1.19
Visuals	0.167	0.389
Feedback	0.500	0.674
Enhanced	0.500	0.674

0.042. The largest effect was seen from the addition of step navigation with subjects making an average of 1.67 fewer sequential deviations when the enhancement was present.

As previously stated, the use of a highlighted step indicator encourages subjects to complete the steps in order by forcing workers to physically acknowledge and click through each step. This form of procedural navigation also discourages "random" scrolling during the task which was only common when subjects used the PDF Procedure. This scrolling action was observed by the study proctor and involved subjects scrolling, sometimes pages, both backwards and forwards throughout the PDF Procedure. Overall, the step highlighter and navigation buttons had a impact on task performance by reducing the number of sequential deviations subjects made. The enhanced visuals and the combined sensor feedback and laser guidance also significantly decreased the number of sequential deviations with an average of 0.837 and 0.56 less sequential deviations with the presence of each enhancement respectively.

Table 3.10: Average sequential deviation count per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.500 ± 0.799	2.17 ± 0.937
Enhanced Visuals	0.333 ± 0.565	1.17 ± 1.18
Sensor Feedback and Laser Guidance	0.500 ± 0.659	1.06 ± 1.22

To further identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of sequential deviations was not dependent on laser guidance, $\chi^2(1) < 0.001$, p = 1, or sensor feedback, $\chi^2(1) = 0.33$, p = 1. This does not match the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance. The more accurate results are from the testing of individual steps because these are not diluted with data from non-relevant steps.

Fragmented

A breakdown of the average number of fragmented deviations depending on procedure type is shown in Figure 3.6 with the values in Table 3.11. A chi-square test was conducted and determined that the number of fragmented deviations was dependent on procedure, $\chi^2(4) = 12.6$, p = 0.014. Post-hoc comparisons identified that the PDF Procedure had significantly more fragmented deviations than the Visuals Procedure (p = 0.011).

To further identify how each enhancement individually affected the number of fragmented deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of fragmented deviations depending on the

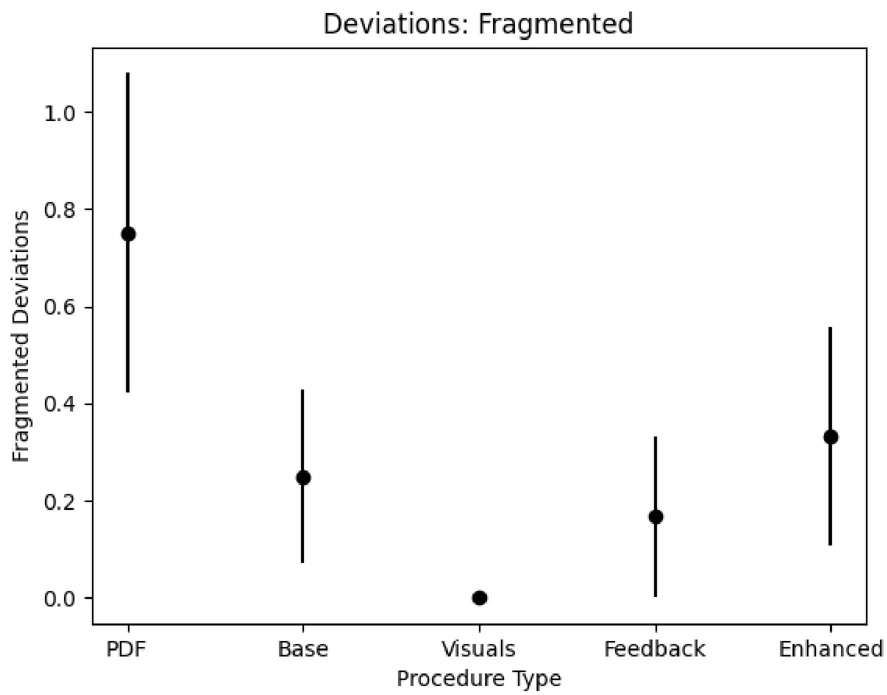


Figure 3.6: Average fragmented deviation count per subject in the UCD study split up by procedure type.

Table 3.11: Number of fragmented deviations per subject in the UCD study divided by procedure type.

Procedure	**Fragmented Deviations**	**Standard Deviation**
PDF	0.750	1.14
Base	0.250	0.622
Visuals	0.00	0.00
Feedback	0.167	0.577
Enhanced	0.333	0.778

presence of each enhancement is outlined in Table 3.12. The effect of step navigation was significant, $\chi^2(1) = 10.1$, $p = 0.003$. The effect of enhanced visuals, $\chi^2(1) = 2.37$, $p = 0.247$, and the combined effect of laser guidance and sensor feedback were not found to be significant, $\chi^2(1) = 0.33$, $p = 1$. On average subjects made 0.562 less fragmented deviations when they used step navigation. This is likely because the step navigation helped subjects complete steps in the correct order by highlighting the current step. Without the step navigation, it was more common for subjects to begin a step at the wrong time before catching their mistake and returning to the current step. Subjects would later finish the step at the right time leading to fragmented deviations.

Table 3.12: Number of fragmented deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.188 ± 0.571	0.750 ± 1.14
Enhanced Visuals	0.167 ± 0.565	0.389 ± 0.838
Sensor Feedback and Laser Guidance	0.250 ± 0.676	0.333 ± 0.793

To more accurately identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of fragmented deviations was not dependent on laser guidance, $\chi^2(1) = 0.042$, p = 1, or sensor feedback, $\chi^2(1) < 0.001$, p = 1. This matches the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance.

Execution

A breakdown of the average number of execution deviations depending on procedure type is shown in Figure 3.7 with the values in Table 3.13. A chi-square test was conducted and determined that the number of execution deviations was dependent on procedure, $\chi^2(4)$ = 16.5, p = 0.002. Post-hoc comparisons identified that the Base Procedure had significantly more execution deviations than the Visuals Procedure (p = 0.031) and the Enhanced Procedure (p = 0.003).

To further identify how each enhancement individually affected the number of execution deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of execution deviations depending on the presence of each enhancement is outlined in Table 3.14. The effect of enhanced visuals was significant, $\chi^2(1) = 13.8$, p < 0.001, however the effect of step navigation, $\chi^2(1) = 0.03$, p = 1, and the combined effect of laser guidance and sensor feedback, $\chi^2(1) = 0.53$, p = 0.933, were not. With a decrease of 0.777 execution deviations, enhanced visuals statistically decreased the

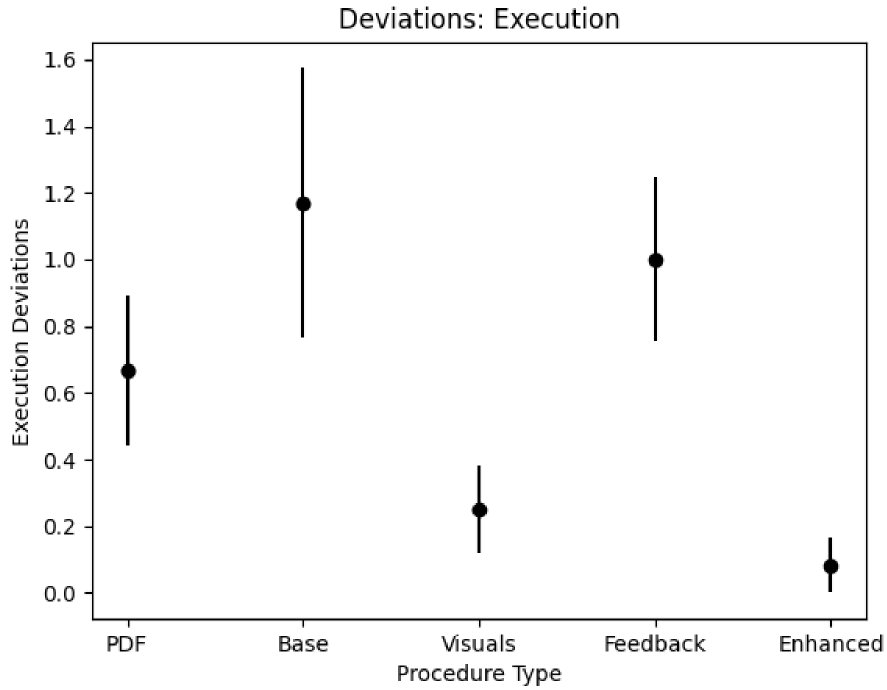


Figure 3.7: Average execution deviation count per subject in the UCD study split up by procedure type.

Table 3.13: Number of execution deviations per subject in the UCD study divided by procedure type.

Procedure	Execution Deviation	Standard Deviation
PDF	0.667	0.778
Base	1.17	1.00
Visuals	0.250	0.00
Feedback	1.00	1.00
Enhanced	0.0833	0.00

number of execution deviations subjects made by more clearly illustrating what needed to be completed in each step.

To more accurately identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of execution deviations was not dependent on laser guidance, $\chi^2(1) < 0.001$, p = 1, or sensor

Table 3.14: Number of execution deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.625 ± 0.959	0.667 ± 0.778
Enhanced Visuals	0.167 ± 0.381	0.944 ± 1.04
Sensor Feedback and Laser Guidance	0.542 ± 0.779	0.694 ± 1.01

feedback, $\chi^2(1) < 0.001$, p = 1. This matches the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance.

Partial Omit

A breakdown of the average number of partial omit deviations depending on procedure type is shown in Figure 3.8 with the values in Table 3.15. A chi-square test was conducted and determined that the number of partial omit deviations was dependent on procedure, $\chi^2(4) = 9.67$, p = 0.046, however post-hoc comparisons did not identify a procedure which had a significantly different number of partial omit deviations than another. On average however, the PDF Procedure and Base Procedure had the largest number of partial omit deviations.

Table 3.15: Number of partial omit deviations per subject in the UCD study divided by procedure type.

Procedure	Partial Omit Deviations	Standard Deviation
PDF	0.417	0.793
Base	0.417	0.669
Visuals	0.0833	0.289
Feedback	0.0833	0.289
Enhanced	0.00	0.00

To further identify how each enhancement individually affected the number of partial omit deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of partial omit deviations depending on the presence of each enhancement is outlined in Table 3.16. The effect of enhanced visuals,

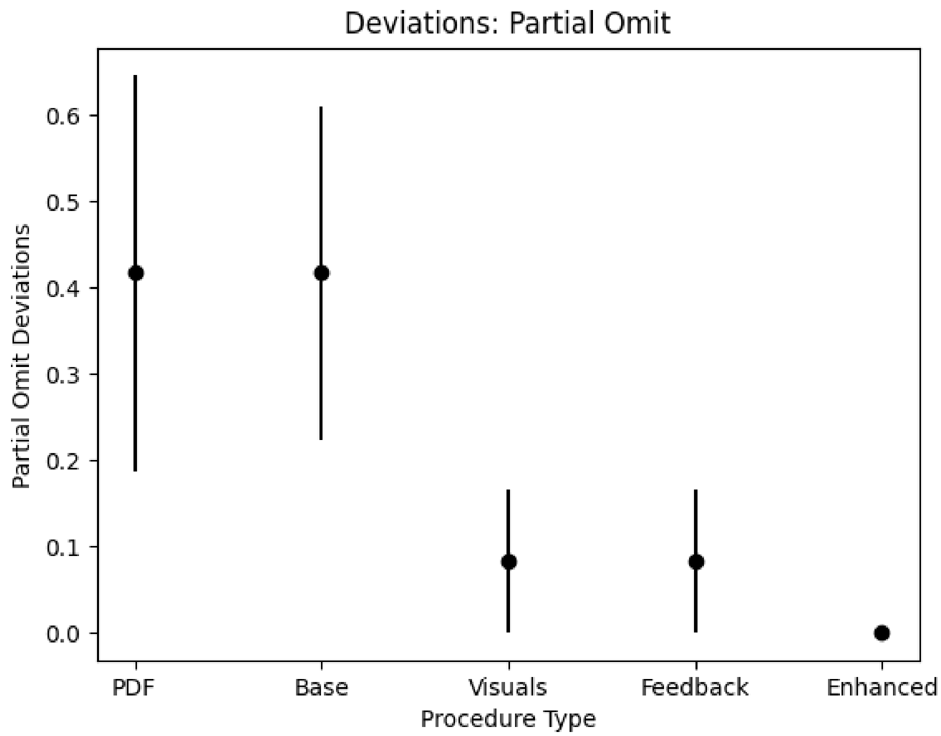


Figure 3.8: Average partial omit deviation count per subject in the UCD study split up by procedure type.

$\chi^2(1) = 5.01$, p = 0.050, and the combined effect of laser guidance and sensor feedback were found to be significant, $\chi^2(1) = 5.01$, p = 0.050, however the effect of step navigation was not, $\chi^2(1) = 3.52$, p = 0.121. On average subjects made 0.264 less partial omit deviations when provided with enhanced visuals.

Table 3.16: Number of partial omit deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.146 ± 0.412	0.417 ± 0.793
Enhanced Visuals	0.0417 ± 0.204	0.306 ± 0.624
Sensor Feedback and Laser Guidance	0.0417 ± 0.204	0.306 ± 0.624

To further analyze the impacts of sensor feedback and laser guidance separately, the steps involved were isolated and further chi-square tests were conducted. Results indicate that laser guidance does not affect the number of partial omit deviations, $\chi^2(1) < 0.001$, p =

1, nor does sensor feedback, $\chi^2(1) < 0.001$, p = 1. This does not match the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance. The more accurate results are from the testing of individual steps because this data is not diluted with data from non-relevant steps.

Omitted

A breakdown of the average number of omitted deviations depending on procedure type is shown in Figure 3.9 with the values in Table 3.17. A chi-square test was conducted and determined that the number of omitted deviations was dependent on procedure, $\chi^2(4) = 14.0$, p = 0.007. Post-hoc comparisons identified that the PDF Procedure resulted in significantly more omitted deviations than the Feedback Procedure (p = 0.046) and Visuals Procedure (p = 0.006).

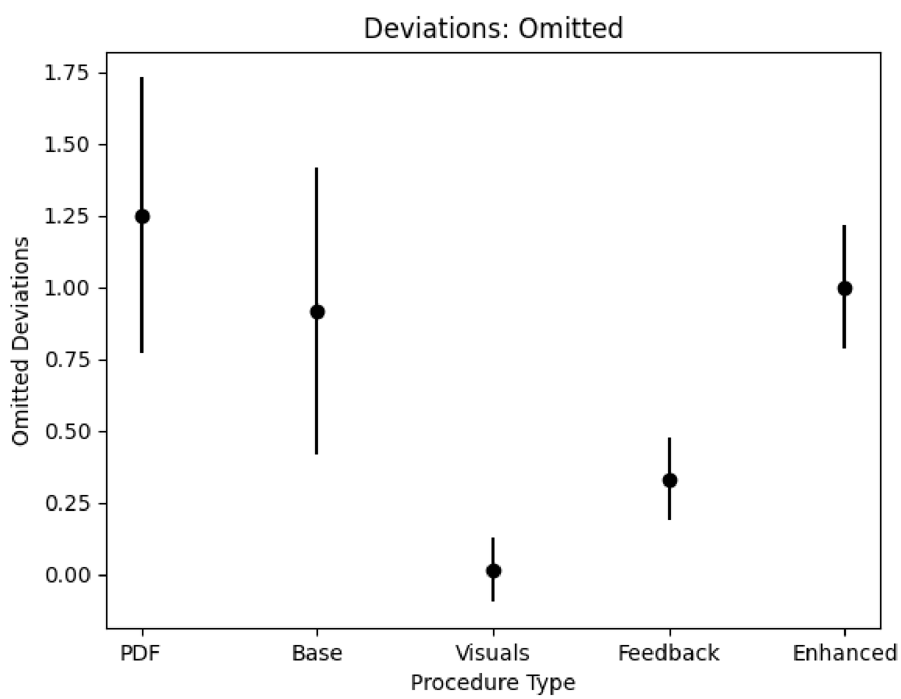


Figure 3.9: Average omitted deviation count per subject in the UCD study split up by procedure type.

To further identify how each enhancement individually affected the number of omitted deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of omitted deviations depending on the presence

Table 3.17: Number of omitted deviations per subject in the UCD study divided by procedure type.

Procedure	Omitted Deviations	Standard Deviation
PDF	1.25	1.66
Base	0.917	1.73
Visuals	0.017	0.389
Feedback	0.333	0.492
Enhanced	1.00	0.739

Table 3.18: Number of omitted deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.604 ± 1.03	1.25 ± 1.66
Enhanced Visuals	0.583 ± 0.717	0.833 ± 1.42
Sensor Feedback and Laser Guidance	0.667 ± 0.702	0.778 ± 1.44

of each enhancement is outlined in Table 3.18. The effect of step navigation was found to be significant, $\chi^2(1) = 5.46$, p = 0.039. The effect of enhanced visuals, $\chi^2(1) = 1.23$, p = 0.536, and the combined effect of laser guidance and sensor feedback were not significant, $\chi^2(1) = 0.24$, p = 1. These results confirm the hypothesis that the step navigation enhancement reduces the number of omitted deviations subjects make during task execution. This is a major finding which greatly supports the usage of step highlighters in all future procedure systems.

To more accurately identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of omitted deviations was not dependent on laser guidance, $\chi^2(1) < 0.001$, p = 1, or sensor feedback, $\chi^2(1) = 0.83$, p = 0.723. This matches the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance.

Extra Action

A breakdown of the average number of extra action deviations depending on procedure type is shown in Figure 3.10 with the values in Table 3.19. A chi-square test was conducted and determined that the number of extra action deviations was dependent on procedure, $\chi^2(4) = 30.1$, $p < 0.001$. Post-hoc comparisons identified that the Base Procedure had significantly more extra action deviations than the Enhanced Procedure ($p = 0.005$), the Feedback Procedure ($p = 0.049$), and the Visuals Procedure ($p < 0.001$) while the PDF Procedure had significantly more extra action deviations than the Enhanced Procedure ($p = 0.019$) and the Visuals Procedure ($p < 0.001$). Overall, the Visuals Procedure and Enhanced Procedure had the lowest number of extra action deviations while the Base Procedure and PDF Procedure had the highest.

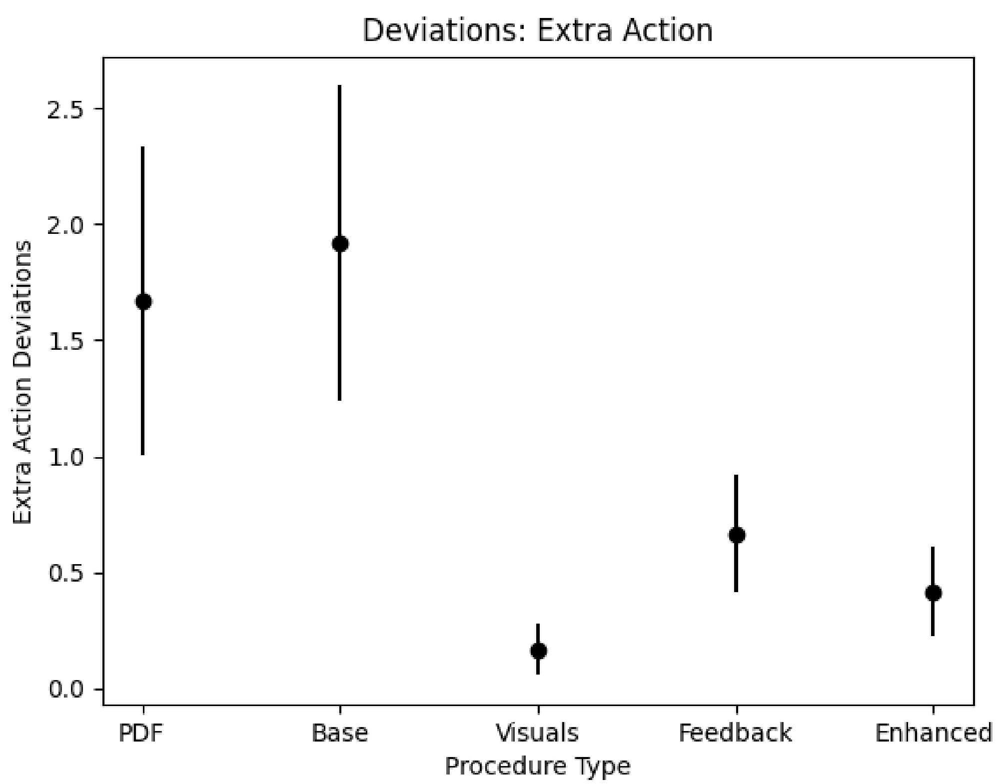


Figure 3.10: Average extra action deviation count per subject in the UCD study split up by procedure type.

To further identify how each enhancement affected the number of extra action deviations

Table 3.19: Number of extra action deviations per subject in the UCD study divided by procedure type.

Procedure	Extra Action Deviations	Standard Deviation
PDF	1.67	2.31
Base	1.92	2.35
Visuals	0.167	0.389
Feedback	0.667	0.888
Enhanced	0.417	0.669

subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of extra action deviations depending on the presence of each enhancement is outlined in Table 3.20. The effects of step navigation $\chi^2(1) = 7.60$, p = 0.012, enhanced visuals $\chi^2(1) = 18.9$, $p < 0.001$, and the combined effect of sensor feedback and laser guidance, $\chi^2(1) = 7.47$, p = 0.013, all significantly impacted the number of extra action deviations. The step navigation and enhanced visuals features decreased the number of extra action deviations by 0.88 and 1.13 respectively.

Table 3.20: Number of extra action deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.792 ± 1.44	1.67 ± 2.31
Enhanced Visuals	0.292 ± 0.550	1.42 ± 1.99
Sensor Feedback and Laser Guidance	0.542 ± 0.779	1.25 ± 2.02

To further understand the individual impacts of the laser guidance and sensor feedback, relevant steps were isolated and further chi-square tests were performed. Neither the laser guidance, $\chi^2(1) = 2.00$, p = 0.315, or the sensor feedback, $\chi^2(1) = 0.22$, p = 1, were found to statistically reduce the number of extra action deviations. While these results do not match those of the previous chi-square test, the previous test analyzed the entire procedure while these only included relevant steps. As such, this secondary analysis is a more accurate representation of the enhancements' effects.

Interrupt

A breakdown of the average number of interrupt deviations depending on procedure type is shown in Figure 3.11 with the values in Table 3.21. Subjects were only interrupted by the proctor during the task if they were manipulating hardware in such a way that the equipment was likely to break. A chi-square test was conducted and determined that the number of interrupt deviations was dependent on procedure, $\chi^2(4) = 10.9$, p = 0.027, however post-hoc comparisons did not identify any procedure that was significantly different from another. Despite this, the Visuals Procedure and Enhanced Procedure had the least number of interrupt deviations present while the PDF Procedure and Feedback Procedure had the most.

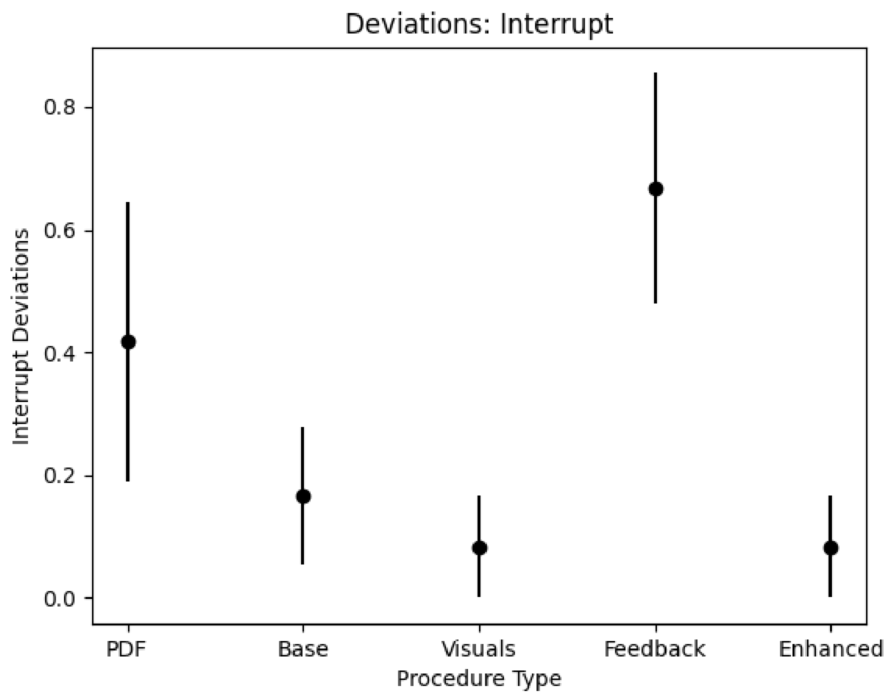


Figure 3.11: Average interrupt deviation count per subject in the UCD study split up by procedure type.

To further identify how each enhancement individually affected the number of interrupt deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of interrupt deviations depending on the presence of each enhancement is outlined in Table 3.22. The effect of enhanced visuals, $\chi^2(1) = 5.65$,

Table 3.21: Number of interrupt deviations per subject in the UCD study divided by procedure type.

Procedure	Interrupt Deviations	Standard Deviation
PDF	0.417	0.793
Base	0.167	0.389
Visuals	0.0833	0.289
Feedback	0.667	0.651
Enhanced	0.0833	0.289

$p = 0.035$, was found to be significant while the effect of step navigation, $\chi^2(1) = 0.94$, $p = 0.664$, and the combined effect of sensor feedback and laser guidance, $\chi^2(1) = 1.19$, $p = 0.552$, were not significant. When provided with enhanced visuals, subjects made on average 0.34 less interrupt deviations. This data confirms that the enhanced visuals allowed subjects to complete the task more effectively, reducing the chances of them incorrectly manipulating equipment in an inappropriate manner.

Table 3.22: Number of interrupt deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.250 ± 0.484	0.417 ± 0.793
Enhanced Visuals	0.0833 ± 0.282	0.417 ± 0.649
Sensor Feedback and Laser Guidance	0.375 ± 0.576	0.222 ± 0.540

To more accurately identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of interrupt deviations was not dependent on laser guidance, $\chi^2(1) < 0.001$, $p = 1$, or sensor feedback, $\chi^2(1) < 0.001$, $p = 1$. While these results do not match those of the previous chi-square test, the previous test analyzed the entire procedure while these only included relevant steps. As such, this secondary analysis is a more accurate representation of the enhancements' effects.

Assist

A breakdown of the average number of assist deviations depending on procedure type is shown in Figure 3.12 with the values in Table 3.23. Assist deviations occurred when subjects requested assistance relating to task completion from the study proctor. A chi-square test was conducted and determined that the number of assist deviations was not dependent on procedure, $\chi^2(4) = 4.31$, p = 0.366. While not statistically significant, the Visuals Procedure and Enhanced Procedure had the lowest number of assist deviations present.

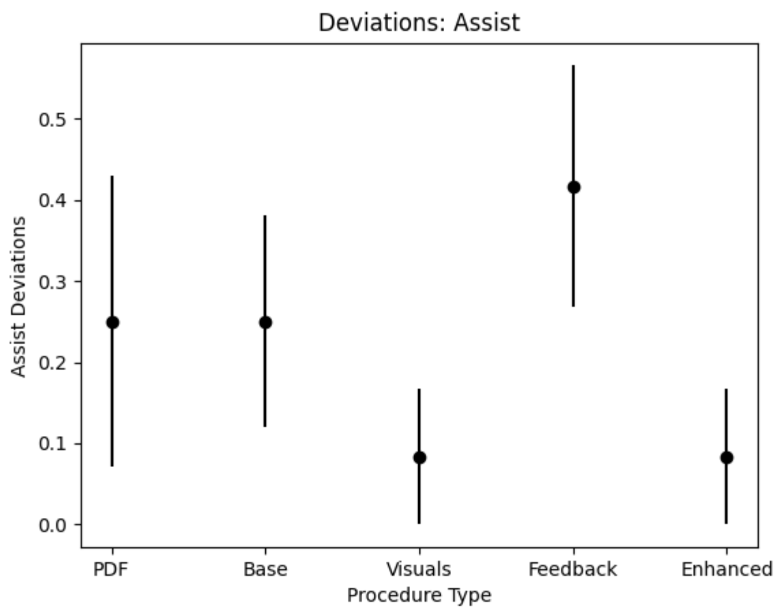


Figure 3.12: Average assist deviation count per subject in the UCD study split up by procedure type.

To further identify how each enhancement individually affected the number of assist deviations subjects made, chi-square tests were conducted based on the procedures enhancement breakdown in Table 3.1. The number of assist deviations depending on the presence of each enhancement is outlined in Table 3.24. The effect of step navigation $\chi^2(1) = 0.08$, p = 1, enhanced visuals $\chi^2(1) = 3.28$, p = 0.140, and the combined effect of sensor feedback and laser guidance $\chi^2(1) = 0.21$, p = 1, did not significantly impact the number of assist deviations. While the enhanced visuals had the greatest impact in reducing the number of assist deviations, the enhancement did not statistically impact the results at this n value.

Table 3.23: Number of assist deviations per subject in the UCD study divided by procedure type.

Procedure	Assist Deviations	Standard Deviation
PDF	0.250	0.622
Base	0.250	0.452
Visuals	0.0833	0.289
Feedback	0.417	0.515
Enhanced	0.0833	0.289

Table 3.24: Number of assist deviations per subject in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	0.208 ± 0.410	0.250 ± 0.622
Enhanced Visuals	0.0833 ± 0.282	0.306 ± 0.525
Sensor Feedback and Laser Guidance	0.250 ± 0.442	0.194 ± 0.467

To more accurately identify the individual impacts of sensor feedback and laser guidance, the steps involved were isolated and chi-square tests were performed. The number of assist deviations was not dependent on laser guidance, $\chi^2(1) = 0.89$, p = 0.692, or sensor feedback, $\chi^2(1) = 0.67$, p = 0.828. This matches the previous results from the chi-square test that analyzed the combined sensor feedback and laser guidance.

3.2.6 Subjective Performance

Workload

Workload was assessed through the NASA-TLX (12); subjects completed the survey after completion of the repair task. The results are illustrated in Figure 3.13 and are numerically displayed in Table 3.25 with the scale going from 1-20: 1 indicating a minimal workload and 20 indicating an extensive workload. On average, subjects who used the PDF Procedure had the highest perceived workload while subjects who used the Visuals Procedure had the lowest. A Kruskal-Wallis test however determined that workload was not dependent on procedure for this n value, $\chi^2(4) = 7.00$, p = 0.136.

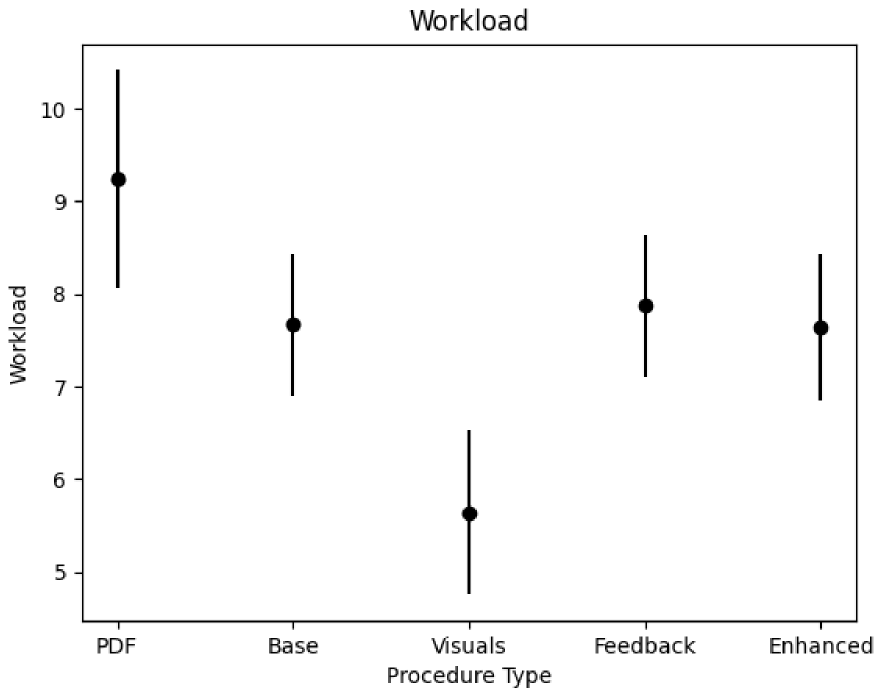


Figure 3.13: Subjective workload in the UCD study determined by the NASA-TLX split up by procedure type.

Table 3.25: Subjective workload in the UCD study determined by the NASA-TLX split up by procedure type.

Procedure	Workload	Standard Deviation
PDF	9.24	4.08
Base	7.67	2.66
Visuals	5.64	3.08
Feedback	7.77	2.64
Enhanced	7.17	2.73

Additional Mann-Whittney U tests were completed to identify the effects of each individual enhancement on the subjects' reported workload. While the step navigation had an effect on workload W = 371, p = 0.048, the enhanced visuals W = 560, p = 0.054, and combined sensor feedback and laser guidance W = 404.5, p = 0.684, did not have a significant effect. The workload values depending on the presence of each enhancement are outlined in Figure 3.26. On average, when subjects completed the procedure with step navigation, subjects reported a statistically lower workload with a reduction of 2.0. This indicates that the step navigation allowed subjects to complete the task with less effort as the enhancement removed some of the subjects' mental burden of keeping track of their place in the procedure.

Table 3.26: Subject workload in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	7.21 ± 2.84	9.24 ± 4.08
Enhanced Visuals	6.64 ± 3.03	8.26 ± 3.19
Sensor Feedback and Laser Guidance	7.76 ± 2.63	7.52 ± 3.56

Trust

After completing the task, subjects were asked to complete a trust survey to analyze their trust in the procedure system (15). The results are shown in Figure 3.14 with a scale ranging from 1 (High distrust) - 7 (High trust). The numerical values are also shown in Table 3.27. A Kruskal-Wallis test determined that the level of trust in the system was not dependent on procedure $\chi^2(4) = 7.21$, p = 0.125. On average though, subjects had the least trust in the PDF Procedure and the most trust in the Enhanced Procedure.

Table 3.27: Subjects' self-rated trust in the procedure system from the UCD study split up by procedure type.

Procedure	Trust	Standard Deviation
PDF	5.48	1.32
Base	5.90	0.900
Visuals	6.22	0.734
Feedback	5.75	0.788
Enhanced	6.40	0.706

Further Mann-Whitney U tests were completed to identify how each individual enhancement affected trust. The numerical values depending on each enhancement are shown in Table 3.28. Results indicate that subjects had a greater trust in the system when provided

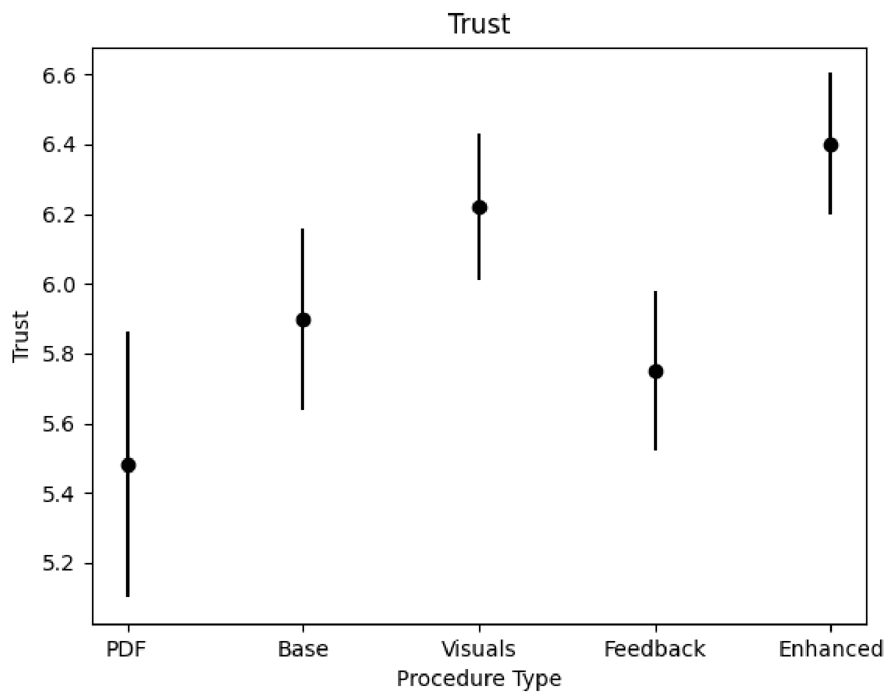


Figure 3.14: Results of the trust survey from the UCD study for each of the procedure types. The scale is continuous from 1-7 with higher values indicating a higher level of trust in the system.

with enhanced visuals W = 264, p = 0.011. Step navigation W = 217.5, p = 0.195, and the combined effects of laser guidance and sensor feedback W = 395.5, p = 0.587, did not change subjects' level of trust in the system. These results indicate that subjects trusted the procedure more when they were provided with enhanced visuals.

Table 3.28: Subjects' trust in the procedure system from the UCD study depending on the presence of different enhancements.

Enhancement	**Enhancement Present**	**Enhancement Absent**
Step Navigation	6.07 ± 0.803	5.48 ± 1.32
Enhanced Visuals	6.31 ± 0.710	5.71 ± 1.01
Sensor Feedback and Laser Guidance	6.07 ± 1.03	5.87 ± 0.803

3.2.7 Post-Task Survey Responses

After completing the task, subjects completed a post-task survey gathering their opinions on the procedure they used and how they felt when completing the task. Subjects were specifically asked to rate their performance on the task, how confident they felt when completing the task, and how helpful they found the procedure. They were additionally asked how helpful they found each enhancement and how potentially helpful each enhancement could be outside the context of the generator repair task.

Perceived Performance

After completing the task, subjects were asked to self-rate their performance on a scale from 1 (Performed very poorly) - 7 (Performed very well). The results are shown in Figure 3.15 with the values shown in Table 3.29. A Kruskal-Wallis test determined that subjects' perceived performance was not dependent on procedure type $\chi^2(4) = 7.81$, p = 0.099. Despite this, subjects on average rated their performance lower when using the PDF Procedure and Feedback Procedure.

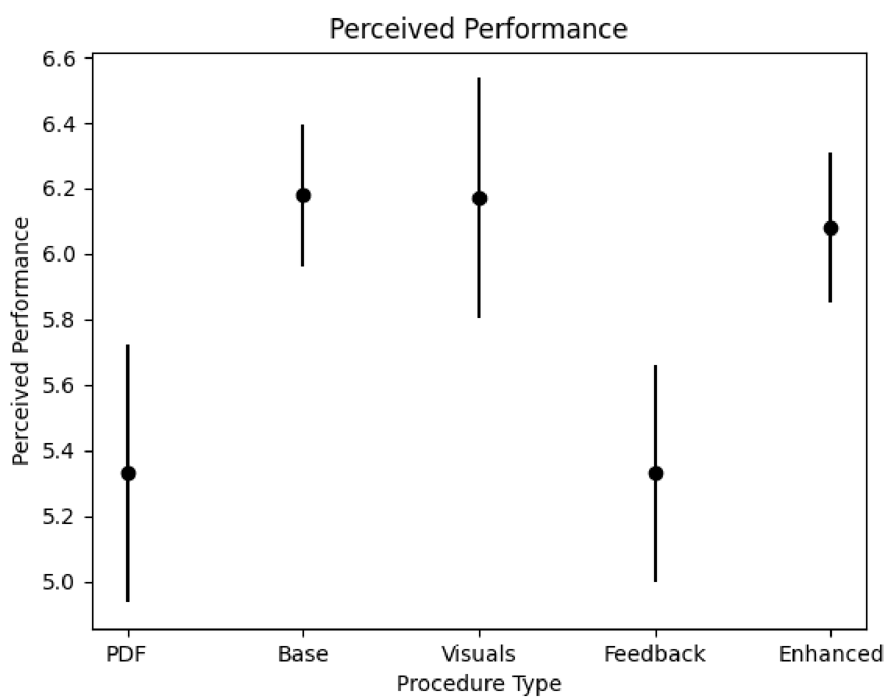


Figure 3.15: Subjects' self-rated task performance from the UCD study for each procedure type on a scale from 1 (Very poor) - 7 (Very well).

Table 3.29: Subjects' self-rated task performance from the UCD study for each procedure type on a scale from 1 (Very poor) - 7 (Very well).

Procedure	Perceived Performance	Standard Deviation
PDF	5.33	1.37
Base	6.18	0.751
Visuals	6.17	1.27
Feedback	5.33	1.15
Enhanced	6.08	0.793

Table 3.30: Subjects' perceived performance in the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	5.94 ± 1.05	5.33 ± 1.37
Enhanced Visuals	6.13 ± 1.03	5.60 ± 1.17
Sensor Feedback and Laser Guidance	5.71 ± 1.04	5.88 ± 1.21

Further Mann-Whitney U tests were completed to identify how each individual enhancement affected subjects' perceived performance. The numerical values depending on the presence of each enhancement are shown in Table 3.30. Results indicate that subjects' perceived performance was not dependent on step navigation W = 205.5, p = 0.134, enhanced visuals W = 303, p = 0.060, or the combined effect of sensor feedback and laser guidance W = 477, p = 0.362. Essentially, none of the enhancements or procedure types statistically changed how subjects felt they performed on the task.

Perceived Confidence During Task Execution

After finishing the procedure, subjects were asked to self-rate how confident they were while completing the task on a scale from 1 (Very low confidence) - 7 (Very high confidence). The results are shown in Figure 3.16 with the values tabulated in Table 3.31. A Kruskal-Wallis test determined that subjects' perceived confidence was not dependent on procedure type $\chi^2(4) = 6.01$, p = 0.199. On average though, subjects reported the highest level of confidence when working with the Visuals Procedure and the lowest confidence when working with the PDF Procedure.

Further Mann-Whitney U tests were completed to identify how each individual enhancement affected subjects' perceived confidence during task execution. The numerical values depending on the presence of each enhancement are shown in Table 3.32. Results indicate that subjects' perceived performance was not dependent on step navigation W = 204.5, p = 0.110, enhanced visuals W = 318, p = 0.074, or the combined effect of sensor feedback and laser guidance W = 469, p = 0.566. Essentially, none of the enhancements or procedure types statistically improved how confident subjects felt when completing the task.

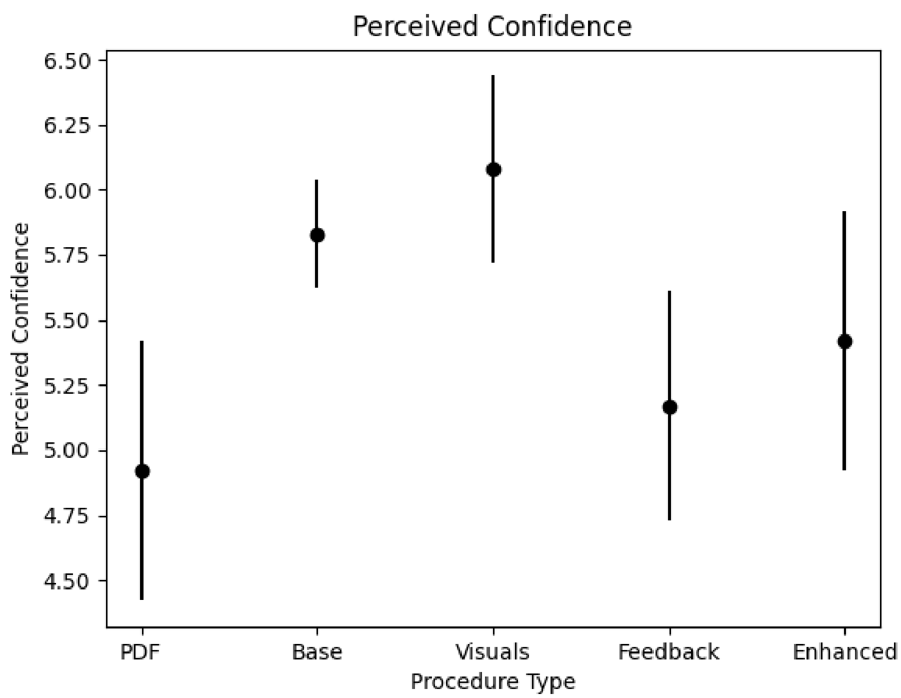


Figure 3.16: Subjects' self-rated confidence during task execution of the UCD study for each procedure type on a scale from 1 (Very low) - 7 (Very high).

Perceived Procedure Helpfulness

After finishing the task, subjects were asked to self-rate how helpful they found the procedure during the task. The results are shown in Figure 3.17 on a scale from 1 (Not useful) - 7 (Very useful) with the values tabulated in Table 3.33. A Kruskal-Wallis test determined that subjects' perceived procedure helpfulness was not dependent on procedure type $\chi^2(4) = 7.94$, p = 0.094). While not significant, subjects on average found the PDF

Table 3.31: Subjects' self-rated confidence during task execution of the UCD study for each procedure type on a scale from 1 (Very low) - 7 (Very high).

Procedure	Perceived Confidence	Standard Deviation
PDF	4.92	1.73
Base	5.83	0.718
Visuals	6.08	1.24
Feedback	5.17	1.53
Enhanced	5.42	1.73

Table 3.32: Subjects' self-rated confidence during task execution of the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	5.63 ± 1.36	4.92 ± 1.73
Enhanced Visuals	5.75 ± 1.51	5.31 ± 1.41
Sensor Feedback and Laser Guidance	5.29 ± 1.60	5.61 ± 1.36

Procedure the least helpful and the Visuals Procedure the most helpful.

Table 3.33: Subjects' rated procedure helpfulness of each procedure type for the UCD study on a scale from 1 (Not useful) - 7 (Very useful).

Procedure	Procedure Helpfulness	Standard Deviation
PDF	5.00	2.22
Base	6.17	0.577
Visuals	6.75	0.452
Feedback	6.33	0.492
Enhanced	6.08	1.68

Further Mann-Whitney U tests were completed to identify how each individual enhancement affected subjects' perceived procedure helpfulness. The numerical values depending on the presence of each enhancement are shown in Table 3.34. Results indicate that subjects' perceived procedure helpfulness was dependent on enhanced visuals $W = 289.5$, $p = 0.019$,

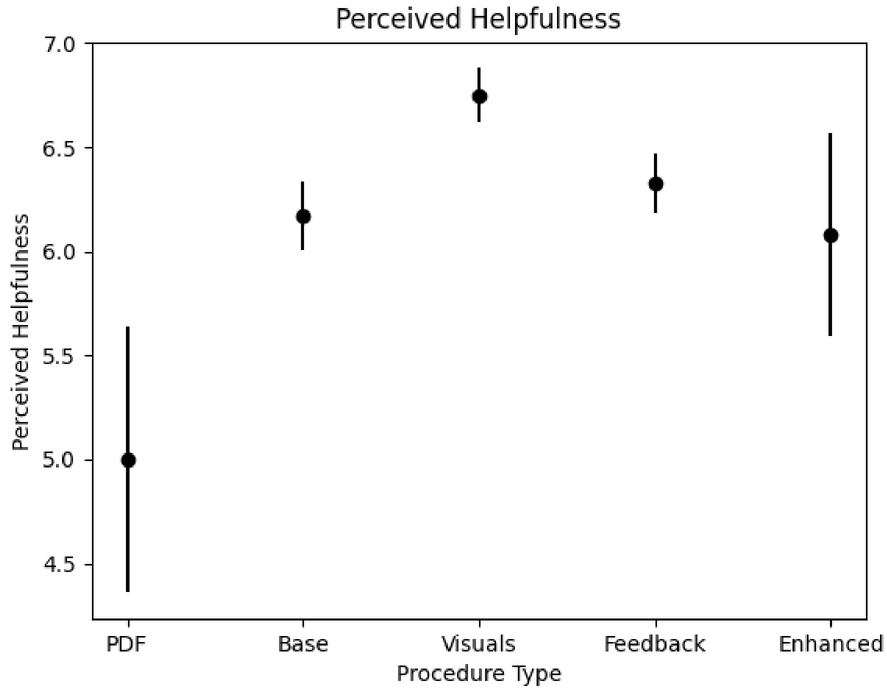


Figure 3.17: Subjects' rated procedure helpfulness of each procedure type for the UCD study on a scale from 1 (Not useful) - 7 (Very useful).

but not dependent on the presence of step navigation W = 205.5, p = 0.098, or the combined effect of sensor feedback and laser guidance W = 419.5, p = 0.843). These results suggest that subjects found the procedure statistically more helpful when they were provided with enhanced visuals.

Table 3.34: Subjects' rated procedure helpfulness during the UCD study depending on the presence of different enhancements.

Enhancement	Enhancement Present	Enhancement Absent
Step Navigation	6.33 ± 0.953	5.00 ± 2.22
Enhanced Visuals	6.42 ± 1.25	5.83 ± 1.44
Sensor Feedback and Laser Guidance	6.21 ± 1.22	5.97 ± 1.50

Enhancements

In the post-task survey, subjects were asked to rate on a scale from 1 (Not useful) - 7 (Very useful) how helpful they found each of the enhancements they worked with. The results are shown in Figure 3.18. In order of most helpful to least, the enhancements were: sensor feedback, videos, enhanced photos, laser guidance, and step navigation.

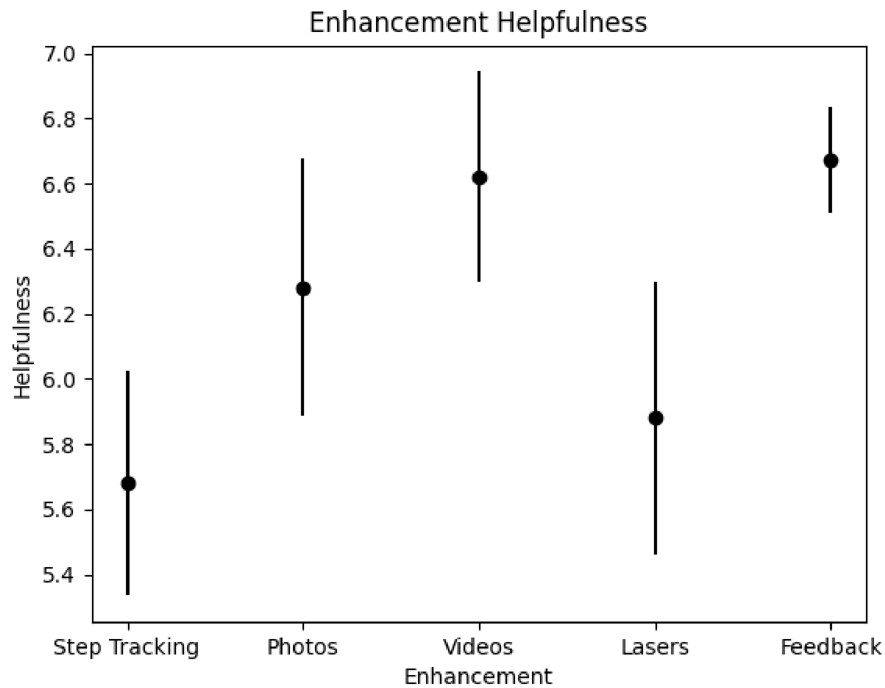


Figure 3.18: Reported enhancement helpfulness in the UCD study on a scale from 1 (Not useful) - 7 (Very useful).

Subjects were similarly asked how potentially helpful they believed each enhancement could be on a scale from 1 (Not useful) - 7 (Very useful) with the results shown in Figure 3.19. This question varied from the previous in that it asked subjects to consider the potential benefits of each enhancement outside of the context of this specific generator repair task. This is valuable because while the generator repair task is a complex manual task, the task is still relatively straight forward due to its contained environment with limited tools to choose from. Future spacecraft repair tasks will likely require more tools and involve the manipulation of increasingly complex machinery which could potentially make certain enhancements like the laser guidance and enhanced videos more helpful. The main difference between the results for the helpfulness and the potential helpfulness questions was seen for the photos and lasers with subjects believing these enhancements could potentially be more helpful in a context outside of the generator repair procedure.

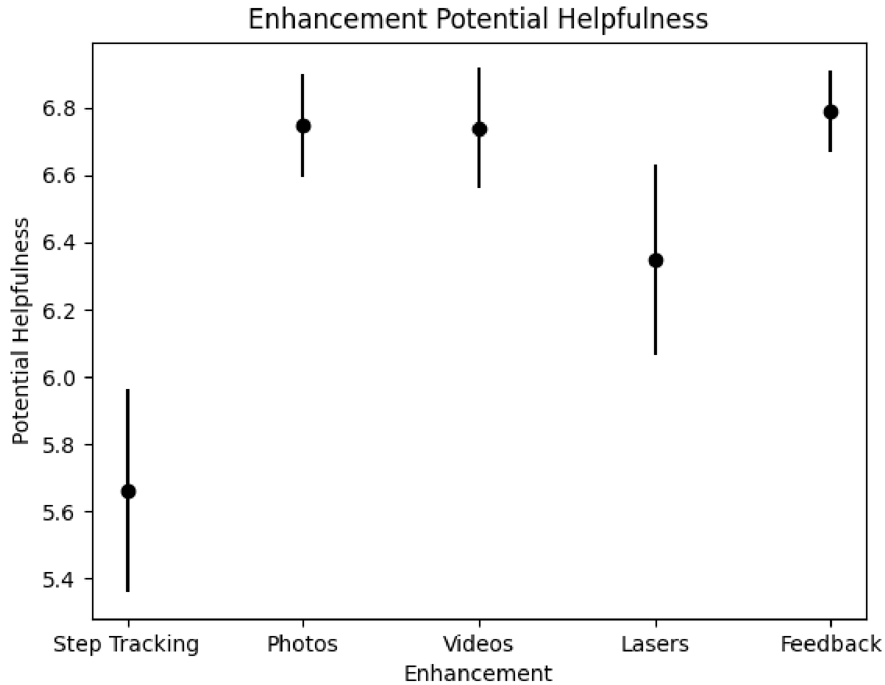


Figure 3.19: Reported potential helpfulness for each enhancement in the UCD study on a scale from 1 (Not useful) - 7 (Very useful).

Free Response

In the post-task survey subjects were given multiple opportunities to answer free response questions such as: "How was the procedure helpful?" and "How could the procedure be more helpful?". Questions were open ended allowing for subjects to elaborate on anything that stuck out as important to them. These responses were organized and tallied so trends could be observed.

To begin with, photos were regularly mentioned; subjects provided with the enhanced visuals regularly noted that they found the pictures helpful for identifying parts of interest. Specifically, 7/12 subjects using the Visuals Procedure and 4/12 subjects using the Enhanced Procedure mentioned that they found the photos helpful when completing the task. Subjects using the base visuals still found the basic images helpful with 8/12 subjects using the PDF Procedure, 7/12 subjects using the Base Procedure, and 5/12 subjects using the Feedback Procedure specifically noting that the provided pictures were useful.

Despite the base visuals being noted as helpful, 9/12 subjects using the PDF Procedure, 10/12 subjects using the Base Procedure and 7/12 subjects using the Feedback Procedure noted that they would like more photos to be included. For these three procedures (which utilized the base images), subjects were only provided with necessary photos. Once parts were identified in an image, subjects were expected to remember the information and photos were not repeated. The large number of subjects requesting additional images indicates that subjects did not always remember the previously provided information. In fact, 2 subjects specifically wrote that they had to return to past steps in order to review old photos which wastes time and increases the chances of subjects losing their place in the procedure. Procedure writers should not assume that workers remember past information and procedures should provide all relevant step information whenever possible to reduce confusion and prevent subjects from having to review past steps.

9 subjects additionally mentioned that they appreciated being able to zoom in on pictures to prevent eye strain and better study the images. This is a simple feature to add and one that should be included in all electronic procedures. 5 subjects wrote that they liked being able to see the next step on the screen while they completed the current step because it gave them an idea of where the procedure was heading. This is an interesting observation because while some future information can be helpful to see where the task is moving, too much information can be distracting and detract attention away from the current step. User feedback and visual observation from the proctor of subjects scrolling through the procedure suggest that it is optimal to have the current step and one future step visible on the screen during procedure execution (with the current step highlighted). By including one future step on the screen, workers are provided with "before and after" information which allows them to focus on the current step while also satisfying their natural curiosity of what is coming up next.

Regarding videos, 4/12 subjects using the Visuals Procedure and 5/12 subjects using the Enhanced Procedure specifically mentioned that they found the videos helpful. Videos were included for complex steps and user feedback confirmed that workers appreciated having the recordings to better illustrate how to complete the steps. 3 subjects who were not provided with the enhanced visuals (1 subject who used the PDF Procedure and 2 subjects who used the Base Procedure) mentioned that they would like videos added into the procedure.

This feedback from both those that used the videos and those that believed it would be a helpful addition confirms that videos can be a powerful tool to demonstrate how to complete complex or critical steps. As CAD modeling continues to improve, gifs showing hardware manipulation could begin to replace traditional videos of experts completing steps. Videos are an excellent method to visually illustrate how to complete steps but they can have issues with lighting or items blocked from the camera view. Future enhanced visuals will likely include a combination of enhanced images, real-life videos, and CAD videos generated by software which highlight relevant parts and show how they should be manipulated.

The step navigation was also consistently mentioned in subjects' free responses. For those provided with the step navigation feature, 5/12 subjects using the Base Procedure, 4/12 subjects using the Visuals Procedure, 5/12 subjects using the Feedback Procedure, and 1/12 subjects using the Enhanced Procedure noted that they found the step highlighter and navigation arrows helpful. Subjects reported that it helped them stay on track with the step highlighter helping prevent them from losing their place. Subjects consistently emphasized that the feature helped them complete the procedure "step by step" and some noted that they liked how the navigation buttons jumped the procedure page to focus on the current step. Conversely, there were 4 subjects who disliked the step navigation system (1 used the Visuals Procedure, 2 used the Feedback Procedure, and 1 used the Enhanced Procedure). Feedback indicated that they found the system annoying and would have preferred to simply scroll through the task. Two of the subjects who disliked the system conceded that they found it annoying because they kept forgetting to use the navigation arrows and then had to jump forward to their current step. Other subjects' feedback indicated they simply did not believe that the feature would have any benefits. So while a small portion of subjects did not find the system intuitive or helpful, the majority of subjects felt either positive or neutral about the feature. 1 subject who used the PDF Procedure, and was therefore not provided with the step navigation enhancement, even requested for a similar feature to be included so they could check off steps and keep track of their progress.

Out of the 24 subjects who were provided with real-time sensor feedback, 7/12 subjects who used the Feedback Procedure and 6/12 subjects who used the Enhanced Procedure noted positive sentiments in their free responses regarding the sensor feedback. Subjects consistently stated that they felt more confident about their step completion after the sensors confirmed they completed the steps correctly. The sensor feedback made them feel "satisfied" and more comfortable moving on in the procedure knowing that they were "moving in the right direction". 6 subjects (2 subjects using the Feedback Procedure and 4 subjects using the Enhanced Procedure) noted that they felt there should be more sensor feedback in the procedure. This is further positive feedback indicating that subjects want to see more of the feature and supports the continued study of how sensor feedback can improve task execution. As previously mentioned, this procedure had a limited number of sensor feedback steps because of difficulties embedding sensors into the COTS generator so future experiments can expand on this study by designing a hardware system that can accommodate more embedded sensors.

Similarly to the sensor feedback, of the 24 subjects provided with laser guidance, 5/12 subjects who used the Feedback Procedure and 2/12 subjects who used the Enhanced Procedure mentioned in their free responses that they liked the laser guidance. Subjects reported that it helped them rapidly identify the bolts which needed to be removed, saving them time and increasing their confidence that they were removing the correct hardware. While values were not recorded, it was also incredibly common after the task for subjects to verbally mention to the proctor how fun the lasers were to use.

12 subjects reported in their free responses that they believed both the sensor feedback and laser guidance would be more helpful for more complex tasks. Similar to the sensor feedback, there were again limited steps which supported the use of laser guidance, so despite having only two steps using the feature, there was a notable amount of positive feedback regarding the enhancement. The free responses which noted the benefits of the laser guidance, combined with the overwhelming number of subjects who reported enjoying the enhancement, support the continued exploration of laser guidance as well as other tools which can quickly highlight objects of interest.

Lastly, 4 subjects noted in their free responses that they would like to see the addition of audio features. One suggestion included having the procedure read the current step's instructions aloud so that subjects did not need to look away from the hardware system. Additionally, the use of voice recognition for procedure navigation was suggested. Rather than physically clicking a button, workers could instead state “Next” or “Previous” to navigate throughout the task, allowing them to keep their hands free and preventing them from having to put down tools. These are interesting suggestions which should be considered for future multimodal electronic procedures.

3.3 Discussion

The goal of the Davis study was to identify how each enhancement individually affected task performance. Beginning with the enhanced visuals, the feature notably decreased task completion time with subjects completing the task an average of 4.4 min faster when provided with the enhancement. This is likely because the enhanced pictures with colored overlays and videos for complex steps provided additional information compared to the base images, allowing subjects to more rapidly familiarize themselves with the hardware and understand what they were required to do. By consistently providing and displaying all relevant step information on the screen through enhanced visuals, subjects found it easier to efficiently complete the task becuase they did not have to remember or review information from past steps.

The enhanced visuals also statistically increased how accurately subjects completed the task. Looking at how each enhancement effected the different deviation types in Table 3.35, the enhanced visuals significantly reduced the number of sequential, execution, partial omit, extra action, and interrupt deviations. The enhanced visuals are the only enhancement which reduced the number of execution deviations. While many of the deviation types are associated with subjects completing steps out of order, execution deviations are caused by incorrectly completing a step. Execution deviations are the closest deviation type to traditional “errors” and are one of the most likely deviation types to significantly impact procedure success. This study suggests that the added information presented through enhanced images can statistically reduce these execution deviations allowing subjects to complete their task more accurately.

Similarly, the enhanced visuals are the only enhancement that significantly decreased the number of interrupt deviations. Both the assist deviation and interrupt deviation relate directly to crew autonomy. While the assist deviation is associated with subjects requesting

Table 3.35: Breakdown of how the different enhancements decreased each deviation type in the UCD study.

	Sequential	Fragmented	Execution	Partial Omit	Omitted	Extra Action	Interrupt	Assist
Step Navigation	X	X			X	X		
Enhanced Visuals	X		X	X		X	X	
Sensors and Lasers	X			X		X		
Sensors								
Lasers								

assistance, the interrupt deviation was initiated by the proctor when they saw hardware being manipulated inappropriately. This type of interruption is similar to how MCC currently observes astronauts from video feed on the ground and interjects when they see mistakes. This oversight will be drastically reduced in future LDEMs as time delays increase, so this data which indicates interrupt deviations can be statistically reduced with enhanced visuals is a major observation. Essentially, providing enhanced visuals increases subjects awareness of the task enough to increase subjects ability to safely and autonomously complete the task. Additionally, subjects reported the videos and enhanced images to be the second and third most helpful enhancements indicating that subjects recognized their benefits.

Analysis also determined that the enhanced visuals significantly increased subjects' trust in the procedure system, as well as how helpful they felt the procedure was (Table 3.36). Subjects trusted the procedure system more when it provided enhanced visuals, likely because it reliably displayed all the information they needed. Essentially, the enhanced images improved almost all aspects of task performance by decreasing task completion time, increasing task accuracy, and increasing their trust in the system. As such, future procedures should include enhanced visuals for each step to better orient workers to the hardware they are working with and illustrate how steps should be completed. While including an image for every step may seem redundant because it sometimes repeats information previously provided, it is better to have all relevant step information visible on the screen and have subjects not need it, rather than wrongly assume subjects will remember past information. If subjects do not know how to complete a step and the necessary information is not provided, they will be forced to search back through the procedure or guess what the step is referencing, two scenarios that increase workers' chances of making mistakes.

Continuing to look at accuracy, both the enhanced visuals and step navigation enhance-

Table 3.36: Breakdown of how the different enhancements improved subjects' subjective opinions in the UCD study.

	Workload	**Trust**	**Perceived Performance**	**Confidence**	**Helpfulness**
Step Navigation	X				
Enhanced Visuals		X			X
Sensors and Lasers					

ments statistically increased subjects' task completion accuracy. With an increase of 9.7% and 8.1% accuracy with the step navigation and enhanced visuals respectively, these enhancements made a significant impact on subjects' task performance.

Looking at the different deviations in Table 3.35, step navigation statistically reduced the number of sequential, fragmented, omitted, and extra action deviations. All of these deviations are associated with either missing steps or completing steps out of order which supports the conclusion that step navigation helps workers "stay on track". It also discourages workers from scrolling through the procedure to review past steps or study future ones. This feature essentially limits the information subjects are provided at any given time to the relevant information displayed on screen for the current step. By limiting the available information to only what is currently necessary, workers are able to focus their attention on each individual step, removing distractions and increasing accuracy.

The step navigation also significantly decreased subjects' perceived workload, likely because the limited information on screen removed distractions and the highlighter prevented subjects from having to re-find where they were in the procedure each time they looked away. The step highlighter is a small but powerful tool that can be incorporated into any electronic procedure to keep workers focused and ensure they are completing steps in the correct order. While the IPV currently used aboard the ISS does have step highlighter capabilities, astronauts are not required to use the feature. The results of this study lead us to the recommendation that all future astronaut procedures should have step highlighting, and that astronauts should be strongly encouraged to use the feature. Despite positive free response feedback, the step navigation was consistently rated the least helpful enhancement in the Davis study. So while workers may not necessarily feel that it is as beneficial as the other enhancements, the results of this study prove how major its benefits are. While subjects reported a significantly lower workload when using the step navigation, because step navigation was consistently perceived as the least helpful enhancement, workers might need to be required to use the feature to ensure they are reaping its benefits. It is common for many people to have high self confidence, believing they will not miss steps or complete steps out of order, however this is a common occurrence in procedural execution and therefore needs to be addressed.

The real-time sensor feedback was rated as the most useful enhancement in the Davis study, however it did not statistically improve the efficiency or accuracy with which subjects completed the task. Subjects reported enjoying the feature and noted that it made them more confident in their step completion. Similarly, the laser guidance had no significant impact on subjects' task performance despite positive feedback. To better understand the effects of real-time sensor feedback and laser guidance, a task that allows for more steps to utilize the enhancements would greatly improve the power of future studies. For this task, it was difficult to find many locations where sensors could be embedded in to the COTS generator as the generator design minimizes free space inside the machine. There were only five steps that provided sensor feedback and two steps with laser guidance, so another study with more of this feedback could help to further identify both enhancements' benefits. This would likely involve developing a new task with a unique hardware system built in-house that could have sensors embedded into the system from the design stage. By building a hardware system from scratch, voids could be specially designed into the system to hold sensors and route wires through. This would also be more similar to how future spacecraft will be built and designed with monitoring sensors in mind.

Chapter 4

HERA Analog Experiment

4.1 Experimental Design

To test our EPV in a spacecraft analog environment, we participated in NASA HERA Campaign 6. The Human Exploration Research Analog (HERA) is a habitat (shown in Figure 4.1) located at the NASA Johnson Space Center (JSC) used as an analog for isolation, confinement, and remote conditions in exploration scenarios (55). During a single mission, 4 crewmembers spend 45 days simulating a space journey by living in the habitat and completing relevant tasks and experiments under observation. One HERA Campaign comprises 4 missions, totaling 16 crewmembers per campaign.

Figure 4.1: HERA, the Human Exploration Research Analog, is a three-story, closed habitat at NASA's Johnson Space Center used to simulate long-duration human spaceflight missions (55).

Figure 4.2: Experimental setup in HERA; the EPV is displayed on the tablet while the crewmember uses the tools on the right to perform the mechanical repair task on the generator.

For our HERA experiment we had crewmembers complete the float inspection task using the modified Honda generator system. The experiment setup in HERA is shown in Figure 4.2. Because the crew was in an isolated environment, the subjects both disassembled the generator and reassembled it after checking the float. Two types of procedures were compared:

1. Traditional, non-interactive procedures (the Base Procedure), and

2. Multimodal, interactive procedures (the Enhanced Procedure)

Both the Base Procedure and Enhanced Procedure are web-based procedures with identical wording for each of the steps. Step navigation was also included for both procedures because the Raspberry Pi step navigation system records whenever a subject clicks one of the navigation arrows to change steps, allowing for automatic step-timing recording. Since the crewmembers were on a simulated exploration mission, real-time observation of the experiment was not possible so this automatic time recording was beneficial for data collection. A breakdown of the different enhancements for both the Base and Enhanced Procedures is shown in Table 4.1.

Note that while the Base Procedure does not have the Enhanced Visuals feature, the procedure does have basic pictures. Some of these images are from the original generator manual which are not colored and identify numerous parts per picture as shown in Figure 3.1. While

Table 4.1: HERA procedural enhancements breakdown for both procedure types.

	Step Navigation	Sensor Feedback	Enhanced Visuals	Laser Guidance
Base	X			
Enhanced	X	X	X	X

the Base Procedure only has the step navigation enhancement, the Enhanced Procedure has all of the EPVs' multimodal features including step navigation, sensor feedback, enhanced visuals, and laser guidance.

This was a within-subjects experiment with all crewmembers completing the task with both procedure types during the course of their mission. There were two groups: one group began the experiment with the Base Procedure while the other started with the Enhanced Procedure. To determine the correlation between power and sample size for our study, a power calculation was completed using R, the results of which are shown in Figure 4.3.

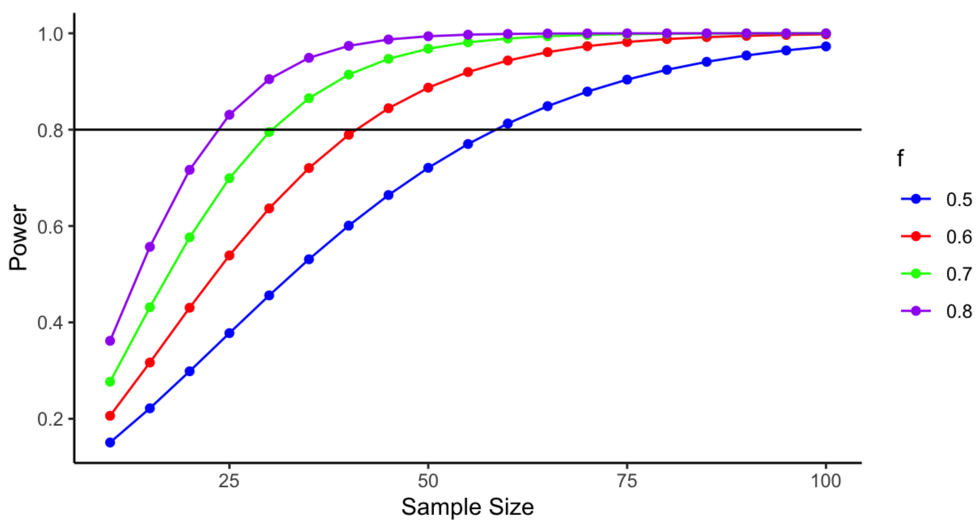


Figure 4.3: Power analysis to determine the correlation between total sample size and effect size.

Assuming the largest effect size of f = 0.8, to achieve a power of 0.8, the study would require a total sample size of 25. Conversely, if we were to assume a small effect size of f = 0.5, the total sample size would need to be around 60 to achieve a power of 0.8. Due to the limitations of HERA, each 4 mission campaign has a total of 16 crewmembers resulting in n = 16. To increase the power of the study, each crewmember was tasked to complete the task 8 times, 4 with the Enhanced Procedure and 4 with the Base Procedure. Therefore, the total sample size was increased to n = 128 with n = 64 for each procedure type. This schedule is illustrated in Figure 4.4.

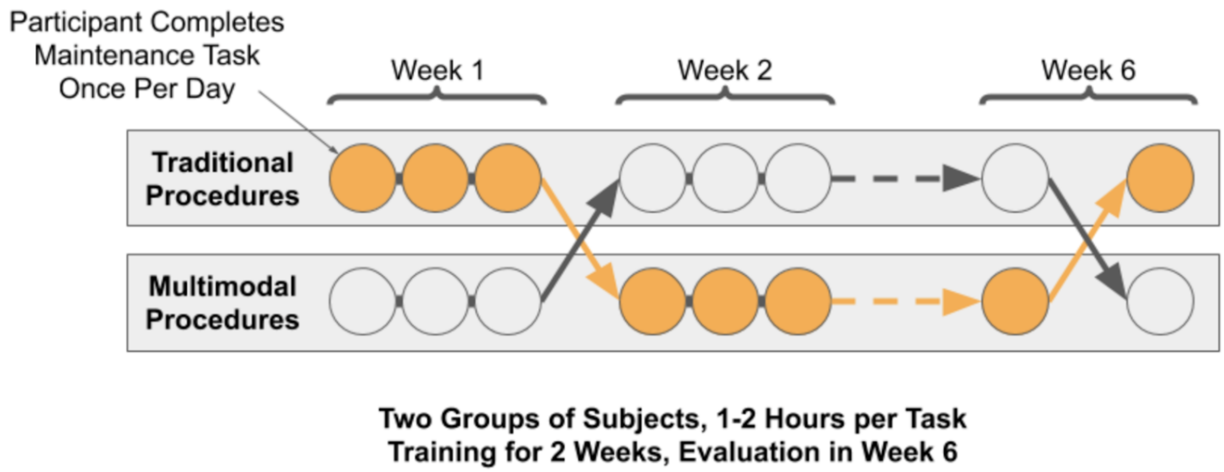


Figure 4.4: HERA experiment schedule breakdown.

The subjects were randomly split into two groups, Group 1 started with the Base Procedure while Group 2 started with the Enhanced Procedure. Each crewmember completed the task three times their first week on separate days (Trials 1-3) with their assigned procedure. In the second week all crewmembers switched to the other procedure type and completed the task three more times (Trials 4-6), again never completing the task more than once per day. The crew then took a month long break before finally completing the task two more times (Trials 7-8), once with each procedure type. This schedule was chosen so that each crewmember would interact with both the Base and Enhanced Procedures. By scheduling the experiment so the crew interacted with both procedure types, each subject could provide feedback on the differences between the Base and Enhanced Procedures and the study could further identify how task performance changed between them. The final two trials after the break were included so the procedure effectiveness could be analyzed in the context of refresher training.

Crewmembers were asked to complete numerous surveys throughout the experiment. Prior to mission ingress, subjects signed a consent form and completed a demographic survey asking questions about their background (gender, age, education), aptitude for using technologies (computer self-efficacy), and aptitude for mechanical repair/maintenance. Subjects then watched a training video which introduced them to the EPV, the tools they would be using during the task, and the overall goal of the manual repair task: to locate and check the float valve tip. Subjects were encouraged to ask questions at this time and reminded of key points such as reading every step and consistently using the step navigation.

In mission, after the completion of each trial, subjects were given three post-task surveys:

1. NASA TLX survey - to assess their cognitive workload
2. Trust survey - to assess their trust in the procedure system
3. Post-Task survey - to collect their thoughts on the procedure system and their task performance

Finally, after the mission was completed, crewmembers completed a Post-Mission survey and had an individual 10 minute debrief with study personnel to collect their feedback on the different procedures.

The study's primary independent variables are subject, trial and procedure type. The dependent variables for human performance are efficiency, accuracy, and workload. We will also be considering the subject's level of trust in the system and how they perceived their performance to be with each procedure type.

4.1.1 Hypotheses

We hypothesized that when subjects used the Enhanced Procedure, compared to the Base Procedure participants would:

1. Complete the task at the same rate; while the enhancements might increase situational awareness, they might also require extra time to interact with.
2. Complete the task with greater accuracy due to the enhancements providing additional relevant task information.
3. Report a decreased subjective workload as the enhancements reduce the difficulty of locating parts and increase confidence by confirming adequate step completion.
4. Report the same level of trust in the system.

4.2 Experimental Results

4.2.1 Participants

NASA recruited 16 subjects (8 males, 8 females) from the general population for the HERA C6 campaign. Recruited subjects were between the ages of 30-55 and required to have an advanced degree in a STEM field so they would represent the astronaut population. All subjects gave informed consent in accordance with the NASA Institutional Review Board and subjects were compensated for their time at an hourly rate. Their average age was 38 $\pm$ 6.43 (mean $\pm$ sd) years overall (38.6 $\pm$ 6.28 years for males, 37.4 $\pm$ 6.95 years for females). Of the 16 subjects, 16 were very familiar with operating a laptop, 9 were mechanics or had experience with manual repair tasks, and 9 used mechanical hand tools often.

4.2.2 Analysis

Because HERA is inherently an isolated and confined environment, real-time observation of the study is impossible. During task execution, the Raspberry Pi microcomputer recorded all sensor data and EPV navigation (usage of the up and down arrow step navigation buttons) to provide a basic time log. Crew were also tasked with manually recording the time at certain predetermined intervals as a backup precaution. These recorded real-time metrics however do not adequately monitor the accuracy of task completion or account for lapses in the usage of the step navigation. To remedy this, video analysis was completed for each trial to record accurate step timing and accuracy. This video analysis was completed by a single individual to ensure consistency in the analysis process and was completed in the same manner as the video analysis from the Davis experiment.

To record step time, the reviewer made note of the beginning time for each individual step. The end of one step therefore corresponded to the beginning of the following step which ensured the entirety of the task completion time was recorded. Because the EPV had a navigation system, steps were defined as "starting" when the subject clicked the "Next" button, thereby highlighting the "new" current step in question. For instances when subjects were not consistently using the provided navigation system, steps were defined as starting when subjects finished the previous step and redirected their attention back to the procedure tablet. If a subject completed multiple steps without checking the procedure tablet, the current step was defined by what the subject was actively touching or focusing their attention on. All time spent on the task (time from beginning step 1 to clicking the "End Trial" button) was accounted for unless the subject actively left the generator and the camera's field of view. This did occur in several instances when subjects left the area to request other crewmembers' assistance. In this scenario, task timing was paused until the subject re-entered the camera's field of view. Time when the subject was outside of the camera's view was not included as there was no way to determine what the subject was doing in this time period.

To assess task accuracy, the accuracy of each step was considered. Any deviation from the procedure was noted and classified according to the previously mentioned PDA methodology (54). For the purposes of this study, we also added our own Deviation type called "assist" for whenever a subject asked for assistance in order to complete the task. In training we discouraged subjects from interacting with others while completing the task, but they did occasionally ask MCC or other crewmembers for assistance. Self-reported metrics including workload, trust, perceived performance, perceived confidence, and perceived procedure helpfulness were recorded in post-task surveys.

Based on the data collected, analysis of variance (ANOVA) tests were completed to analyze the independent variables (Procedure and Trial). For continuous dependent variables (efficiency, accuracy, workload and trust), a general linear mixed model was used to assess the statistical differences in performance resulting from the use of the two types of procedures. A mixed model was used to account for how crewmembers were completing the task multiple times, with Subject as a repeated measure. When analyzing the "count" occurrences of the different types of deviations, a chi-square test was completed with a Poisson mixed model.

For the following data, the base procedure has n = 63 and the enhanced procedure has n = 63. There was meant to be a total of n = 128 trials, however 2 trials were not video recorded and therefore could not be analyzed to confirm timing and accuracy. There were also compliance issues during the campaign where crewmembers completed the task with the wrong procedure type (subjects did not follow the posted schedule). An overview of the as-run schedule in HERA C6 is shown in Table 4.2 outlining when subjects used each procedure type. These compliance issues were most notable early in the missions when the crew were getting used to the HERA task scheduling system. The largest instance of this is in trial 1 where the Base Procedure had n = 11 while the Enhanced Procedure had n = 5. This discrepancy is unfortunate because the most significant performance difference between the Enhanced and Base Procedures was observed during these early trials when subjects were completing the task for the first time.

Table 4.2: Breakdown of n values for each procedure type and trial in HERA C6.

Trial Number	Enhanced Procedure	Base Procedure
Trial 1	5	11
Trial 2	7	9
Trial 3	8	7
Trial 4	9	7
Trial 5	9	7
Trial 6	9	6
Trial 7	8	8
Trial 8	8	8
Total n	63	63

Further compliance issues were seen when subjects occasionally failed to complete part, or all, of the 3 post-task surveys. Trials in which subjects did not fully complete the surveys are still included for efficiency and accuracy analysis, but resulted in each of the surveys having different n values depending on who completed them. The impacts of these compliance issues with the surveys are further discussed in the results section.

Another issue that impacted the effectiveness of this experiment was the reliability of the embedded sensors for the real-time sensor feedback. While the generators were fully checked prior to each mission, in mission observation and maintenance of the generators was impossible due to the isolated nature of HERA experiments. Sensor anomalies could only be observed by the crew onboard and while they were asked to report any malfunctions, in practice this was uncommon. This resulted in some of the missions having inconsistent and unreliable sensor feedback. Post-mission analysis determined that at some point, one sensor malfunctioned in Mission 1 and three sensors malfunctioned in Mission 3. As will be discussed further in the results section, this unreliable sensor feedback greatly impacted the impression crewmembers had of the enhancement and the Enhanced Procedure.

Further, it should be noted that in Mission 3, around halfway through the experiment, a hardware piece was dropped and lost. The piece in question was a carburetor gasket that was not necessary for successful completion of the task, but was involved in two separate steps. The crewmember who lost the part did not report it and while the other crewmembers also looked for it in later trials, the gasket was not located until after the mission ended. Because no one reported the missing piece to MCC, it was not replaced. To not negatively skew the data or penalize the crewmembers for failing to complete the two steps involving the gasket, once the gasket was lost, the following trials assume adequate step completion of the two steps in question. This issue did not statistically impact the study conclusions.

4.2.3 Task Performance: Efficiency

For this analysis, efficiency, or task completion time, is defined as the time it took for the subject to complete the manual repair task. Using R, a linear mixed model was used to determine if task completion time was dependent on either procedure or trial with one within variable, subject. The plot of total task completion time broken up by trial, procedure, and group is shown in Figure 4.5. As previously mentioned, Group 1 started with the Base Procedure while Group 2 started with the Enhanced Procedure.

The linear mixed model determined there was an interaction effect between procedure and trial ($F(7, 97.0) = 3.43$, $p = 0.003$). Post hoc tests with the Bonferroni correction show that in trial 1, the subjects' efficiency was dependent on procedure ($p < 0.001$) with subjects completing the task faster with the Enhanced Procedure. It should be noted that in trial 1, due to subject compliance issues, the base procedure had $n = 11$ while the enhanced procedure had $n = 5$. Despite this, in trial 1 it took crewmembers an average of 44.9 ± 11.2 min to complete the task with the Base Procedure compared to 32.4 ± 7.3 min with the Enhanced Procedure. This is a significant difference with subjects using the Enhanced Procedure saving 12.5 minutes. For the main effect of procedure, there was no statistical effect on subjects' task completion time ($F(1, 96.8) = 2.11$, $p = 0.150$). It took crewmembers an average of 22.3 min (1340s) to complete the task with the Base Procedure compared to 19.2

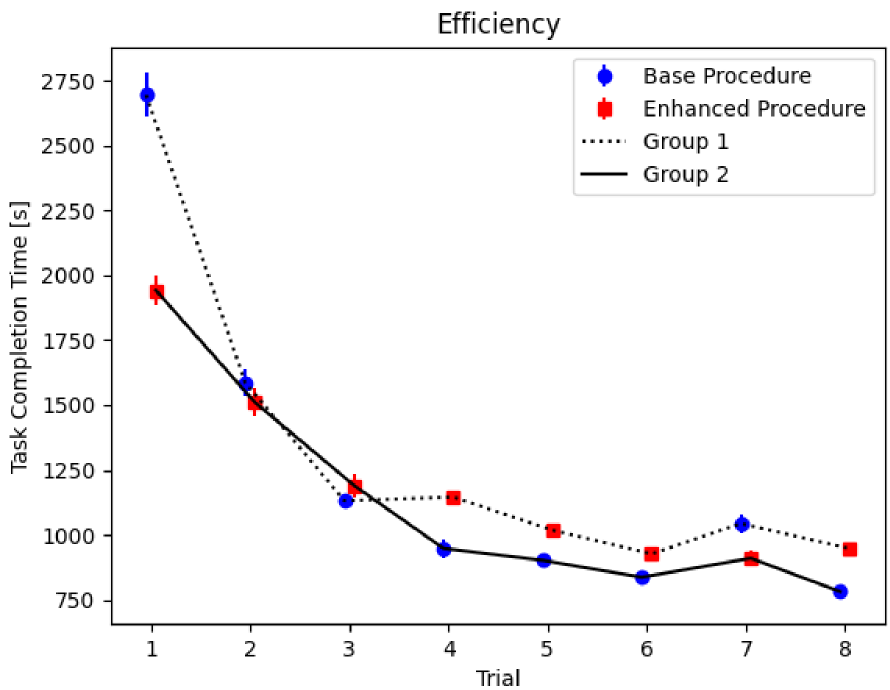


Figure 4.5: Task completion time of HERA C6 split up by trial, group, and procedure type.

min (1150s) with the Enhanced Procedure. Task completion time was however dependent on the main effect of trial ($F(7, 96.3) = 61.4$, $p < 0.001$); specifically trial 1 and trial 2 were completed slower than all other trials ($p < 0.001$), and trial 3 was slower than trial 5 ($p = 0.043$), trial 6 ($p = 0.007$), and trial 8 ($p = 0.002$).

This increased efficiency depending on procedure type in trial 1 was not anticipated in our hypothesis because the enhancements add more features for the workers to interact with. While laser guidance and enhanced images may improve workers' abilities to locate objects of interest, other features such as the videos and sensor feedback require more time for subjects to observe and interpret. Despite this potential increase in time required to interact with the enhancements, the benefits of the features allowed subjects to much more quickly become familiar with the new procedure and generator, thereby allowing them to more quickly complete the task. This efficiency increase does not remain in later trials however when the crew is already familiar with the procedure. This indicates that the Enhanced Procedure might be most beneficial for novel tasks, just in time training, or tasks that astronauts have not reviewed for a long period.

4.2.4 Task Performance: Accuracy

Accuracy in this analysis is defined as the total number of correct steps out of the total number of steps used to complete the task. While there is a specific required number of steps to complete the task, subjects typically took more steps than this due to procedural deviations. The accuracy for each trial split by goup and procedure type is illustrated in Figure 4.6. A linear mixed model was used to determine if task accuracy was dependent

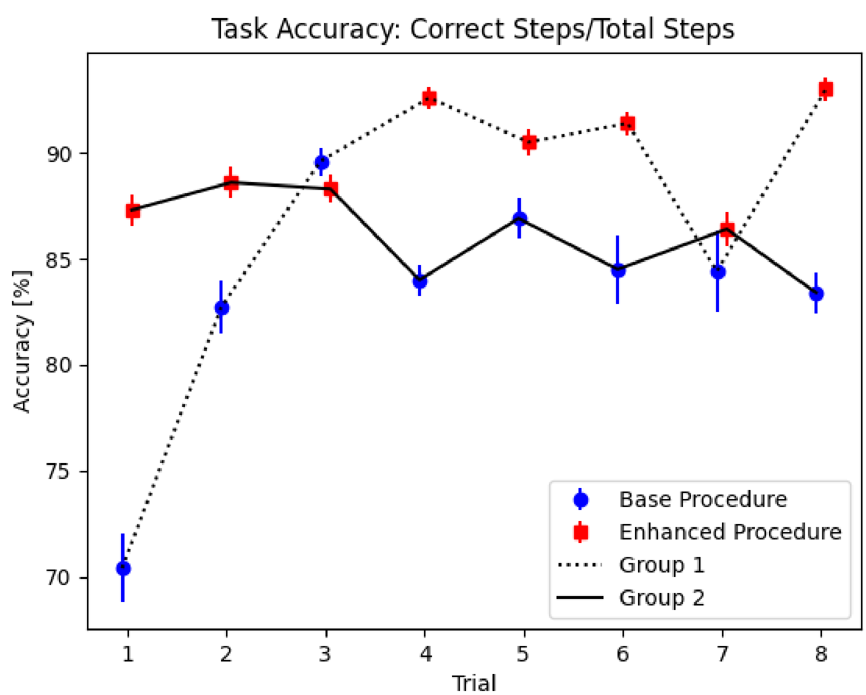


Figure 4.6: Task accuracy (correct steps/total steps) for HERA C6 split up by trial, group, and procedure type.

on two between variables (procedure and trial) and one within variable (subject). Analysis determined that the interaction effect between procedure and trial was not significant ($F(7, 95.0) = 1.63$, $p = 0.137$). Accuracy was dependent on the main effect of procedure ($F(1, 96.9) = 29.74$, $p < 0.001$) and trial ($F(7, 96.2) = 3.84$, $p < 0.001$). Specifically, subjects had a statistically better accuracy when they used the Enhanced Procedure. On average subjects completed the task with $82.4 \pm 11.6\%$ accuracy when using the Base Procedure, as compared to the Enhanced Procedure where subjects achieved an average accuracy of $90.0 \pm 5.4\%$. This is a difference of around 7.6%. Post hoc comparisons with the Bonferroni correction identified that the accuracy for trial 1 was statistically lower than the accuracy for trial 3 ($p = 0.003$), trial 4 ($p = 0.003$), trial 5 ($p = 0.002$), trial 6 ($p = 0.002$), and trial 8 ($p = 0.004$). Further comparisons exploring the interaction effects between procedure and trial indicate there was a difference in accuracy dependent on procedure for trial 1 ($p = 0.014$). Specifically, in trial 1 the accuracy was significantly lower when subjects used the Base Procedure ($70.4 \pm 13.1\%$) compared to the Enhanced Procedure ($87.3 \pm 6.0\%$).

This 16.9% difference in accuracy in trial 1 is an important observation because in non-emergency scenarios, it might not always matter how fast tasks are accomplished, but it always matters how accurately they are completed. With the largest difference in accuracy again being observed in the first trial, support for the use of Enhanced Procedures when completing new complex manual repair tasks continues. And while the difference in accuracy between the Enhanced and Base Procedures decreases as the trials progress, the Enhanced Procedure continues to allow crewmembers to complete the task more accurately for the duration of the experiment.

Accuracy can also be defined as the total number of procedural deviations the subject made during task execution. The total number of deviations broken down by group, trial, and procedure type is shown in Figure 4.7. A Poisson generalized mixed model was used to analyze the total number of deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of deviations was dependent on the interaction effect between procedure and trial, $\chi^2(7) = 17.5$, $p = 0.015$. Additionally, the main effect of procedure was significant, $\chi^2(1) = 31.4$, $p < 0.001$, as well as the main effect of trial, $\chi^2(7) = 124.2$, $p < 0.001$. On average, subjects made 15.1 ± 12.3 procedural deviations when completing the task with the Base Procedure compared to 7.7 ± 4.4 deviations when using the Enhanced Procedure. Post hoc tests identified that subjects on average made more deviations in trial 1 compared to trial 2 ($p < 0.001$), trial 3 ($p < 0.001$), trial 4 ($p < 0.001$), trial 5 ($p < 0.001$), trial 6 ($p < 0.001$), trial 7 ($p = 0.017$), and trial 8 ($p < 0.001$). Additional tests determined that there were significantly more procedural deviations dependent on procedure type with subjects making more deviations when using the Base Procedure compared to the Enhanced Procedure in trial 1 ($p < 0.001$), trial 2 ($p < 0.001$), trial 4 ($p < 0.001$), trial 6 ($p = 0.008$), trial 7 (p =

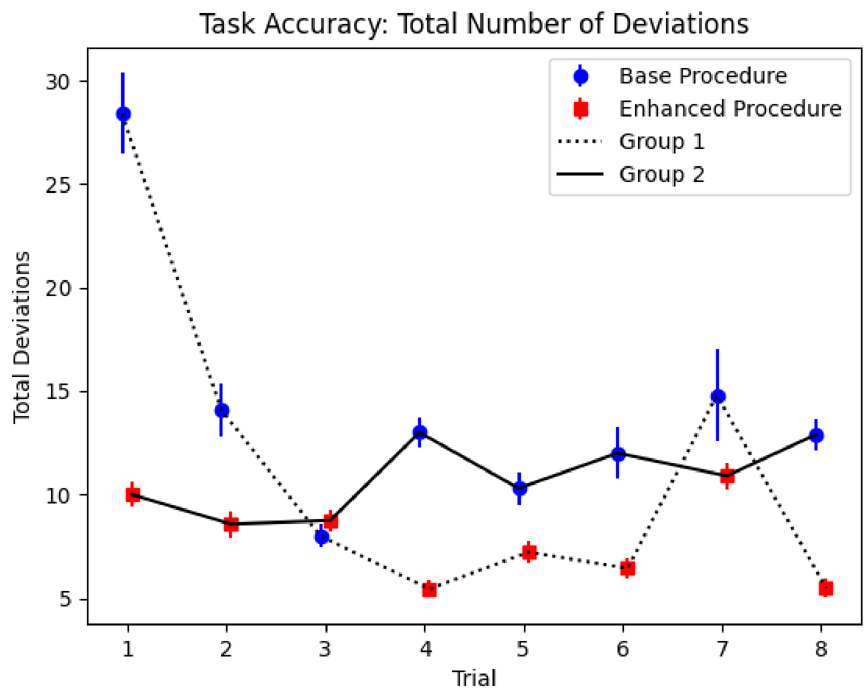


Figure 4.7: Task accuracy (number of deviations) for HERA C6 split up by trial, group, and procedure type.

0.003), and trial 8 (p = 0.002).

The largest difference was seen in the first trial where subjects made on average 28.4 ± 15.6 deviations when using the Base Procedure and 10.0 ± 4.7 deviations with the Enhanced Procedure. This staggering difference of 18.4 deviations between the Enhanced and Base Procedures in trial 1 further supports the conclusion that the Enhanced Procedure system allows subjects to more accurately complete novel procedures. The consistent increased accuracy associated with the Enhanced Procedure throughout the duration of the study supports using enhanced procedures for every-day tasks as well.

4.2.5 Task Performance: Deviations

By classifying the deviations based on PDA (54), the subjects' task completion and interaction with the EPV can be understood on a deeper level. A breakdown of the average count for each deviation mode is shown in Figure 4.8. For each deviation mode the Base Procedure on average had more deviations than the Enhanced Procedure. Chi-square tests were conducted to determine if the number of deviations was dependent on either procedure or trial.

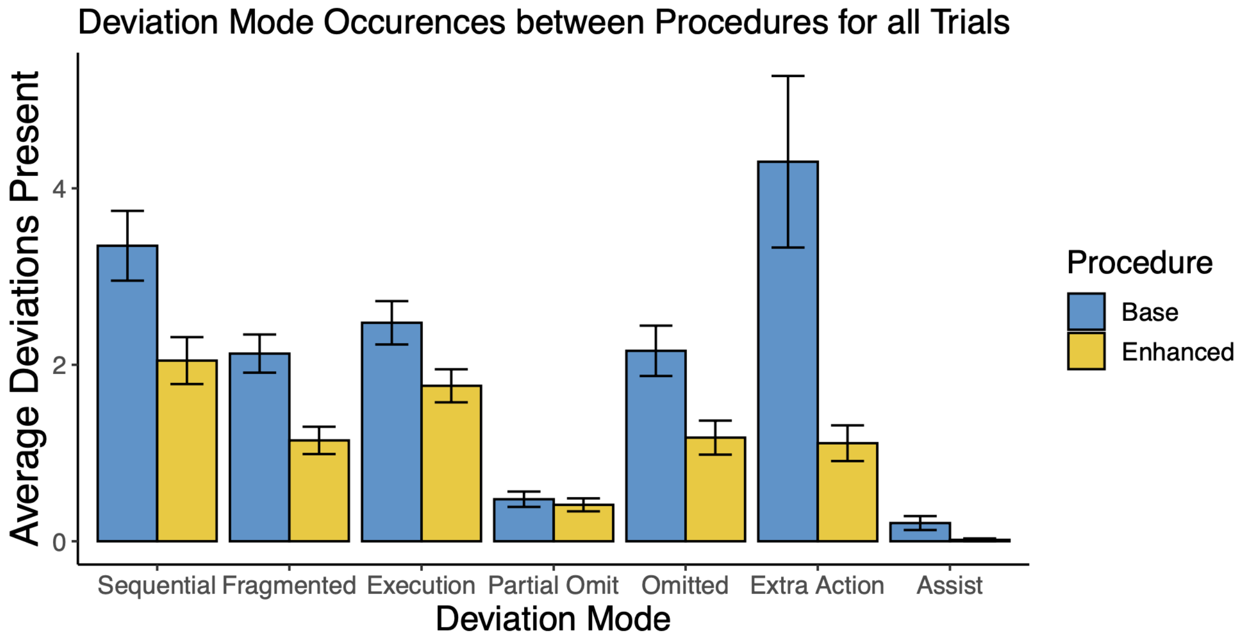


Figure 4.8: Deviation mode count in the HERA C6 study split up by procedure type.

Sequential

A breakdown of the average number of sequential deviations depending on procedure type is shown in Figure 4.9. A Poisson generalized mixed model was used to analyze the total number of sequential deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of sequential deviations was not dependent on the interaction effect between procedure and trial, $\chi^2(7) = 2.53$, p = 0.925, or the main effect of trial, $\chi^2(7) = 4.13$, p = 0.032. The number of sequential deviations was however dependent on procedure, $\chi^2(1) = 4.60$, p = 0.020. On average, subjects made 3.4 ± 3.1 sequential deviations when they used the Base Procedure compared to 2.1 ± 2.1 sequential deviations with the Enhanced Procedure.

Because the Base Procedure did not have an image for every step and the images required more vertical space than the text instructions (thereby spacing steps out farther), there were frequently more steps visible on the screen at any given time when subjects used the Base Procedure compared to the Enhanced Procedure. The decrease in text information and visible steps on the screen is one potential explanation for why the Enhanced Procedure

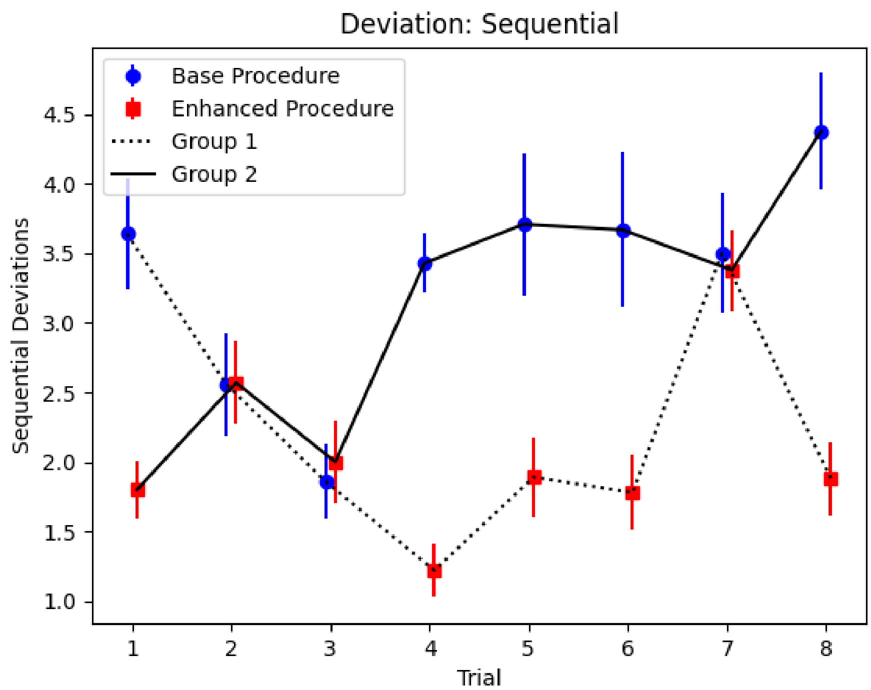


Figure 4.9: Average sequential deviation count in HERA C6 study split up by trial, group, and procedure type.

reduced the number of sequential deviations.

Fragmented

A breakdown of the average number of fragmented deviations depending on procedure type is shown in Figure 4.10. A Poisson generalized mixed model was used to analyze the total number of fragmented deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of fragmented deviations was not dependent on the interaction effect between procedure and trial, $\chi^2(7) = 4.78$, p = 0.686, or the main effect of trial, $\chi^2(7) = 11.1$, p = 0.136. The number of fragmented deviations was however dependent on procedure, $\chi^2(1) = 5.44$, p = 0.020. On average, subjects made 2.1 $\pm$ 1.7 fragmented deviations when they used the Base Procedure compared to 1.1 $\pm$ 1.2 fragmented deviations with the Enhanced Procedure.

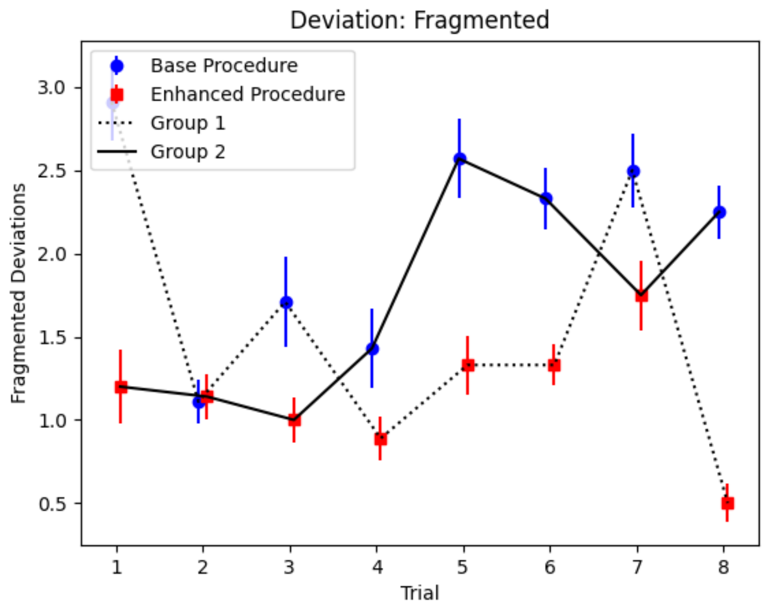


Figure 4.10: Average fragmented deviation count in HERA C6 study split up by trial, group, and procedure type.

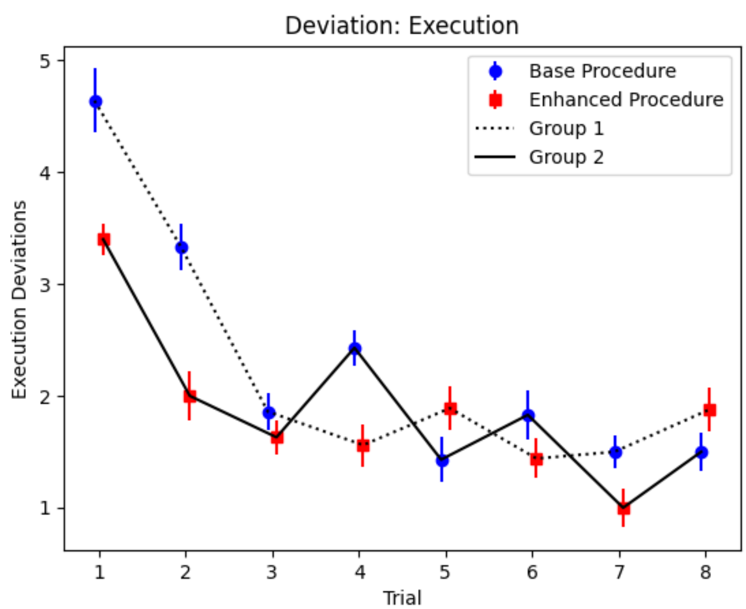


Figure 4.11: Average execution deviation count in HERA C6 study split up by trial, group, and procedure type.

Execution

A breakdown of the average number of execution deviations depending on procedure type is shown in Figure 4.11. A Poisson generalized mixed model was used to analyze the total number of execution deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of execution deviations was not dependent on the interaction effect between procedure and trial, $\chi^2(7) = 2.72$, $p = 0.910$, or the main effect of procedure, $\chi^2(1) = 0.42$, $p = 0.517$. On average subjects made 2.5 ± 2.0 execution deviations when they completed the task with the Base Procedure compared to 1.8 ± 1.5 execution deviations when subjects used the Enhanced Procedure. The number of execution deviations was however dependent on trial, $\chi^2(7) = 26.4$, $p < 0.001$. Post hoc analysis with the Bonferroni correction identified that there were statistically more execution deviations in trial 1 compared to trial 3 ($p = 0.005$), trial 4 ($p = 0.006$), trial 5 ($p = 0.002$), trial 6 ($p = 0.002$), trial 7 ($p < 0.001$), and trial 8 ($p = 0.002$). Essentially, as subjects repeated and became more familiar with the task, they made fewer execution deviations.

Partial Omit

A breakdown of the average number of partial omit deviations depending on procedure type is shown in Figure 4.12. A Poisson generalized mixed model was used to analyze the total number of partial omit deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of partial omit deviations was not dependent on the interaction effect between procedure and trial, $\chi^2(7) = 2.29$, $p = 0.942$, the main effect of procedure, $\chi^2(1) = 0.12$, $p = 0.730$, or the main effect of trial, $\chi^2(7) = 3.88$, $p = 0.793$. On average subjects made 0.48 ± 0.69 partial omit deviations when they used the Base Procedure compared to 0.41 ± 0.59 partial omit deviations with the Enhanced Procedure.

Omitted

A breakdown of the average number of omitted deviations depending on procedure type is shown in Figure 4.13. A Poisson generalized mixed model was used to analyze the total number of omitted deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number

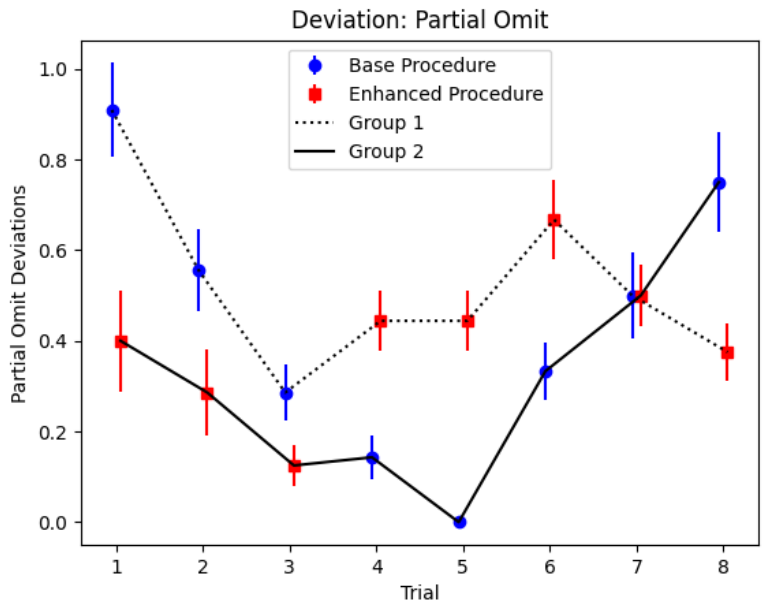


Figure 4.12: Average partial omit deviation count in HERA C6 study split up by trial, group, and procedure type.

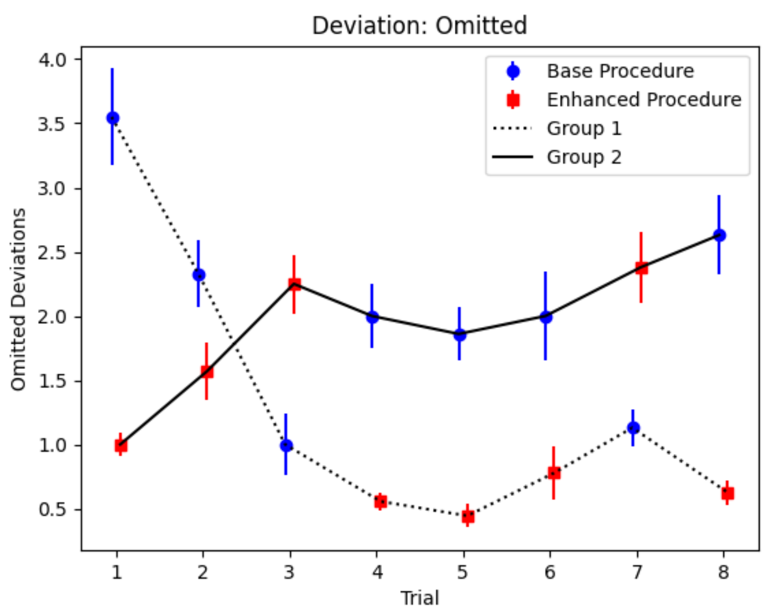


Figure 4.13: Average omitted deviation count in HERA C6 study split up by trial, group, and procedure type.

of omitted deviations was not dependent on the interaction effect between procedure and trial, $\chi^2(7) = 8.49$, $p = 0.292$, the main effect of procedure, $\chi^2(1) = 1.69$, $p = 0.193$, or the main effect of trial, $\chi^2(7) = 10.41$, $p = 0.166$. On average, subjects made 2.2 ± 2.3 omitted deviations when they used the Base Procedure compared to 1.2 ± 1.5 omitted deviations with the Enhanced Procedure.

Extra Action

A breakdown of the average number of extra action deviations depending on procedure type is shown in Figure 4.14. A Poisson generalized mixed model was used to analyze the

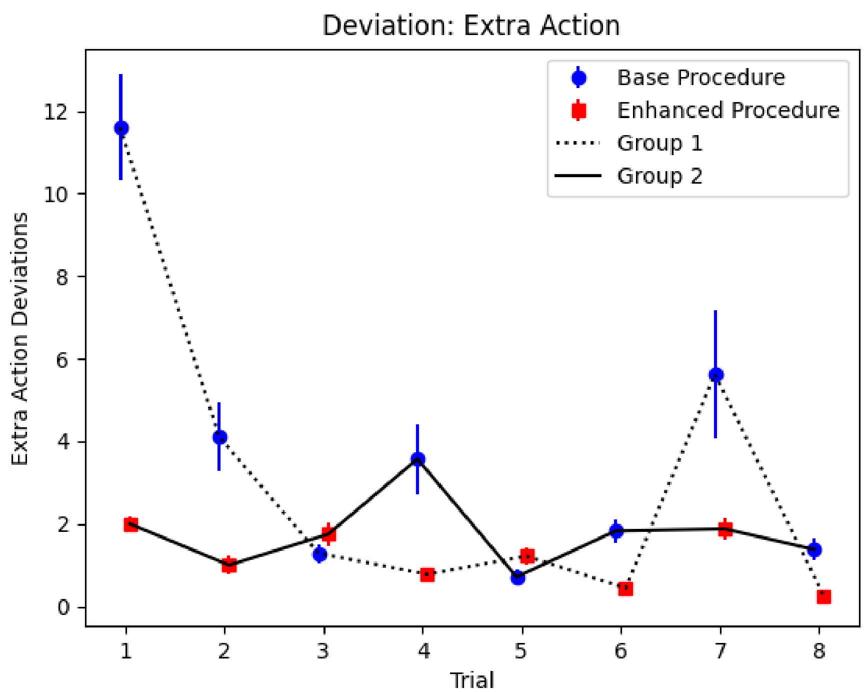


Figure 4.14: Average extra action deviation count in HERA C6 study split up by trial, group, and procedure type.

total number of extra action deviations depending on procedure and trial with subject as a repeated measure. The main interaction effects were analyzed and determined that the total number of extra action deviations was dependent on the interaction effect between procedure and trial, $\chi^2(7) = 25.9$, $p < 0.001$, the main effect of procedure, $\chi^2(1) = 24.7$, $p < 0.001$, and the main effect of trial, $\chi^2(7) = 123.7$, $p < 0.001$. On average subjects made 4.3 ± 7.7 extra action deviations when they used the Base Procedure compared to 1.1 ± 1.6 extra action deviations with the Enhanced Procedure. Post hoc comparisons identified that there were statistically more extra action deviations in trial 1 compared to trial 2 ($p = 0.039$), trial 3 ($p = 0.001$), trial 4 ($p = 0.004$), trial 5 ($p < 0.001$), trial 6 ($p < 0.001$), and trial 8 ($p < 0.001$). Additionally, trial 7 had statistically more extra action deviations than trial 3 ($p = 0.006$), trial 4 ($p = 0.020$), trial 5 ($p < 0.001$), trial 6 ($p < 0.001$), and trial 8 ($p < 0.001$). Post hoc comparisons analyzing the interaction effect between procedure and trial found that trial 1 ($p < 0.001$), trial 2 ($p =0.001$), trial 4 ($p = 0.002$), trial 6 ($p = 0.035$), and trial 7 ($p < 0.001$) all had significantly more extra action deviations when subjects used the Base Procedure compared to when subjects used the Enhanced Procedure.

Overall, the extra action deviation has the largest difference depending on if subjects used the Base or Enhanced Procedure out of all the deviation modes with subjects making an average of 3.2 more deviations when using the Base Procedure. The range of observed extra action deviations was also extremely varied between the procedures with a maximum of 36 extra action deviations made by a subject during one trial with the Base Procedure, compared to the maximum of 6 extra action deviations made with the Enhanced Procedure. Through observation it was determined that in the cases with large numbers of extra action deviations, subjects typically forgot a step when putting back together the generator. How many extra action deviations they would have would depend on when they realized they missed something because once they realized it, they would have to go back and repeat steps to disassemble the generator to return to the point where they could correct their mistake. Interestingly, this forgetting a reassembly step occurred most when subjects used the Base Procedure which is clearly represented in the large discrepancy between the number of extra action deviations between the Base and Enhanced Procedures.

Assist

A breakdown of the average number of assist deviations depending on procedure type is shown in Figure 4.15. Because there were so few assist deviations in the later trials, a Poisson generalized mixed model was used for trials 1 and 2 to compare the average number of assist deviations between the Base Procedure and the Enhanced Procedure with subject as a repeated measure. The test determined that the number of assist deviations was dependent on procedure type, $\chi^2(1) = 6.89$, $p = 0.009$. When subjects performed the task in trial 1 with

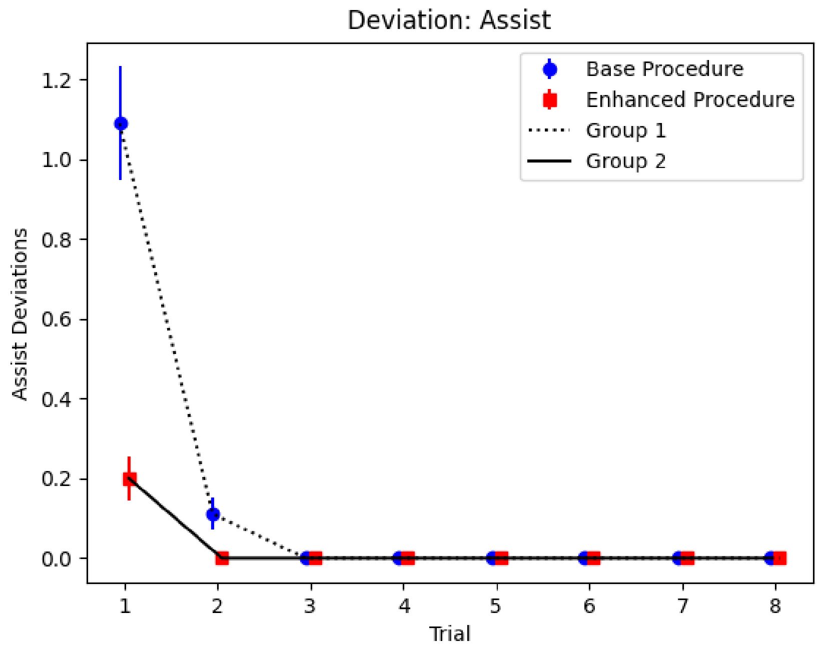


Figure 4.15: Average assist deviation count in HERA C6 study split up by trial, group, and procedure type.

the Base Procedure, they had an average number of 1.09 ± 1.14 assist deviations compared to 0.200 ± 0.447 with the Enhanced Procedure. This is a significant metric because this deviation mode relates directly to the autonomy of the crewmember. A higher number of assist deviations indicates crewmembers needed an increased amount of outside assistance to complete the task.

In total, 12 out of the 13 assist deviations occurred in trial 1, and 12 out of the 13 deviations occurred with the Base Procedure. In trial 1, because of the compliance issues with the schedule, there was a notable difference in n values. However, even after taking that into account, 7 out of the 11 participants (64%) who started with the Base procedure asked for some sort of assistance related to the completion of the task. That is significantly more than the 1 out of 5, or 20%, of the subjects who started with the Enhanced Procedure. Looking at this first attempt, the difference in how many times subjects requested assistance is further evidence that the Enhanced Procedure allows subjects to complete the repair task more autonomously.

4.2.6 Workload

A breakdown of the subjects' perceived workload is shown in Figure 4.16. Workload was measured according to the NASA-TLX survey on a scale of 1-20 with 1 correlating to a minimal workload and 20 correlating to an extensive workload (11). A linear mixed

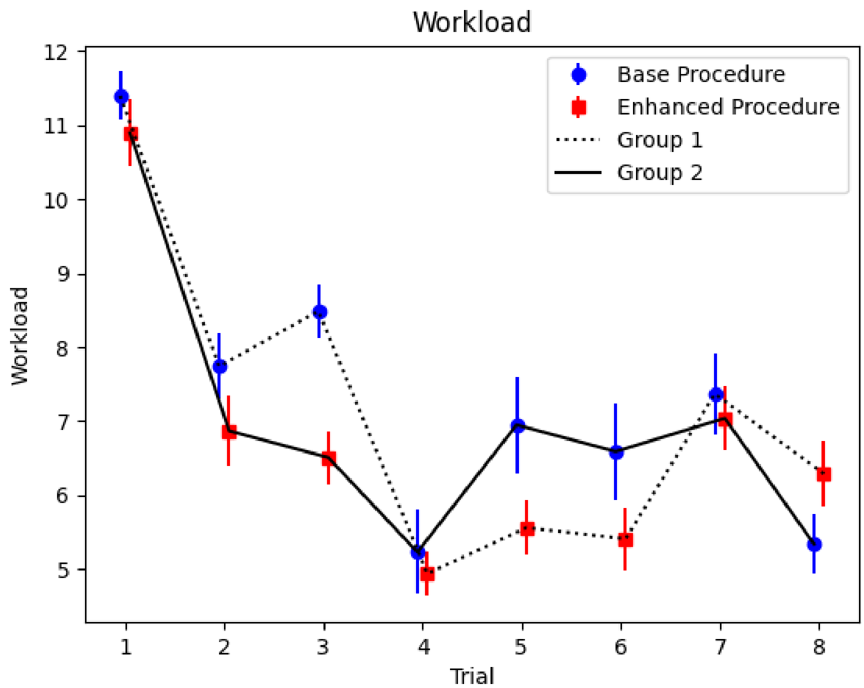


Figure 4.16: Workload for HERA C6 split up by trial, group, and procedure. The scale is continuous from 1-20 with higher values indicating a higher perceived workload.

model was used to determine if subjective workload was dependent on either procedure or trial with one within variable, subject. There was no interaction effect between procedure and trial ($F(7, 94.4) = 1.00$, $p = 0.436$). The main effect of procedure was not significant ($F(1, 91.5) = 3.57$, $p = 0.062$), however the main effect of trial was ($F(7, 91.1) = 8.53$, $p < 0.001$). On average subjects had a workload of 6.42 ± 3.40 when they used the Enhanced Procedure, compared to 7.60 ± 4.23 with the Base Procedure. Subjects therefore did on average have a slightly lower workload when they used the Enhanced Procedure, however it was not determined to be statistically significant. Further analysis identified that the subjects' reported workload in trial 1 was statistically higher than for trial 2 ($p = 0.001$), trial 3 ($p = 0.003$), trial 4 ($p < 0.001$), trial 5 ($p < 0.001$), trial 6 ($p < 0.001$), trial 7 (p

= 0.001), and trial 8 ($p < 0.001$). Essentially, the subjects perceived workload during trial 1 was statistically higher than for every other trial. This confirms that crewmembers felt they had to work harder the first time they completed the task, likely because it was a novel procedure with unfamiliar hardware.

4.2.7 Trust

After completion of the manual repair task, subjects completed the Trust in Automated Systems Survey (15) to evaluate their trust in the EPV. For this survey, the Enhanced Procedure had n = 63 while the Base Procedure had n = 56. The average results from the survey are shown in Figure 4.17. The scale is continuous from 1-7 with higher numbers relating to higher levels of trust. A linear mixed model determined that the interaction effect between procedure and trial was not significant ($F(7, 88.7) = 2.01$, $p = 0.063$. Trust was however dependent on the main effects of both procedure ($F(1, 91.5) = 5.05$, $p = 0.027$) and trial ($F(7, 89.3) = 3.60$, $p = 0.002$). On average, subjects reported a statistically higher level of trust in the Base Procedure (4.26 ± 0.94) than the Enhanced Procedure (3.90 ± 1.01).

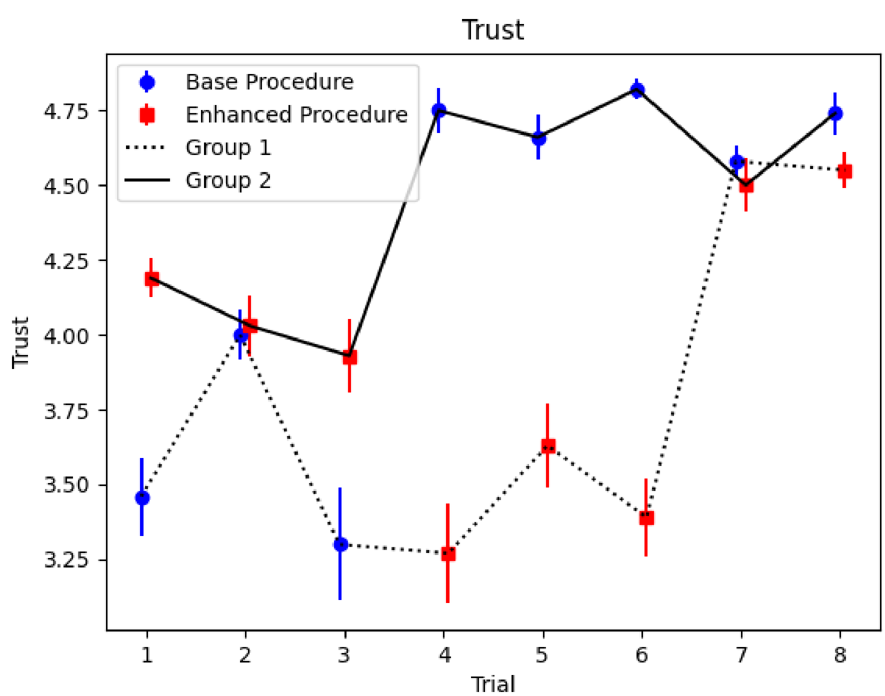


Figure 4.17: Trust in the system for HERA C6 split up by trial, group, and procedure. The scale is continuous from 1-7 with higher values indicating a higher level of trust in the system.

Post hoc comparisons with the Bonferroni correction determined that the average trust was significantly different between trial 1 and trial 8 ($p = 0.040$), trial 3 and trial 7 ($p = 0.028$), and trial 3 and trial 8 ($p = 0.007$). Looking at the trust breakdown for each trial, it is apparent that subjects in the first 3 trials trusted the Enhanced Procedure slightly more than the Base Procedure. However, at trial 4 there is a dramatic shift with group 2 suddenly having a much higher trust in the Base procedure. Trial 4 is when the groups switched procedure types which can account for some of this change. Another possible explanation for the decrease in trust in the Enhanced Procedure is sensor malfunction. It was reported in post-mission debriefs for Missions 1 and 3 that some of the sensor feedback was unreliable. Evaluation of the generator identified that some sensors had been slightly dislodged from their position resulting in inconsistent sensor readings. These sensor malfunctions likely impacted the trust subjects had in the EPV as the system incorrectly reported the state of the generator hardware, making crewmembers second guess their work. Subjects were asked in training to report such instances so the sensors could be replaced to ensure the best possible EPV system setup for the task; however this was not always followed in practice. Interestingly, this large discrepancy in trust between procedure types is not seen in the final two trials after the long break.

4.2.8 Post-Task Survey Responses

After completion of the task, subjects completed a post task survey gathering their opinions on the procedure and how they felt when completing the task. Subjects were specifically asked to rate their performance on the task, how confident they felt when completing the task, and how helpful they found the procedure.

Perceived Performance

In the Post-Task Survey, subjects were asked to self-rate their performance on a scale from 1 (Performed very poorly) - 7 (Performed very well) with the results shown in Figure 4.18. Subjects' perceived performance was not dependent on the interaction effect between procedure and trial ($F(7, 79.3) = 0.365$, $p = 0.920$) or the main effect of procedure ($F(1, 90.6) = 0.345$, $p = 0.558$). Subjects' perceived performance was however dependent on the main effect of trial ($F(7, 88.4) = 3.582$, $p = 0.002$). Post hoc comparisons determined that subjects' perceived performance for trial 1 was significantly lower than trial 5 ($p = 0.002$), trial 6 ($p = 0.002$), trial 7 ($p = 0.008$), and trial 8 ($p = 0.004$). These values, as well as the visual upward trend on the graph, indicate that subjects felt their task performance increased as they became more familiar with the task. This trend also correlates with how subjects felt their workload decreased as the trials continued.

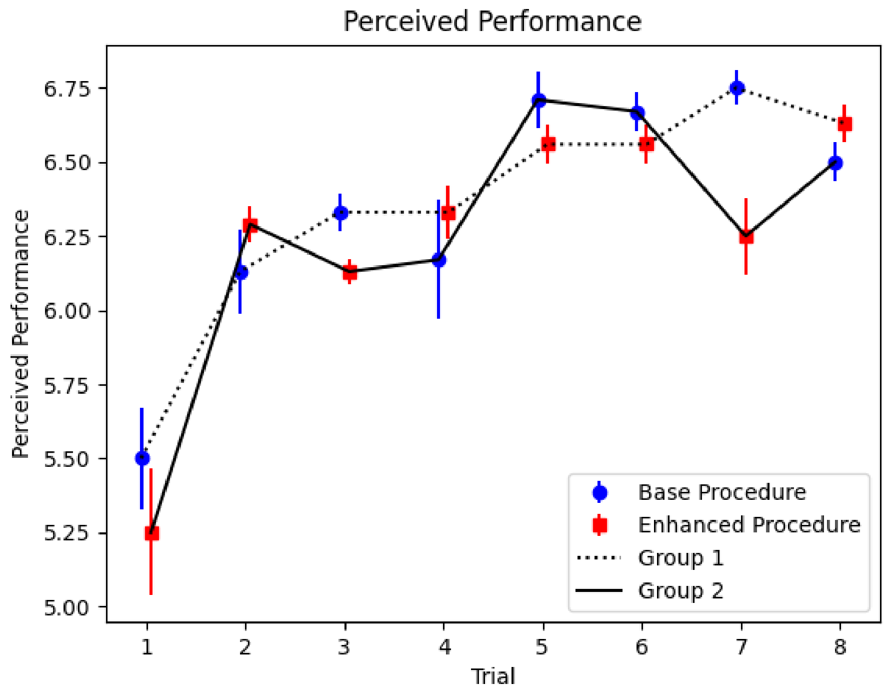


Figure 4.18: Self-perceived task performance for HERA C6 experiment split up by trial, group, and procedure. The scale is continuous from 1-7 with higher values indicating an improved performance.

Self-Rated Confidence During Task Execution

After each trial, crewmembers were also asked to rate their perceived confidence as they completed the task from 1 - 7, with higher numbers indicating an increased level of confidence. Subjects' perceived confidence during task execution is shown in Figure 4.19. Subjects' confidence was not dependent on the interaction effect between procedure and trial ($F(7, 80.6) = 0.263$, $p = 0.966$), or the main effect of procedure ($F(1, 91.7) < 0.001$, $p = 0.981$), however it was dependent on the main effect of trial ($F(7, 89.4) = 4.40$, $p < 0.001$). Post hoc comparisons identified that the crewmembers perceived confidence in trial 1 was significantly lower than in trial 2 ($p < 0.001$), trial 3 ($p = 0.009$), trial 4 ($p = 0.019$), trial 5 ($p < 0.001$), trial 6 ($p = 0.037$), trial 7 ($p < 0.001$), and trial 8 ($p = 0.001$). After the crew's first trials, their confidence during the mechanical repair task increased and then remained relatively constant.

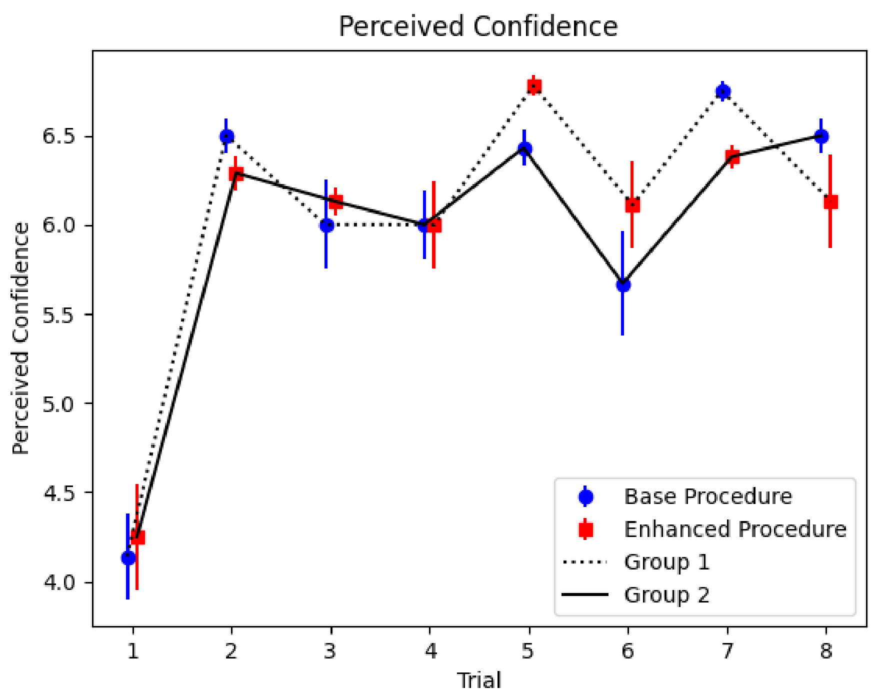


Figure 4.19: Self-rated confidence during task execution for HERA C6 experiment split up by trial, group, and procedure. The scale is continuous from 1-7 with higher values indicating an increased confidence level.

Perceived Procedure Helpfulness

In the Post-Task Survey crewmembers were asked to rate how helpful they found the procedure they used on a scale from 1 (Not useful) - 7 (Very useful). Subjects' self-reported procedure helpfulness is shown in Figure 4.20. Subjects' self-reported procedure helpfulness was not dependent on the interaction effect between procedure and trial ($F(7, 71.1) = 0.510$, $p = 0.824$) or the main effect of procedure ($F(1, 92.9) = 0.794$, $p = 0.375$), however it was dependent on the main effect of trial ($F(7, 89.1) = 2.657$, $p = 0.015$). Post hoc comparisons identified that reported procedure helpfulness was significantly lower for trial 1 compared to trial 2 ($p < 0.001$), trial 3 ($p = 0.010$), trial 4 ($p = 0.021$), trial 5 ($p = 0.006$), trial 7 ($p <$

0.001), and trial 8 (p = 0.008).

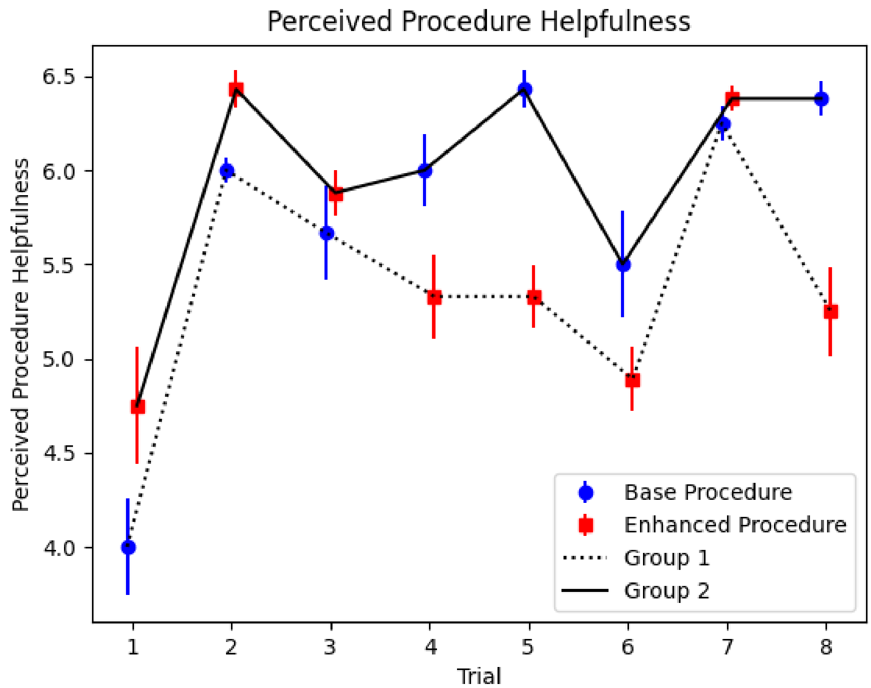


Figure 4.20: Self-perceived procedure helpfulness for HERA C6 experiment split up by trial, group, and procedure. The scale is continuous from 1-7 with higher values indicating an increased procedure helpfulness.

4.2.9 Post-Mission Survey Responses

After the mission was complete, the crew was asked to complete a Post-Mission Survey to collect their overall thoughts on the different procedure systems. The subjects were asked to rate the helpfulness of the EPV and each of the enhancements. They were also asked how potentially helpful each of the enhancements could be outside of the context of the generator repair task and they were given numerous free response questions about how, when, and what was helpful about the EPV.

Enhancements

The crewmembers were asked to rate on a scale from 1 (Not useful) - 7 (Very useful) how helpful they found each of the enhancements. The results are shown in Figure 4.21. In order from most to least helpful, the enhancements were rated: enhanced photos, step tracking, videos, lasers, and sensor feedback.

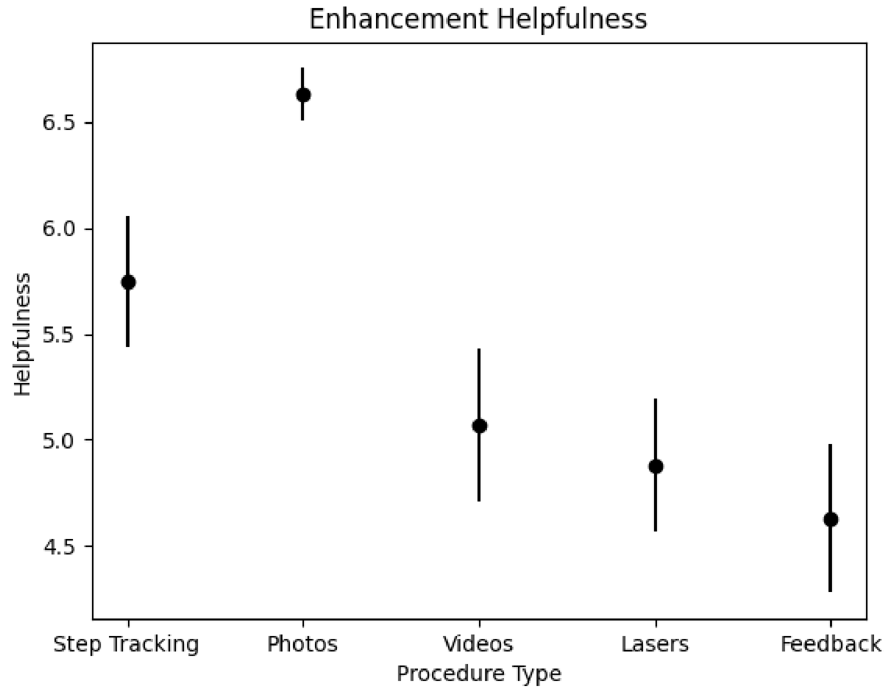


Figure 4.21: HERA reported enhancement helpfulness on a scale from 1 (Not useful) - 7 (Very useful).

One reason for the low score for sensor feedback helpfulness is that there were times during the HERA study when the sensor feedback failed to respond properly. Because live observation and anomaly resolution was not possible during the HERA campaign due to the isolated nature of the experiment, our team was usually only notified after missions that sensors failed to work as expected. The crew was asked in training to report sensor malfunctions so that the main generator could be replaced with a backup to address hardware issues, however in practice this advice was rarely followed resulting in certain missions failing to have reliable sensor feedback. This inconsistent sensor feedback significantly impacted the crews' opinion of the enhancement and was specifically noted by subjects in the free response section which will be discussed next. Increased sensor robustness and improved hardware design that better integrates sensors into the machinery could help mitigate this issue in future experiments and other IoT sensor applications.

Subjects were also asked to rate on a scale from 1 (Not useful) - 7 (Very useful) how helpful they thought the enhancements could be outside of the context of the generator

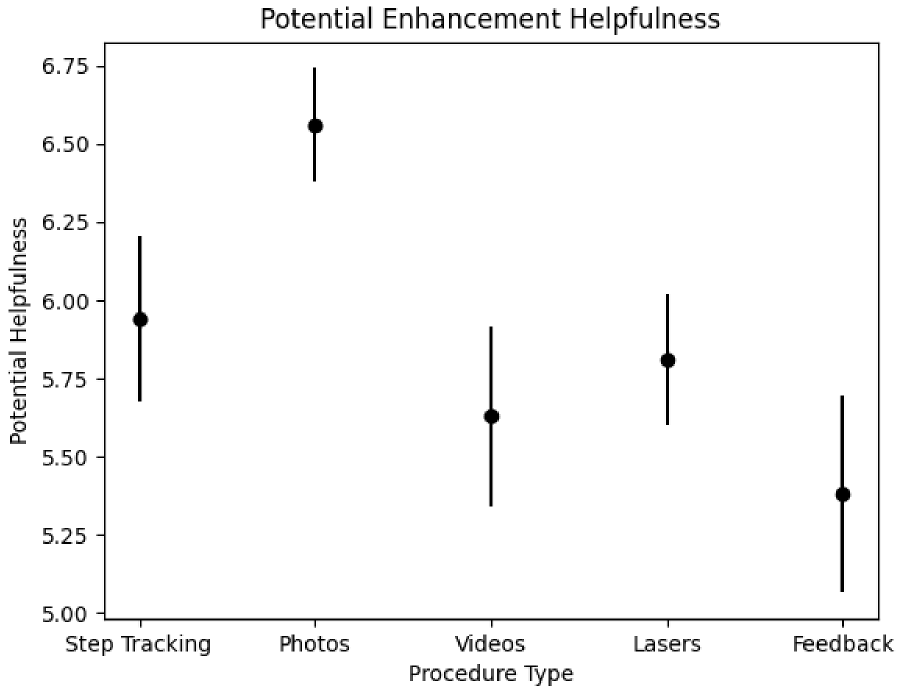


Figure 4.22: HERA reported enhancement potential helpfulness on a scale from 1 (Not useful) - 7 (Very useful).

repair task. The results are shown in Figure 4.22 and match a similar trend to the previously reported procedure helpfulness with the only order change being an increased potential helpfulness from the laser indicators compared to the videos. The new enhancement order from potentially most helpful to least is: enhanced photos, step tracking, laser indicators, videos, and sensor feedback. When a free response question in the survey asked why subjects changed their answers, 11 subjects reported that they thought the enhancements could be more helpful for longer procedures or more complex tasks.

Free Response

In the Post-Mission survey, subjects were asked numerous free response questions about when the EPV was helpful, how it was helpful, what was helpful about it, and what could be improved. These free responses were organized and tallied depending on the number of subjects that mentioned the same themes. Crewmembers were also interviewed for 15 minutes after the mission and asked similar questions in person to further clarify their opinions.

In general, crewmembers reported that the enhancements provided helpful information which increased their situational awareness and made task completion easier and "more fun". 13/16 crewmembers reported that the Enhanced Procedure was, or would have been, the most helpful during the earlier run-throughs of the task when the procedure was new and they were less familiar with the task and hardware. It was also emphasized that the enhancements would be more beneficial for more complex tasks, for tasks that workers had not been previously trained on, or as a tool for training refreshers. Five crewmembers said the Enhanced Procedure was more helpful than the Base Procedure after the extended month-long break, however this varied depending on how well each subject remembered the original task.

Interestingly, two subjects also specifically noted that they didn't realize they were consistently completing a step wrong the first three times they used the Base Procedure. They realized their previous errors when they switched to using the Enhanced Procedure on the 4th trial, indicating that the Enhanced Procedure did indeed provide more information than the Base Procedure in such a way that it allowed more accurate task completion.

Because the enhancements were primarily optional, around half of the subjects preferred the Enhanced Procedure over the Base Procedure even after they were familiar with the task. There were however seven crewmembers that preferred the Base Procedure after the initial trials because it was "nimble" and allowed them to "complete the task faster". This preference for the Base Procedure was also notably impacted by the reliability of the sensor feedback. Seven crewmembers mentioned that they had an issue at one point with the sensor feedback failing to adequately acknowledge step completion. The sensor failure rate varied by mission with 1 sensor having issues in the first mission and 3 sensors having issues in the 3rd mission. While the crew was asked to switch to the backup generator should sensors malfunction, in practice this was not followed which did notably impact subjects' opinions of the enhancement. When the sensors performed unreliably, subjects reported that they spent more time second guessing their step completion and became frustrated with the procedure system. Despite this, six crewmembers mentioned that they thought the sensor feedback was useful, giving them confidence they had successfully completed the steps in question. Multiple crewmembers further suggested that more steps should have sensor feedback to increase the impact of the enhancement.

The most commonly mentioned enhancement was the enhanced images. Eleven crewmembers mentioned they found the enhanced images helpful while five mentioned they liked the videos for complex steps. Because the photos were on the right hand side of the EPV and did not take up a large footprint of the screen, it was consistently noted that the crew appreciated the ability to zoom in on pictures to get a better view of what they were working on. Subjects also mentioned that they liked being able to see the following steps' picture simultaneously to see the "before and after" of the current step as it gave them a better understanding of what they were trying to accomplish. Additionally, six crewmembers referenced the laser indicators, reporting that they enjoyed using the lasers but thought the feature would be more beneficial when working with more complex hardware or when parts were "in darker areas of the machine or hidden away".

4.2.10 Outlier Removal Considerations

There are a multitude of ways to consider or determine outliers. For this experiment we plotted a histogram of total task completion time, task accuracy, the number of procedural deviations, and the number of extra action deviations as shown in Figures 4.23, 4.24, 4.25, and 4.26 respectively. Around the point where the normal distribution ended, a line was drawn indicating where the outliers were removed for the following analysis.

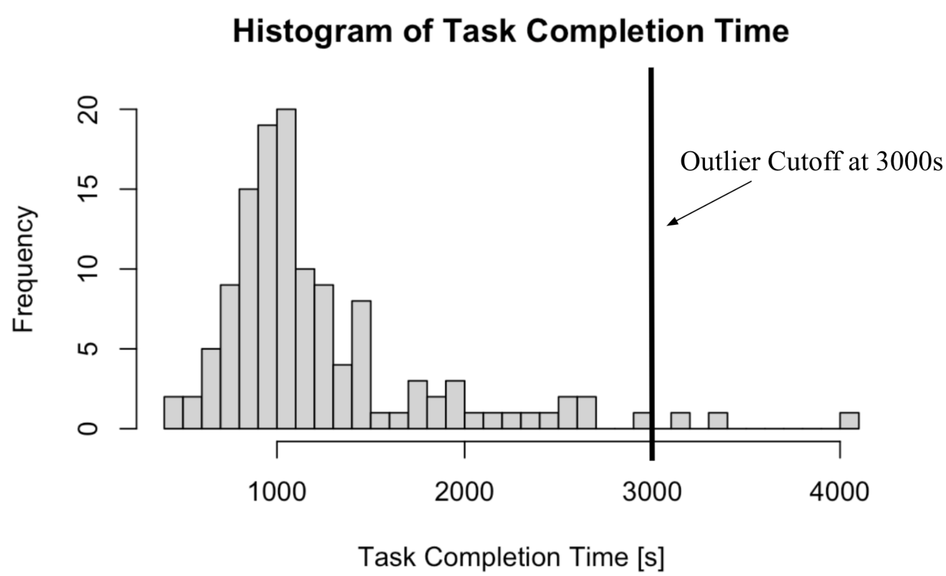


Figure 4.23: A histogram showing the task completion time distribution from HERA C6.

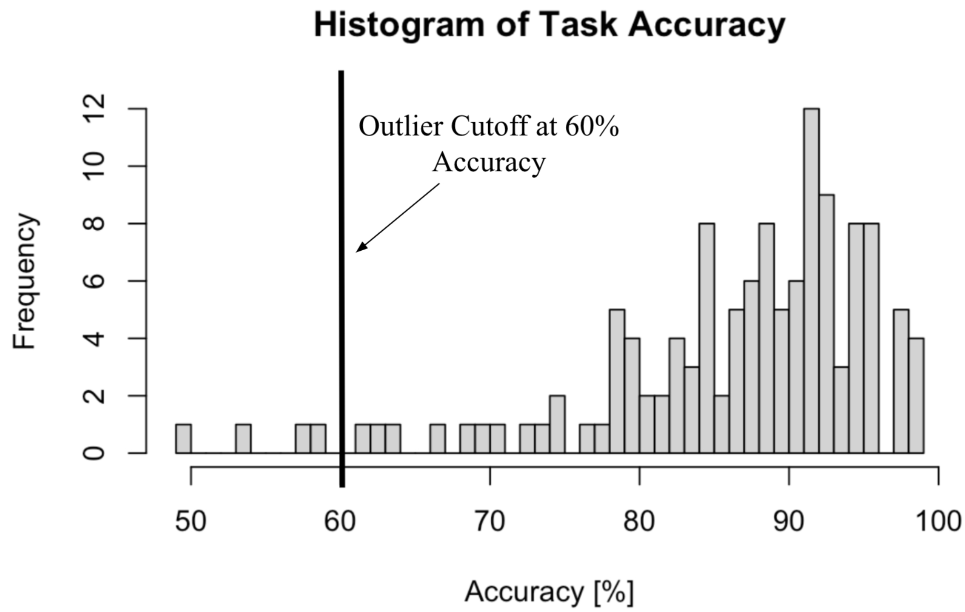


Figure 4.24: A histogram showing the accuracy distribution from HERA C6.

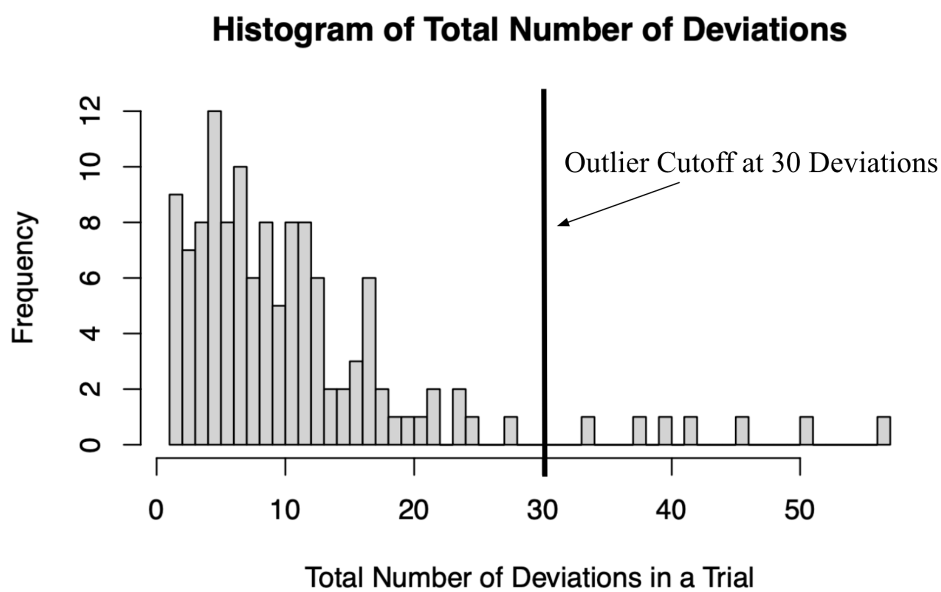


Figure 4.25: A histogram showing the total number of deviations distribution from HERA C6.

Many of the outliers for each metric were outliers for the other metrics as well. Therefore, removing the outliers from one group removed most of the outliers from the others. When removing outliers in the above cases however, it was determined that all outliers were from

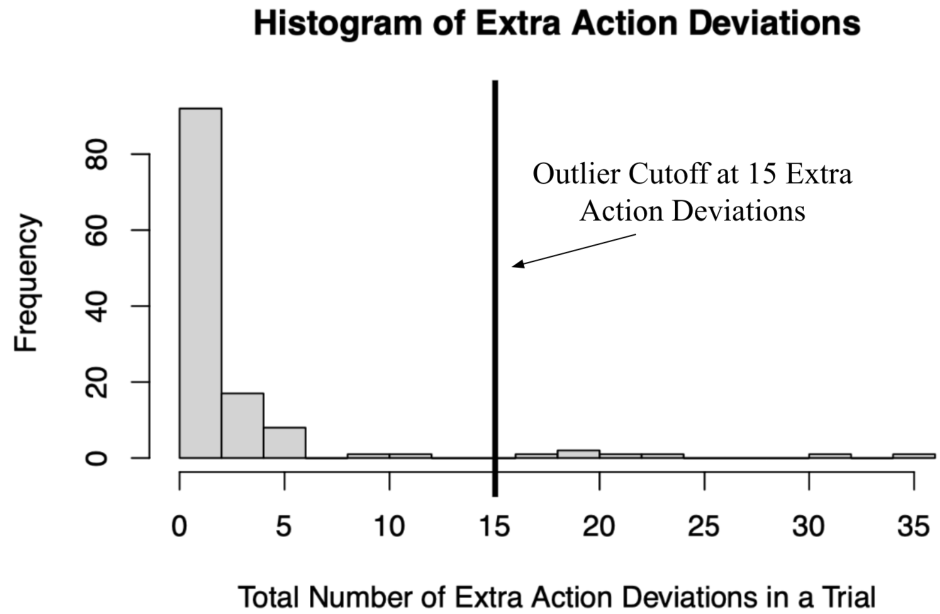


Figure 4.26: A histogram showing the total number of extra action deviations distribution from HERA C6.

the Base Procedure category, primarily from the first trial. Removing the outliers skewed the n values and also proved to have no statistical effect on the main study results. For example, when removing the outliers from the number of Extra Action deviations (Figure 4.26), outliers were considered to be when subjects made more than 15 extra action deviations during one trial. This was typically caused by a subject forgetting a step and moving on with the procedure, forcing them to repeat steps to undo their mistake. Removal of the extra action deviation outliers removed seven trials from the data and resulted in a new n value breakdown as shown in Table 4.3. In this data set there is a total of n = 63 for the Enhanced Procedure and n = 57 for the Base Procedure, with a total n = 119. Four out of seven outliers were removed from trial 1 and every outlier occurred with the Base Procedure. This finding that all outliers were from trials using the Base Procedure is notable, highlighting the extreme difference in performance generated between the Base and Enhanced Procedures.

With the outliers removed, a linear mixed model ANOVA was conducted on task completion time with two between variables (procedure and trial) and one within variable (subject). The mixed ANOVA again determined there was no statistical difference in timing dependent on procedure ($F(1, 1) = 0.102$, $p = 0.750$) but that it was dependent on trial ($F(1,7) =$

Table 4.3: Breakdown of n values for each procedure type and trial in HERA C6 after outliers have been removed. For changed values, previous n values before the removal of outliers are stated in [ ].

Trial Number	Enhanced Procedure	Base Procedure
Trial 1	5	7 [11]
Trial 2	7	8 [9]
Trial 3	8	7
Trial 4	9	6 [7]
Trial 5	9	7
Trial 6	9	6
Trial 7	8	7 [8]
Trial 8	8	8
Total n	63	56 [63]

71.715, $p < 0.001$). Continuing with accuracy, a mixed ANOVA again determined that task accuracy was dependent on both procedure ($F(1, 1) = 29.37$, $p < 0.001$), trial ($F(1, 7) = 4.14$, $p < 0.001$), and the interaction effect between procedure and trial ($F(1, 7) = 2.12$, $p = 0.049$). It is notable that these main study findings come to the same conclusions as before the outliers were removed. Similar testing was performed when removing outliers from the other metrics which resulted in the same conclusions. As such, it was decided to not remove outliers from this experiment as it does not change the study results and skews the data by only removing data from the Base Procedure trials.

4.3 Discussion

In our hypothesis we anticipated that the Enhanced Procedure would allow subjects to complete the task in the same amount of time, with better accuracy, lower workload, and the same level of trust in the system. ANOVA testing determined that the Enhanced Procedure did allow for subjects to complete the task with increased accuracy without changing the time it took to complete the task. Crewmember workload however did not statistically change between the two procedure types and the crew reported a lower level of trust in the

Enhanced Procedure.

While the Enhanced Procedure did not statistically help the subjects complete the task faster compared to the Base Procedure, the enhancements did not slow crewmembers down either. Because of this, the added enhancements in the Enhanced Procedure could be used as an autonomous assistance source when astronauts are completing any task, including everyday tasks. The extra visuals to study, videos to watch, and sensor feedback to engage with are all optional features of the EPV and therefore do not need to be interacted with if subjects do not require extra aid. While crewmembers did report they thought the Base Procedure allowed them to complete the task quicker in later trials, the timing data does not support this, indicating there is no temporal downside to including these enhancements in the procedure. By providing extra information in the form of these optional enhancements, workers are provided with more detailed information should they need it, without being burdened with unnecessary assistance.

In regards to accuracy, the crew had an average accuracy 7.6% greater when they used the Enhanced Procedure compared to the Base Procedure, with an average of 7.4 less deviations. This was determined to be a statistically significant difference and potentially the greatest finding from this study. In many non-emergency scenarios, the time it takes to complete a task is not necessarily critical, however it always matters how accurately tasks are completed. Due to the limited ability for MCC to catch procedural deviations, any procedural enhancements that can increase task completion accuracy should be heavily invested in as space travel continues to require more autonomy from its astronauts.

When looking at crewmembers' procedural accuracy and task completion time, there is a major difference between trial 1 and the following trials. This large discrepancy in timing and accuracy is not surprising because in the first trial the crew were completing a novel task they had never seen before. The crew had previously been introduced to the generator in a short training video but had never personally interacted with the analog previous to mission ingress, keeping them unfamiliar with part locations and procedure steps. For trials 2-8 the task was no longer new and the study became more of a long term training experiment compared to rapid JITT. In the first trial, subjects completed the task statistically faster when using the Enhanced Procedure. On average, subjects completed the task 12.5 minutes, or 38.6%, faster. At the same time, subjects completed the task with an average improved accuracy of 16.9% with 18.4 less deviations when they used the Enhanced Procedure. This drastic difference in performance during trial 1 indicates how the Enhanced Procedure could be especially helpful for JITT, or when subjects are less familiar with the task or equipment they are working with. The enhancements clearly helped to introduce and orient subjects faster and more effectively to the new hardware and allowed them to more efficiently and accurately complete the repair task. With future LDEMs likely requiring astronauts to complete tasks they never practiced before ingress, the EPV features could be an excellent tool to help ease the strain of completing unfamiliar tasks.

Another interesting observation from this study is that despite subjects having a statistically greater task accuracy when using the Enhanced Procedure (Figure 4.6), they did not perceive this improved performance (Figure 4.18). While this thesis defines performance as a combination of efficiency, accuracy and workload, crewmembers were asked to rate their performance in the Post-Task survey where performance was not defined, thereby leaving it up to their own interpretation. Despite this, it was surprising that crewmembers did not perceive their statistically improved accuracy as an increase in their perceived performance. This disconnect between measured performance and self-perceived performance is an important observation that leads to the recommendation: when analyzing the effectiveness of procedures, it is important to check quantitative values as opposed to simply taking into account subjects' perceptions.

While we hypothesized that the Enhanced Procedure would decrease subjects' workload, the difference in workload between the procedure types was not found to be statistically significant in this study. Crewmembers did however have a significantly lower trust in the Enhanced Procedure than in the Base Procedure which was not anticipated in our hypothesis. As previously mentioned, a large factor in this discrepancy was likely caused by sensor feedback malfunctions which greatly impacted crewmembers opinions of the real-time sensor feedback. To improve the sensor feedback and retain workers' trust in enhanced procedures in future applications, embedded sensors must be installed in such a way that they are robust enough to handle accidental manual manipulation, vibration, and power fluctuations. Due to the geometry of the HONDA generator, a limited number of sensors could be embedded into the system and many of these had to be secured with velcro or glue to the outside of different parts. By designing a system with embedded sensors in mind, areas can be left open and available to more securely house these monitoring devices in both secure and maintainable ways. Similarly, wire pathways can be preplanned to ensure a cleaner and more robust embedded sensor monitoring system. Ensuring that these sensors are reliable even after years of use will likely improve the trust that workers have in them.

Overall, while subjects had a lower level of trust in the Enhanced Procedure and the EPV did not improve crewmembers' average workload or task efficiency, the procedure did have the important benefit of helping workers complete the repair task more accurately. Also, during subjects' first time completing the task, the Enhanced Procedure significantly improved both workers efficiency and accuracy indicating the procedural enhancements could be an important tool to help workers more effectively complete new or unfamiliar tasks.

Chapter 5

Discussion

While both the HERA study and the Davis study used the same task and procedure system, they were very different studies due to their experimental design. The HERA study was a within-subjects experiment comparing two procedure types: the Base Procedure and Enhanced Procedure. Due to the isolated nature of HERA, subjects disassembled and re-assembled the generator and there was very limited proctor oversight during the duration of the 45 day experiments. The HERA analog allowed for enhanced procedures to be studied in an isolated and confined environment with little or no MCC communication. However, because of the limited number of subjects, crewmembers repeated the experiment 8 times to increase the n value. In comparison, the Davis study had a between-subjects design comparing five different procedure types with varying enhancements. The Davis study allowed for continuous proctor oversight during the experiment and had no limits on the number of subjects so the individual enhancements could be studied separately with each subject only using one procedure type.

In both the Davis and HERA experiments, there was a statistically significant decrease in task completion time associated with enhanced procedures. During the HERA experiment, subjects completed the task 12.5 minutes faster with the Enhanced Procedure during the first trial compared to those who used the Base Procedure, and in the Davis experiment subjects completed the task 4.4 minutes faster when they were provided with enhanced visuals. These results indicate that when provided with enhanced visuals (including an image for every step and videos for complex steps), subjects can complete novel tasks more efficiently. Providing more images, and even repeating previous pictures helps ensure subjects have all the required information to complete each step. While some procedures provide information and then expect workers to remember it, the results of these studies indicate that subjects can complete tasks quicker if they do not need to remember or review past information. In the HERA study the photos were rated as the most helpful enhancement and the videos were rated third. Similarly, in the Davis study the videos and photos were ranked the second and third most helpful enhancements respectively. With the enhanced visuals' photos and videos being consistently ranked in the top three enhancements, it is clear that subjects appreciated the information these features provided. The Davis study further determined that including these enhanced visuals increased the trust subjects had in the procedure system.

While subjects in the Davis study did not report significantly different levels of trust with the five procedure types, in the HERA study crewmembers on average trusted the Base Procedure more than the Enhanced Procedure. Free response questions determined that subjects lost trust in the Enhanced Procedure when the sensor feedback provided unreliable step completion information due to sensor malfunctions. This highlights an important factor regarding embedded sensor feedback: it must be reliable and accurate. Providing incorrect feedback is not only unhelpful, but can damage subjects' confidence, making them doubt their performance and the procedure system itself. As time delays increase and astronauts receive less communication with MCC, feedback from the spacecraft itself will be of great assistance, but only if it is accurate and trusted. With astronauts living and depending on the spacecraft, all embedded sensor systems need to be reliable. Adding redundant sensors could help astronauts identify when sensors fail, and keeping sensors accessible could allow for easy replacement, however the best way to reduce weight and ensure consistency is to use high quality, reliable sensors.

Real-time sensor feedback is a powerful tool that subjects in the Davis study found consistently helpful. Out of all the enhancements, the sensor feedback was rated the most helpful despite it not significantly increasing task efficiency or accuracy. With 13/24 subjects noting it in their free responses, subjects consistently reported that they felt more confident in their step completion after the sensor feedback confirmed they completed the steps correctly. In comparison, subjects in the HERA study rated the sensor feedback as the least helpful enhancement. Previously mentioned sensor malfunctions likely played a significant role in this low rating with seven crewmembers noting in their free responses that they had an issue with unreliable sensor feedback at some point during their mission. Despite this, 6/16 subjects still noted in the free response section that they found the sensor feedback helpful as it increased their confidence. Sensor malfunctions were not present in the Davis experiment because there was continuous proctor oversight and maintenance available between trials ensuring consistent and fully functional equipment. Overall, the studies indicate that there is great potential for incorporating real-time sensor feedback into procedures. When feedback is consistently reliable, subjects find the enhancement helpful. However, caution must be taken to ensure that all embedded sensors are robust.

What was consistent between the HERA study and the Davis study was the impact procedural enhancements had on task completion accuracy. In both the HERA and Davis studies, subjects completed the task significantly more accurately with the Enhanced Procedure compared to the Base Procedure. Subjects completed the task 7.6% and 6.8% more accurately when using the Enhanced Procedure in the HERA and Davis studies respectively. Even more significantly, subjects completed the task 10.9% more accurately when using the Enhanced Procedure compared to the traditional PDF Procedure in the Davis study. In HERA, the largest difference in accuracy was observed in the first trial with subjects completing the task 16.9% more accurately with the Enhanced Procedure. With both studies concluding enhanced procedures improve accuracy, this thesis strongly supports the continued use of enhanced procedures to improve task performance, especially for first time procedure execution.

Further analysis of the Davis study determined it was the step navigation and enhanced visuals enhancements which improved accuracy. Subjects performed the task 9.7% and 8.1% more accurately when provided with the step navigation and enhanced visuals respectively. The enhanced visuals decreased the number of sequential, execution, partial omit, extra action and interrupt deviations. While the sequential and extra action deviations relate primarily to errors in step order, the execution, partial omit, and interrupt deviations are directly associated with successful step completion. Enhanced visuals was the only enhancement that reduced the number of execution and interrupt deviations indicating it is the only enhancement that truly improved subjects understanding of the task. While sensor feedback can help confirm step completion, only the enhanced visuals and laser guidance enhancements provide additional step information to aid workers in initial step completion. Therefore, while all the enhancements can improve task performance, these features have the most potential to increase crew autonomy as shown in how enhanced visuals decreased the number of interrupt deviations in the Davis study.

It is important to note there is a difference between asking for help and needing help. In the Davis study, subjects asked for help (assist deviation) at a consistent rate independent of enhancements or procedure. Subjects were interrupted less however (stopped from inappropriately handling equipment) when they were provided with enhanced visuals. In the HERA study the interrupt deviation was not present because there was no real-time proctor oversight. The number of assist deviations however was shown to be dependent on procedure type with 64% of crewmembers starting the task with the Base Procedure asking for assistance compared to 20% of the crewmembers who started with the Enhanced Procedure. Combining the results of these two studies shows that enhanced procedures can increase how autonomously workers are able to complete novel tasks.

Looking at step navigation in the Davis study, there was an almost 10% increase in accuracy correlated with the addition of a simple step highlighter and navigation buttons. As previously mentioned, the step tracker reduced the number of sequential, fragmented, omitted, and extra action deviations in the Davis study, proving how the feature can help workers complete tasks more accurately. Additionally, the step navigation system also significantly decreased subjects' workload.

Despite these task performance benefits and 15 subjects noting they liked the step navigation (compared to 4 who wrote they did not), step navigation was rated as the least helpful enhancement in the Davis study. Interestingly however, procedures that had step navigation were consistently rated as more helpful than procedures that did not. Comparatively, in the HERA study crewmembers rated step navigation as the second most helpful enhancement. While experimental analysis makes the benefits of step navigation clear, workers may not appreciate these benefits in the moment. Many workers tend to feel confident that they can complete procedures without missing steps and yet step/part-omissions continue to be a significant cause of aviation accidents (49). By implementing and encouraging workers to use step navigation features, we can increase how accurately workers complete tasks. Importantly, the step navigation system also had no negative impacts. While some subjects did not find the navigation as intuitive as traditional scrolling, data shows that the feature did not increase task completion time or change subjects' level of trust in the procedure system.

Additionally, in the Davis study the proctor observed that when subjects did not have the step navigation, subjects jumped around the procedure more. With the PDF Procedure it was common for subjects to scroll far forward in the procedure, looking to see how much longer the task was and reading future steps. The step navigation discourages this scrolling action by using navigation arrow buttons to move throughout the procedure. By limiting scrolling, we are limiting the amount of information the worker has access to. This can be beneficial to prevent distractions but must not go so far as to blind the worker from the main goal of the procedure. With the step navigation enhancement, subjects could still scroll through the procedure without moving the blue step indicator (and did occasionally when using the Base Procedure to review images from past steps), however subjects rarely scrolled forward to look at future steps.

When using the EPV, the blue highlighted step is shown towards the top of the screen. By having the current step highlighted, subjects attention can be consistently drawn to the correct step which helps subjects avoid missing steps. Depending on how long the step instructions are, the next step is usually shown on the screen as well. Subjects in both the HERA and Davis studies noted that they liked being able to see this "next" step because it provided them with insight into where the procedure was going. This level of information (current and future step) seems to be ideal because it provides the worker with some insight into the task progression without overwhelming them with currently irrelevant information. Subjects in both the HERA and Davis studies also consistently noted they liked being able to zoom in on pictures. This is another simple feature that should be included in all electronic procedures. Overall, both the HERA and Davis studies indicate that enhanced procedures should continue to be developed because they have great potential to increase task performance.

5.1 Potential Downsides of Enhanced Procedures

The purpose of this thesis is to identify how multimodal procedural enhancements could be used as a countermeasure to reduce the impacts of communication time delays between astronauts and MCC. While enhanced procedures have enormous potential to increase crew autonomy, there are potential points of failure that they introduce as well. Generally, in engineering when the complexity of a system increases, more failure points also develop which can make it more difficult to identify and fix problems.

This exact scenario occurred in the HERA study when there were embedded sensor malfunctions. As previously discussed, when the sensors failed they stopped providing reliable feedback. As a result, the crew originally became confused thinking they were making procedural mistakes which slowed down their task progression as they tried to identify the issue. Eventually, the crew came to realize that the procedure system was providing incorrect feedback which resulted in them losing trust in the procedure system. This loss of trust is a major issue because it impacts how the crew interacts with the procedure system. Rather than finding the system a beneficial source of aid and assurance, the system can instead become an unreliable nuisance with feedback the crew can't trust. In the context of this experiment's generator repair task, this loss of trust is problematic, however in the context of a spacecraft that astronauts are relying on to keep them alive, this loss of trust could have deadly consequences.

As such, it could be argued that it is better to have no sensor feedback rather than unreliable sensor feedback. With all of the technology available today however, a better solution would be to ensure that any sensors used are reliable and robust enough to last the entirety of the long duration mission. Short of this, sensors could be redundant or easily replaceable to allow for in-mission maintenance. This would still require astronauts to identify faulty sensors but would at least grant astronauts the opportunity to fix them, thus hopefully allowing for trust repair. Studies looking into the trust repair in human-machine interactions has taken off in recent years thanks to the developments in AI but significant work still remains (56). As this work continues to develop, it's lessons could be incorporated as another way to mitigate any issues that arise with enhanced procedures.

Enhanced multimodal procedures additionally have the same issues current day electronic procedures have. Unlike paper procedures, electronic procedures rely on a technological infrastructure. The procedures are stored as files which could be corrupted. The files are opened by computers which could crash or generally break. Procedure access and uploads depend on a reliable network. Despite these potential failure points, the ISS has mostly moved away from paper procedures and successfully used the electronic IPV system for years. A similar evolution to enhanced multimodal electronic procedures seems like the obvious next step because, while it might introduce more points of failure, it also adds powerful tools to allow for increased astronaut autonomy.

5.2 Recommendation Summary

The following are recommendations for future electronic multimodal procedures:

1. **Step Images**

   An image should be included for as many steps as possible.

   [*Rationale: In the Davis study, enhanced images were found to be the most impactful enhancement, improving all aspects of task performance. Enhanced images statistically increased subjects' task efficiency and accuracy, and increased their trust in the procedure system. Because the step images provided so much assistance to workers, it is recommended that an image should be included for as many steps as possible even if that requires repeating pictures. Procedure writers should not assume workers remember past information because this can force workers to either guess what is required or review past information which can decrease accuracy and efficiency respectively.*]

2. **Image Zoom**

   All images should be capable of zooming.

   [*Rationale: Subjects in both the Davis and HERA studies mentioned they appreciated being able to zoom in on pictures to prevent eye strain and better study the images. When writing electronic procedures, adding zoom functionality to images is an easy addition that should always be incorporated to improve user comfort.*]

3. **Steps Displayed on Screen**

At least two steps should be displayed on the procedure screen: the highlighted current step and the next step.

[*Rationale: Subjects in both the Davis and HERA studies mentioned they appreciated having the following step shown on screen below the current highlighted step. Displaying at least two steps allows workers to see the "before and after" of the current step to give them an idea of where the procedure is heading, satisfying their natural curiosity of what is coming up next. Caution should be taken when providing more than two steps on screen however because this can detract attention away from the current step. It is important for procedures to display enough information on screen for workers to feel confident without providing an overwhelming amount that distracts from the current objective.*]

4. **Step Highlighter**

   Step highlighters should be required when using electronic procedures.

   [*In the Davis study the step navigation enhancement statistically reduced the number of sequential, fragmented, omitted, and extra action deviations. This indicates the enhancement decreased how often subjects either missed steps or completed steps out of order, increasing overall task accuracy by 9.7%. Because of this significant increase in task performance, workers should be strongly encouraged, or required, to use step highlighters to stay on track when moving through electronic procedures.*]

5. **Optional Enhancements**

   It is important to keep enhancements "optional" (other than the step highlighter).

   [*Rationale: Enhancements are designed to enhance workers' abilities to complete tasks so it is important to keep enhancements "optional" (other than the step highlighter) to account for the range of aid different workers need. By making enhancements optional, workers unfamiliar with the procedure can choose when they require extra assistance while workers who are familiar with the procedure are not forced to interact with unnecessary enhancements. In the HERA study, crewmembers reported that the Enhanced Procedure was (or would have been) the most helpful during the earlier runthroughs of* *the task when the procedure was new. After the crew became familiar with the task, around half of the crew preferred the Base Procedure over the Enhanced Procedure because they felt it allowed them to "complete the task faster". While this impression of increased task efficiency with the Base Procedure was common, the data on task completion time did not show any decrease in efficiency associated with the Base Procedure. Nevertheless, this feedback supports the continued use of optional enhancements to avoid slowing down (or giving the impression of slowing down) workers who are familiar with the task. All enhancements should be available as extra assistance workers can use if needed without distracting those who do not require it.*]

6. **Feedback Reliability**

   Sensor feedback provided to workers should be reliable.

   [*Rationale: Real-time feedback can improve workers confidence when completing tasks, however this feedback must be reliable to ensure workers can trust it. In the HERA study there were two missions where sensor malfunctions resulted in inconsistent sensor feedback readings. The crew in these missions reported that this unreliable feedback slowed them down and made them question if they were completing steps correctly. This greatly impacted the crew's opinion of the procedure system and decreased their trust in the Enhanced Procedure. To ensure that workers do not face this issue, all embedded sensors must be robust enough to consistently provide reliable feedback.*]

7. **Analyzing Procedure Effectiveness**

   Both subjective opinions and quantitative metrics should be considered when analyzing procedure effectiveness.

   [*Rationale: When analyzing procedure effectiveness, it is important to look at actual performance metrics, such as efficiency and accuracy, as well subjective opinions. As shown in the HERA study where subjects did not recognize their improved performance with the Enhanced Procedure, sometimes subjective opinions do not align with quantitative data. To ensure accurate procedural effectiveness analysis, both quantitative data and subjective opinion should be considered.*]

Chapter 6

Conclusion

6.1 Summary

The effects of enhanced multimodal procedure countermeasures on workers' task efficiency, accuracy, and workload was investigated. Two experiments were conducted; the NASA JSC HERA study identified how electronic multimodal procedures affected task performance in an isolated spacecraft analog by comparing two procedure types: a Base Procedure and an Enhanced Procedure. The Davis study expanded this research to further identify how each individual enhancement (step navigation, enhanced visuals, sensor feedback, and laser guidance) affected task performance. The following conclusions will compare the initial hypotheses to the studies' results.

In the HERA study we hypothesized that when subjects used the Enhanced Procedure, compared to the Base Procedure:

- *Hypothesis: Participants would complete the task at the same rate; while the enhancements might increase situational awareness, they might also require extra time to interact with.*

  Analysis determined that, on average, there was no statistical difference in task completion time dependent on procedure type. There was however an interaction effect between procedure and trial; in trial 1 crewmembers completed the task significantly faster (12.5 minutes faster) when using the Enhanced Procedure compared to the Base Procedure.

- *Hypothesis: Participants would complete the task with greater accuracy due to the enhancements providing additional relevant task information.*

  Crewmembers consistently completed the repair task with a greater accuracy when using the Enhanced Procedure compared to the Base Procedure. On average crewmembers completed the task 7.6% more accurately with the Enhanced Procedure and in trial 1 crewmembers completed the task 16.9% more accurately.

- *Hypothesis: Participants would report a decreased subjective workload as the enhancements reduce the difficulty of locating parts and increase confidence by confirming adequate step completion.*

  There was no statistically significant difference in the reported workload or confidence of crewmembers dependent on procedure type.

- *Hypothesis: Participants would report the same level of trust in the system.*

  Crewmembers reported a significantly greater level of trust in the Base Procedure compared to the Enhanced Procedure. Crew feedback indicated that this decrease in trust in the Enhanced Procedure was caused by unreliable sensor feedback.

In the Davis study we hypothesized that:

- *Hypothesis: When subjects used the enhanced visuals, they would complete the task more accurately and efficiently as it would be easier to locate the parts mentioned in each step.*

  Subjects completed the repair task 4.4 minutes more efficiently and 8.1% more accurately when they completed the task with enhanced visuals. These are both statistically significant improvements on subjects' task performance.

- *Hypothesis: When subjects used the real-time sensor feedback and laser guidance they would report a higher subjective confidence during task execution as the enhancements confirmed adequate step completion.*

While subjects reported an increased confidence associated with real-time sensor feedback and laser guidance in free response questions, there was no statistically significant difference in subjects' self-rated quantitative confidence levels.

- *Hypothesis: When subjects used the step navigation, they would complete the task more accurately and efficiently as the navigation reduced missed steps.*

  The step navigation had no effect on subjects' task efficiency. Subjects did however complete the task 9.7% more accurately when they used the step navigation enhancement.

- *Hypothesis: Subjects would report a decreased subjective workload when using either the enhanced visuals, sensor feedback, laser guidance, or step navigation.*

  Both the enhanced visuals and step navigation significantly decreased subjects' perceived workload while the sensor feedback and laser guidance did not.

- *Hypothesis: Subjects would not report a change in their level of trust in the system dependent on any of the enhancements.*

  Neither the step navigation, sensor feedback, nor laser guidance changed subjects' trust in the procedure system. The enhanced visuals however improved how much subjects trusted the procedure.

In summary, enhanced multimodal electronic procedure countermeasures were found to significantly improve subjects' task performance when completing complex manual repair tasks. The HERA study proved through the increase in task completion accuracy and the reduction in assist deviations that multimodal procedures can increase crew autonomy when completing procedures in isolated environments. The Davis study explored each enhancement separately to identify how they independently affect task performance. Step navigation increased subjects' task completion accuracy and decreased their perceived workload. Enhanced visuals increased subjects' task completion time, increased task completion accuracy, and increased subjects' trust in the procedure system. Real-time system embedded sensor feedback and laser guidance did not significantly improve task performance but were consistently noted in free response questions as helpful enhancements that increased subjects' confidence. These are exciting conclusions which support the use of multimodal electronic procedure countermeasures to increase crew autonomy in future LDEMs.

6.2 Future Work

Upon completion of this thesis, both the HERA and Davis studies are being expanded to test more procedural enhancements. A new feature of the Enhanced Procedure is haptic feedback, technology that uses physical stimuli, in this case vibrations, to transmit information. Tactile interfaces can be especially helpful when workers are overwhelmed with visual or aural information because it translates information using an alternate sensory channel. In this application, haptic feedback is relayed through a wrist based cuff which alerts subjects to incorrectly completed steps (determined through sensor feedback) as well as caution statements. Different vibration pulses are used to differentiate between the two warning types with a fast rapid pulse indicating an incomplete step and a slower buzz alerting subjects to a caution statement. The wiring diagram for the haptic feedback cuff is shown in Figure 6.1. The cuff is directly connected to the Raspberry Pi for both power and data.

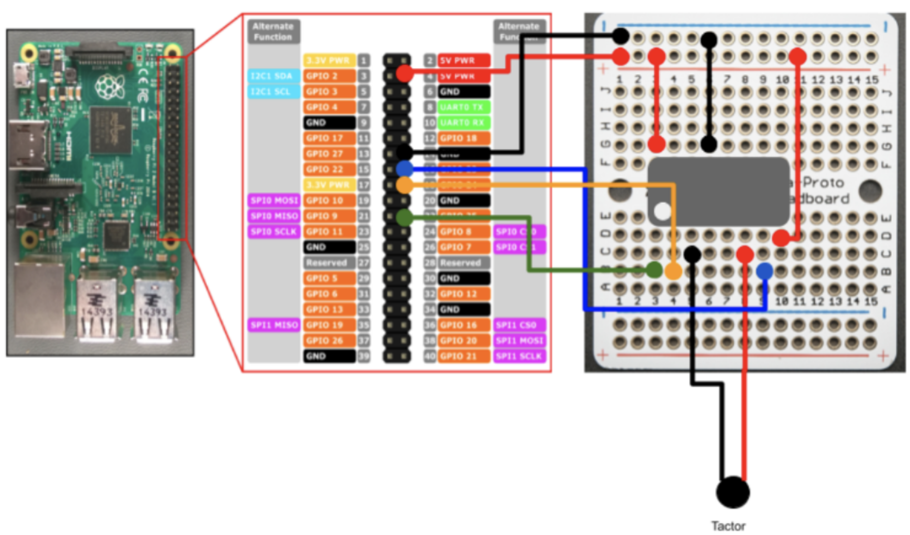


Figure 6.1: Haptic feedback cuff wiring diagram.

Another exciting new enhancement being tested is augmented reality (AR) visual displays. A tablet based Augmented Reality Procedure with live-camera visual overlays using an Apple iPad has been developed. Other researchers have previously studied AR procedures using head mounted displays and received negative feedback from users regarding the comfort, intuitiveness, and usability of the head mounted format. In contrast, our team has developed a tablet-based AR procedure which has the benefit of being more familiar and intuitive to most people compared to a device like the Microsoft HoloLens. An example of the tablet screen during the task is shown in Figure 6.2. On the AR procedure display the written instructions are at the bottom of the screen with the up and down navigation arrows on the left. The live camera feed utilizes most of the screen and each step overlays the relevant AR visuals onto the live image of the generator. Various animations and moving symbols are used to translate information such as unscrewing bolts and removing tubes. These AR overlays can be accurately oriented to the generator by using either fiducial markers or a point cloud scan of the generator; further research is exploring the benefits of each method.

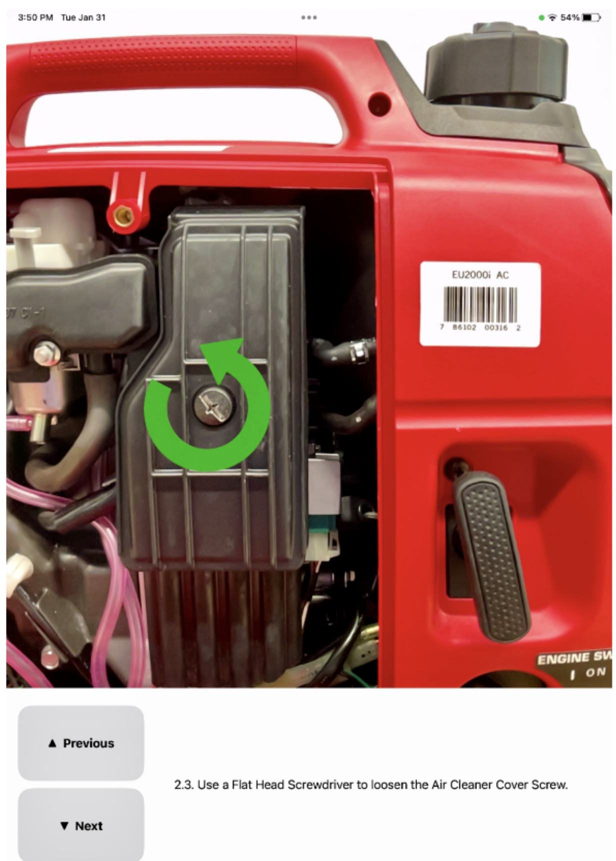


Figure 6.2: Augmented Reality Procedure with live camera feed, navigation buttons, written procedure, and AR overlay.

In order to use the AR tool hands free, the tablet can be held in place with a table mount allowing for free hand movement to complete the manual repair task. An example experimental setup of the tablet based Augmented Reality Procedure is shown in Figure 6.3. This new AR Procedure will be tested and compared to the previously studied 5 procedure types.

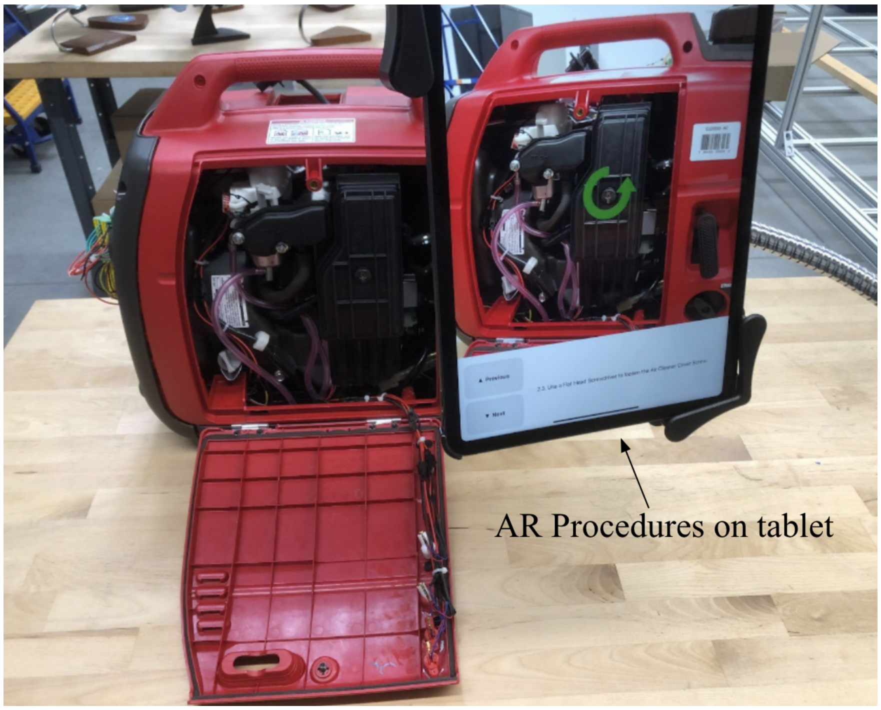


Figure 6.3: Augmented Reality Procedure system setup with tablet mount for hands free use.

Both the haptic feedback and AR Procedure are also currently (2024) being tested in NASA HERA Campaign 7. In addition to the new enhancements, eye tracking via the Pupil Core Eye Tracker has been added to better understand how subjects are using the procedure system. Analyzing subjects' scan patterns is expected to provide valuable insights into the subject's gaze and visual attention as they work through the task.

Future work could also continue to explore how interactive feedback from system embedded sensors affects task performance. The results of this thesis support the use of embedded sensors for real-time feedback, though the study was hindered by the number of sensors which could be placed into the COTS Honda generator. Future work could expand on this research by designing a new procedure and hardware system with embedded sensors in mind. By increasing the number of steps which have sensorized feedback, the task performance effects could be increased to better understand how feedback affects performance.

Bibliography

[1] J. Karasinski, I. Toron Valverde, H. Brosnahan, J. Gale, R. Kim, M. Yashar, and J. Marquez, "Designing Procedure Execution Tools with Emerging Technologies for Future Astronauts," *Applied Sciences*, February 2021.

[2] A. Valinia, J. Allen, D. Francisco, J. Minow, J. Pellish, and A. Vera, "Safe Human Expeditions Beyond Low Earth Orbit (LEO)," *Technical Memorandum NESC-RP-20-01589*, February 2022.

[3] J. Marquez, J. Karasinski, and J. Zheng, "How the Internet of "Space" Things will Disrupt and Transform Astronaut Work-Life," SpaceCHI, 2021.

[4] U. Muellerschkowski, "Complex operations on the International Space Station training and execution in manned and unmanned situations," ESA and Lufthansa Flight Training, October 2013.

[5] T. Panontin, S. Wu, M. Parisi, K. McTigue, and A. Vera, "Human-systems integration architecture needs analysis: On-board anomaly resolution during autonomous operations," Human Research Program Technical Report, 2021.

[6] K. McTigue, M. Parisi, T. Panontin, S.-C. Wu, and A. Vera, "Extreme Problem Solving: The New Challenges of Deep Space Exploration," SpaceChi: Human-Computer Interaction for Space Exploration, May 2021.

[7] B. Drake, *Human Exploration of Mars Design Reference Architecture 5.0.* Technical Report NASA-SP-2009-566, July 2009.

[8] D. Miranda, "2020 NASA Technology Taxonomy," HQ-E-DAA-TN76545, January 2020.

[9] Computer Security Resource Center, "Countermeasures," tech. rep., National Institute of Standards and Technology, 2024.

[10] A. D'Angelo, D. Rutherford, R. Ray, S. Laufer, C. Kwan, E. Cohen, A. Mason, and C. Pugh, "Idle time: an underdeveloped performance metric for assessing surgical skill," *The American Journal of Surgery*, 2015.

[11] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, 1988.

[12] S. Hart, "NASA Task Load Index (TLX)," *Human Performance Research Goup, NASA Ames Research Center*, 1988.

[13] L. Spirkovska, "Summary of Tactile User Interfaces Techniques and Systems," Ames Research Center, NASA/TM-2005-213451, January 2005.

[14] S. Mehrotra, C. Degachi, O. Vereschak, C. Jonker, and M. Tielman, "A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges," *ACM Journal on Responsible Computing*, 2024.

[15] J. Jian, C. Drury, and A. Bisantz, "Foundations for an Empirically Determined Scale of Trust in Automated Systems," *International Journal of Cognitive Ergonomics*, March 2020.

[16] R. Phillips and L. Wang, "The nasa constellation program procedure system," *American Institue of Aeronautics and Astronautics*, 2010.

[17] U. Brauer, F. Plassmeier, P. Kiernan, M. Wolff, and M. Martignano, "Authoring and viewing of electronic ISS crew procedures," *ESA Scientific and Technical Memoranda*, 2003.

[18] L. Landon, K. Slack, and E. Salas, *Psychology and Human Performance in Space Programs.* CRC Press, October 2020.

[19] J. Ingalls and J. Cipolletti, "Standards and Specifications for Ground Processing of Space Vehicles: from an Aviation-based Shuttle Project to Global Application," *American Institute of Aeronautics and Astronautics*, 2011.

[20] M. Martignano, M. Wolff, U. Brauer, and P. Kiernan, "Software assisted authoring and viewing of ISS crew procedures," *Acta Astronautica*, 2007.

[21] United Space Alliance, "Space Shuttle Flight Data File Standards and Specifications," *Space Flight Operaion Contract*, June 2006.

[22] R. Brown, E. Braunstein, R. Brunet, R. Grace, T. Vu, D. Zimpfer, W. Dwyer, and E. Robinson, "Timeliner: Automating Procedures on the ISS," *NASA Johnson Space Center*, 2002.

[23] B. O'Hagan and A. Crocker, "Fault Management Techniques in Human Spaceflight Operations," *NASA Johnson Space Center*, 2006.

[24] J. Karasinski, J. Zheng, M. Yashar, and J. Marquez, "Integrating Mission Timelines and Procedures to Enhance Situational Awareness in Human Spaceflight Operations," *SpaceCHI*, 2022.

[25] E. Salas, S. Tannenbaum, K. Kraiger, and K. Smith-Jentsch, "The Science of Training and Development in Organizations: What Matters in Practice," *Association for Physcological Science*, 2012.

[26] J. Vlasblom, H. Pennings, J. van der Pal, and E. Oprins, "Competence retention in safety-critical professions: A systematic literature review," *Educational Research Review*, 2020.

[27] X. Wang, *Factors Influencing Knowledge and Skill Decay in Organizational Training: A Meta-Analysis.* PhD thesis, University of Oklahoma Graduate College, 2010.

[28] M. Foale, A. Kaleri, A. Sargsyan, D. Hamilton, S. Melton, D. Martin, and S. Dulchavsky, "Diagnostic Instrumentation Aboard ISS: Just-In-Time Training for Non-Physician Crewmembers," *Aviation, Space, and Environmental Medicine*, 2005.

[29] M. Aguzzi, R. Bosca, and U. Mullerschkowski, "Astronaut training in view of the future: A Columbus payload instructor perspective," *Acta Astronautica*, 2009.

[30] M. Ammar, A. Haleem, M. Javaid, S. Bahl, S. Garg, A. Shamoon, and J. Garg, "Significant applications of smart materials and Internet of Things (IoT) in the automotive industry," *Materials Today*, 2022.

[31] L. Lautmann and P. Gallimore, "Control of the crew-caused accident: Results of a 12-operator survey," *Boeing Airliner*, 1987.

[32] D. Gugelmin-Almeida, L. Tobase, T. Polastri, H. Peres, and S. Timerman, "Do automated real-time feedback devices improve CPR quality? A systematic review of literature," *Resuscitation Plus*, 2021.

[33] S. Snodgrass, D. Rivett, V. Robertson, and E. Stojanovski, "Real-time feedback improves accuracy of manually applied forces during cervical spine mobilisation," *Manual Therapy*, 2010.

[34] S. Swinnen, T. Lee, S. Verschueren, D. Serrien, and H. Bogaerds, "Interlimb coordination: Learning and transfer under different feedback conditions," *Human Movement Science*, 1997.

[35] K. Sultanabanu, S. Liyakat, and K. Kazi, "Integrating IoT and Mechanical Systems in Mechanical Engineering Applications," *Journal of Mechanical Robots*, 2023.

[36] M. Pech, J. Vrchota, and J. Bednář, "Predictive Maintenance and Intelligent Sensors in Smart Factory: Review," *Sensors*, 2021.

[37] E. Ranjbar, R. Sedehi, M. Rashidi, and A. Suratgar, "Design of an IoT-Based System for Smart Maintenance of Medical Equipment," *International Conference on Internet of Things and Applications*, 2019.

[38] P. Bedi, S. Goyal, J. Kumar, and C. Shweta, "Smart Automobile Health Monitoring System," in *Multimedia Technologies in the Internet of Things Environment, Volume 2*, 2022.

[39] G. Coulby, A. Clear, O. Jones, and A. Godfrey, "Low-cost, multimodal environmental monitoring based on the Internet of Things," *Building and Environment*, 2021.

[40] L. Zhang, S. Chen, H. Dong, and A. Saddik, "Visualizing Toronto City Data with HoloLens: Using Augmented Reality for a City Model," *IEEE Consumer Electronics Magazine*, 2018.

[41] P. Phupattanasilp and S.-R. Tong, "Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming," *Sustainability*, 2019.

[42] P. Parsafar and P. Baben, "Real-time fault diagnosis system for electrical panels using embedded systems," *Journal of Electrical Systems and Information Technology*, 2023.

[43] Tencom Limited, "Embedded Sensors Help Monitor Composite Materials," 2023.

[44] J. Karasinski and J. Gale, *An Augmented Reality/Internet of Things Prototype for Just-in-time Astronaut Training.* Lecture Notes in Computer Science, May 2017.

[45] T. Boroushaki, M. Lam, W. Chen, L. Dodds, A. Eid, and F. Adib, "Exploiting Synergies between Augmented Reality and RFIDs for Item Localization and Retrieval," *Massachusetts Institue of Technology*, 2023.

[46] Bally Ribbon Mills, Luna Innovations Inc., "Inflatable Space Habitats Use Sensors Embedded in Webbing for Structural Health Monitoring," *Engineering Solutions for Design and Manufacturing*, 2022.

[47] P. Myers III, "Commercial Aircraft Electronic Checklists: Benefits and Challenges (Literature Review)," *International Journal of Aviation, Aeronautics, and Aerospace*, 2016.

[48] D. Boorman, "Safety Benefits of Electronic Checklists: An Analysis of Commercial Transport Accidents," *International Symposium on Aviation Psychology*, 2001.

[49] H. Rashid, S. Place, and G. Braithwaite, "Investigating the investigations: a retrospective study in the aviation maintenance error causation," *Cognition Technology and Work*, 2012.

[50] E. Erdfelder, F. Faul, and A. Buchner, "GPOWER: A general power analysis program," *Behavior Research Methods*, 1996.

[51] R. Adams, D. Klowden, and B. Hannaford, "Virtual training for a manual assembly task," *Haptics-e*, 2001.

[52] P. Linnerooth, D. Houlihan, M. Lenz, and J. Buchanan, "A Comparison of Covert and Videotape Modeling Strategies for Training Complex Mechanical Assembly Tasks," *Journal of Cognitive Engineering and Decision Making*, 2014.

[53] NASA, *Sample Size Specification Guidelines for NASA Human Research Studies.* NASA.

[54] S. O'Meara, K. Jenks, C. Stevens, and S. Mindock, Jennifer Robinson, *Methodology to Quantify Procedure Execution Analysis.* Engineering Psychology and Cognitive Ergonomics, 18th International Conference, July 2021.

[55] NASA, "HERA," January 2024.

[56] C. Esterwood, "Rethinking Trust Repair in Human-Robot Interaction," *Computer Supported Cooperative Work and Social Computing*, 2023.