Improving Usability for Novices in the Design of Mechatronic Devices: A Study Using Arduino Modules

By

Levi Mark Struyk Huyer

A thesis submitted to the Graduate Program in Mechanical and Materials Engineering

In conformity with the requirements for the

Degree of Master of Applied Science

Queen's University

Kingston, Ontario, Canada

February, 2024

Abstract

Due to the complexity of mechatronic system designs, research was conducted with the objectives of identifying issues encountered by novice users while using and learning to use these systems and providing recommendations on how to improve mechatronic system design. As part of the research, a networked mechatronic system that utilized Arduino Mega microcontrollers and custom printed circuit boards was created. A study was then designed using this mechatronic system with the intention of eliciting feedback relevant to the study objectives. To investigate the research objectives, three research questions were identified. These questions investigated what concepts inherent to mechatronic design were identified by novices, what effect real-time feedback had on mental workload, and whether effort was the highest contributor to mental workload. The study provided participants with a scenario where they were to set up the mechatronic system in an 'autopilot' configuration on a model representative of an aircraft. To address the second research question, two instruction sets, or manuals, were created: a control set providing delayed feedback and an experimental set providing instantaneous feedback. The experiment was conducted with 16 student participants identified as novice users, split evenly into a control group and an experimental group based on participant number. Eleven of the participants were from an engineering background and five were from a non-engineering background. Mental workload was measured during the experiment using the raw NASA task load index. Participant information was taken during a pre-experiment questionnaire and responses were recorded during a post-experiment interview and debriefing session. The study found that mental workload was affected by the system designer's ability to communicate information through their design, and that most issues identified by novices were deficiencies in clarity, simplicity, transparency, consistency, and/or context. Similarly, recommendations on how to improve designs for novice users related to the same five concepts. Additionally, it was identified that designers should use clear and concise universal language in all

communication and include accurate pictorial references to enhance user understanding and visualization.

Table of Contents

Abstract i

Table of Contents iii

List of Tables ix

List of Figures x

List of Abbreviations and Symbols xi

Chapter 1: Introduction 1

Chapter 2: Literature Review 3

2.1 Designing for Novices 3

2.2 NASA-TLX in Aerospace Research 5

2.3 Evaluating Arduino in Aerospace 8

2.4 Data Collection and Analysis Methods 10

2.5 Objectives and research questions 15

2.5.1 First Research Objective 15

2.5.1.1 Research Question 1 15

2.5.1.2 Research Question 3 16

2.5.2 Second Research Objective 17

2.5.2.1 Research Question 2 17

2.6 Research Methods 17

Chapter 3: Hardware Design and Integration 19

3.1 Problem Definition...19

3.2 Serial Communications Overview: SPI and I2C 21

3.3 Transfer of Information..22

3.4 Proof of Concept: Modular System Prototype..23

3.5 PCB Shield Overview 24

3.6 Future Design Work for the Mechatronic System 25

3.6.1 Future Design Work: Serial Addressing...25

3.6.2 Future Design Work: EEPROM Implementation 26

3.6.3 Future Design Work: IMU Redundancy 27

Chapter 4: Study Design...29

4.1 Physical Setup 29

4.2 Software for Experimental Testing 30

4.3 Hierarchical Task Analysis 34

4.4 SHERPA Method Error Identification 38

4.5 Instruction Set Design 40

4.5.1 Control Group Instruction Set versus Experimental Group Instruction Set...........................41

4.6 Application of NASA-TLX 45

4.6.1 Demographic Questionnaire 45

4.6.2 Post-Experiment Interview Questions 46

Chapter 5: Experimental Protocol and Methods 48

5.1 Participants 48

5.2 Participant Testing Methods 49

5.3 Pilot Study 50

5.4 Experimental Procedure 53

Chapter 6: Experimental Results and Analysis 54

6.1 Mental Workload Scores 54

6.1.1 Statistical Analysis of NASA-TLX Scores 56

6.1.2 Overall Results from the NASA-TLX MWL Survey 58

6.2 Measured Experimental Success 59

6.3 Participant Interview Responses 60

6.3.1 Attempting to Minimize Observer Bias 64

6.4 Analysis of NASA-TLX Subsections 65

6.4.1 Subsection ANOVA Results 65

6.4.2 Subsection Omittance Investigation 65

6.4.2.1 Statistical Results from Omittance Investigation 67

6.4.2.2 Physical Demand 68

6.4.2.3 Temporal Demand 69

6.4.2.4 Performance 70

6.4.2.5 Analysis Between Perceived Performance and Measured Performance 71

6.4.2.6 Frustration 73

6.4.2.7 Conclusion to Omittance Investigation 74

Chapter 7: Discussion 75

7.1 Issues observed by novices (RQ1) 75

7.1.1 Consistency 77

7.1.2 Context 79

7.1.3 Issues Identified When Installing Mechatronic Components 81

7.2 Timing of Feedback on Mental Workload Results (RQ2) 81

7.2.1 Differences Between Immediate and Delayed Feedback 82

7.2.2 Power Analysis 83

7.3 Effort as a Main Contributor to Mental Workload (RQ3) 84

7.4 Limitations to the Use of the NASA-TLX in This Experimental Protocol 84

7.5 Relevance of Frustration and Performance Variability to Research Objectives 85

7.6 Recommendations 86

7.6.1 Primary Recommendations 86

7.6.2 Recommendation for Communication 87

7.6.3 Recommendation for Visuals 88

Chapter 8: Conclusions and Recommendations 89

8.1 Research Questions 89

8.2 Research Objectives 90

Chapter 9: Limitations and Future Research 92

9.1 Limitations of the Results 92

9.1.1 Population Size 92

9.1.2 Demographic Distribution 92

9.1.3 Personal Factors 93

9.1.4 Environmental Factors 93

9.2 Future Research 94

9.2.1 Improvement of Data Collection Procedures 94

9.2.2 Improvement and Modifications to Experimental Setup 95

9.2.3 Improvement of Documentation and Instruction 96

9.2.4 General Recommendations for Future Research 97

References 99

Appendix A: PCB Schematics and Drawings 102

Appendix B: Combined SHERPA Output 104

Appendix C: Part List 107

Appendix D: Software Transcript 108

Appendix E: Participant Instruction Sets 116

Set A: Control Set (Powered After Setup) 116

Set B: Powered at Start of Setup 118

Appendix F: Data Collection Methods 120

Pre-experiment Questionnaire 120

NASA-TLX Survey..121

Post-experiment Interview 122

Appendix G: GREB Letter of Intent and Consent Form..123

Appendix H: GREB Approval 125

List of Tables

Table 1: Summary of background study articles. 11

Table 2: Breakdown of PCB components and their respective functions. 24

Table 3: Software breakdown. 31

Table 4: LED indicator communication and corresponding meaning. 32

Table 5: SHERPA external error mode taxonomy [5]. 39

Table 6: ANOVA results (5% significance level) with mean and variance values shown. 57

Table 7: Scored rubric measuring experimental success (actual performance). 59

Table 8: Summary of overall positive and negative responses to Q1. 61

Table 9: Summary of interview responses (orange = negative response and blue = positive response). 63

Table 10: MWL scores resulting from the removal of specific subsections. 67

Table 11: Responses sorted under clarity, simplicity, and transparency (italicized comments relate to additional concepts). 76

Table 12: Combined SHERPA output from Figure 3 and Figure 4. Empty rows indicate error mode was insignificant. 104

List of Figures

Figure 1: Flow diagram for device communication. 23

Figure 2: Schematic of the mechatronic system as it was used in the experiment.......... 33

Figure 3: HTA for the first approach where system is powered after installation. 36

Figure 4: HTA for the second approach where system is powered before complete installation. 37

Figure 5: Supplementary diagrams presented to participants during the experiment. 44

Figure 6: Experimental model used in the participant study.......... 50

Figure 7: Yellow square removed from final iteration of the experimental model.......... 52

Figure 8: Comparison of the NASA-TLX subsection data and overall MWL scores of the control and experimental groups. Subsection and MWL scores are unitless. 55

Figure 9: Comparison of NASA-TLX subsection data between participants of non-engineering background and engineering background. 56

Figure 10: Summary of statistical analysis using a 0.05 significance threshold (Note: performance p-value was equal to 1.0).......... 58

Figure 11: Perceived performance compared to experimental success.......... 62

Figure 12: Positive and negative responses to supplementary diagrams given to participants summarized from interview responses. 64

Figure 13: Summary of p-values obtained from omission ANOVAs using a significance threshold of 0.05. Category along x-axis represents primary subsection being omitted while column and colour indicate the secondary subsection being omitted. 68

Figure 14: Scatterplot showing the relationship between experimental success and perceived success. 72

Figure 15: Arduino PCB schematic. 102

Figure 16: Arduino PCB diagram. 103

List of Abbreviations and Symbols

ANOVA: Analysis of variance

CS: Chip select

EEPROM: Electrically Erasable Programmable Read-only Memory

GREB: General Research Ethics Board

HTA: Human Task Analysis

I2C: Inter-integrated circuit

IMU: Inertial measurement unit

POCI: peripheral out/controller in (formerly MISO: master-in-slave-out)

PICO: peripheral in/controller out (formerly MOSI: master-out-slave-in)

MWL: Mental workload

NASA-TLX: National Aeronautics and Space Administration Task Load Index

PCB: Printed Circuit Board

SCL: Serial clock line

SHERPA: Systematic Human Error Reduction and Prediction Approach

SPI: Serial peripheral interphase

Chapter 1: Introduction

Mechanical engineering has changed significantly during the last few decades with one of the most significant additions to the engineer's arsenal being mechatronics, or the combination of mechanical, electrical, and computer engineering. Many fields that were traditionally purely mechanical in nature are now being made more advanced thanks to mechatronics, including the field of aviation. Unfortunately, for those students and amateur tinkerers hoping to learn more about mechanical engineering and aerospace, mechatronic systems can be intimidating to install and use since these systems are often designed with experienced users in mind. To this end, two research objectives were defined. The first was to identify issues that novices encounter when interacting with mechatronic devices (*obj. 1*), and the second was to identify general design recommendations on how to intentionally design mechatronic systems for novice users (*obj. 2*).

The first step to improvement was understanding where novice users identify issues. This could be done using tools like the National Aeronautics and Space Administration's Task Load Index (NASA-TLX). This subjective workload index was designed by NASA to identify areas of improvement to minimize cognitive load especially in aviation related areas such as cockpit environments; command, control, and communication workstations; and supervisory and control processes. Using the NASA-TLX as a basis for this study, three research questions were identified to achieve the research objectives:

RQ1. What aspects of mechatronic design are consistently identified by novices after interaction with a mechatronic system?

RQ2. Is mental workload as measured by the NASA-TLX reduced by real-time device feedback compared to delayed feedback?

RQ3. Is effort the biggest contributor to mental workload as measured by the NASA-TLX when installing a mechatronic system?

This thesis is laid out as follows: Chapter 2 includes the literature review, Chapter 3 outlines the design of the experimental hardware, Chapter 4 outlines the design of the research study, Chapter 5 includes the experimental protocol and methods use, Chapter 6 includes the experimental results and analysis, Chapter 7 includes the discussion of the results, Chapter 8 summarizes the conclusions and recommendations from the study, and Chapter 9 includes the limitations of the study and future research.

Chapter 2: Literature Review

2.1 Designing for Novices

Within human factors engineering, it is common practice to identify stakeholders at the beginning of any project, and novice focused design is no different. Unfortunately, research defining best practices when designing for novice stakeholders is scant; as stated by Dean Hooper, "in light of current understanding of the characteristics and mechanisms of novices through experts in problem solving through basic research, the field of human factors is relatively devoid of design principles that utilize this knowledge base [1]." However, research involving novice and expert interaction with designed systems is a well-established field. A review article examining the use of experts versus novices in usability testing found that of eight articles reviewed, six of the articles found that novices identified more credible usability issues than expert users [2].

Hooper [1] aimed to define design principles when creating user-centric systems for different levels of expertise. In this research, Hooper builds upon the Rasmussen Skills, Rules, and Knowledge (SRK) Model used to characterize human behavior and decision making [3]. SRK represents the three levels of "performance" as described by Rasmussen; skill-based, rule-based, and knowledge-based [3]. Using Hooper's generalizations, skill-based behavior is characterized by trained automatic responses, rule-based behavior is characterized by responses to pattern and state recognition, and knowledge-based behavior is guided by the user's knowledge and experience [1]. Hooper combines these performance levels with the transition between novice to expert: novices exhibit knowledge-based behavior by drawing parallels with the new system and their own knowledge, intermediate users exhibit rule-based behavior as they begin to learn the system, and experts exhibit skill-based behavior due to the length of time using the system [1]. Using empirical analysis and this modified Rasmussen SRK model, Hooper suggested general design principles for the different stages of expertise. The portion relevant to this research, that focusing on novices, states the following [1]:

- Task flow should be step-by-step (linear).
- Design should be useable with very little background knowledge of the domain.
- Novice users should be allowed to step through the tasks themselves with full transparency and visibility as to how the system reacts to user inputs.

A significant takeaway is that to be useable with little domain knowledge, the novice user must be informed on the device's use. This can be accomplished most easily using detailed instruction sets and intuitive design elements, such as limiting the options for the user. While for experts it is beneficial to present extraneous information to help improve the decision-making process, it can be inferred that for novices it is better to limit this information to avoid confusion [1].

One aspect that separates designing for experienced users and designing for novice users is how the instructions on how to use the design are communicated to the user. This is especially true in coding instruction, where the lack of clear communication can be a barrier when learning to program. Denny et al. address this problem with communication as identified in error messages in their research [4]. Though their results focus on issues concerning the readability of error messages, the article provides relevant information on designing novice focused systems, particularly the communication aspects of those systems. Their recommendations are summarized by the following [4]:

1. Remove as much jargon as possible without compromising the essence of the message.
2. Use complete sentences, including the explanation or definitions of symbols relevant to the message (i.e., instead of ";", use "semicolon (";")").
3. Use simple vocabulary for words that are not necessarily considered jargon to reduce confusion.

4. Use as few words as possible without sacrificing the essence of the instruction or message being conferred.

These two articles are not necessarily related to mechatronic system design, since Hooper's article focused on general system design, and the article by Denny et al. was only focused on designing forms of communication. However, these were considered relevant to designing for novices due to the importance of overall system considerations and the importance of communication in system design. Therefore, the two lists of design rules and recommendations were summarized under three concepts that can be applied when designing for novices:

- *Simplicity:* keep the task flow linear, use as few words as possible to educate the user on the system and ensure that simple vocabulary is used when doing so.
- *Transparency:* the system should clearly show how it gets from point A to point B, and how user input affects the system, allowing users to quickly understand the system.
- *Clarity:* minimal background knowledge should be needed to use the system. This includes avoiding the use of jargon and abbreviations.

From a design perspective, these concepts are a logical starting point for designing a beginner-friendly system. If a designer were to make something overly complicated and excessively use jargon, feedback from novice users would illustrate a steep learning curve due to the lack of simplicity and clarity. Likewise, if the designer were to focus on expert users and hide much of the system's processes, novice feedback would likely focus on their lack of understanding.

2.2 NASA-TLX in Aerospace Research

When evaluating or designing any task, especially in critical environments, measuring the mental workload (MWL) of those interacting with the system is a good way to identify user issues. According to N. Stanton [5], "MWL represents the proportion of [finite attentional capacity] demanded by a task or set of tasks.” Since MWL analysis is so important to the development of products and processes, there have been numerous methods developed to measure it. Stanton [5] provides details on fifteen different methods, including the Modified Cooper Harper Scale (MCH), the Subjective Workload Assessment Technique (SWAT), the Subjective Workload Dominance Technique (SWORD), and the NASA Task Load Index (NASA-TLX). Published by the NASA-Ames research center in 1988, the NASA-TLX is a widely used MWL assessment tools, and was the culmination of three years of research attempting to “identify the factors associated in subjective workload within and between different types of tasks [6].” The result was a subjective workload tool still used today. In 2006, the NASA-Ames research center published a survey of 550 articles published during the previous twenty years in which the NASA-TLX had been used [7]. This survey found that many of the studies had been funded by the US government in some form (roughly 40%). The survey also found that it had been effectively used across multiple industries unrelated to aerospace and had been translated and applied successfully across more than eight languages [7].

The NASA-TLX itself is a “subjective workload assessment tool that is used to gather subjective ratings of operator mental workload in man-machine systems [5].” It is a weighted survey with six different subscales: mental demand, physical demand, temporal demand, effort, performance, and frustration. Descriptions summarized from the original NASA-TLX publication are included below [6]:

1. Mental Demand: the level of mental and perceptual activity required. Relates to purely mental portions of the task such as deciding, calculating, searching, etc.
2. Physical Demand: the level of physical activity required. Relates to purely physical portions of the task such as pulling, pushing, etc.
3. Temporal Demand: the level of pressure from the rate at which the task was or needed to be performed.

4. Performance: the perceived level of success with the performed task and the personal level of satisfaction with that level of success.
5. Effort: the perceived amount of physical and mental activity required to achieve the level of performance reached with the task.
6. Frustration: the level of insecurity, discouragement, irritation, stress, and annoyance felt during the task.

According to Stanton [5], before the NASA-TLX is applied, the task under analysis is thoroughly defined and a hierarchal task analysis (HTA) is performed. The HTA is a method of describing the task being analyzed which results in an "exhaustive description of task activity" [5] that in terms of the NASA-TLX can be used to inform the use of the MWL assessment tool. The HTA methodology will be described in greater detail in section 4.3 Hierarchical Task Analysis.

The version of this survey used in this study can be found in Appendix F: Data Collection Methods [8] along with the questions relating to each subsection. In the original version of the NASA-TLX (released 1988), pairwise comparisons were included in which participants were asked to pick which of the factors had a greater contribution to the mental workload of the task. The number of times each category was picked was used to create a weighting factor for the final workload calculation. In the more recently developed NASA Raw Task Load Index (RTLX), this weighting factor is omitted. Studies have indicated that the effectiveness of the NASA-RTLX is similar to the NASA-TLX [7]. The raw NASA-TLX is further supported in its efficacy by research conducted by Said et al. [9] where it was validated across three studies during its application measuring MWL in patient monitoring tasks.

The research example that is most relevant to this project comes from the field of computer science, where the NASA-TLX was used to assess MWL in introductory computer science classes [10]. In the study, the NASA-TLX was conducted within 24 hours of introductory computer science laboratory assignments by student volunteers. Using this data, the perceived workload of the assignments was compared against demographic factors like race and gender to determine if perceived workload could be used to identify vulnerable groups. The perceived workload was compared to a measure of actual workload taken from analyzing aspects of the coding assignments to determine the relationship between perceived and actual workload. The study found that while perceived workload does not necessarily correlate with actual workload, it is an effective way to identify where there are usability issues with the system being studied [10]. For example, the researchers used the MWL results to identify potential trends among demographics which suggested that certain groups struggled more than others within the course being studied [10]. In the school setting in which the research was completed, this could be used by educators to modify their program to help these disadvantaged groups. The similarities between mechatronics/aerospace novices and introductory computer science students make this study’s results relevant to this research: Both groups are presented with the available tools and components needed to complete difficult tasks while having little background knowledge of the field. The similarities continue when considering the application of the NASA-TLX with the purpose of decreasing perceived and actual workload for novice users.

2.3 Evaluating Arduino in Aerospace

Arduino microcontrollers have become an attractive choice for small scale aerospace projects due to their low cost, usability, and versatility. These systems can be used to help students develop autopilot systems [11], or to quickly develop complex algorithms (i.e., for different autopilot pathing methods) [12]. The studies presented in this section show that for well over a decade, Arduino based systems have been used in aerospace research and design to great effect.

Microcontroller use, specifically Arduino, is prevalent among aerospace researchers and hobbyists alike. In this space, there are companies that cater specifically to these users. For example, Ardupilot is a company that produces microcontrollers using Arduino architecture that are built specifically for drone creation, incorporating outputs for servo motors and other commonly used components [13]. Other companies, such as Adafruit and Sparkfun, produce a plethora of electronic components adapted specifically for use with Arduino and similar low-cost microcontrollers. Using resources like these, researchers tackle a wide range of aerospace projects. For example, in a 2012 study, Ardupilot was used to develop a system for "high level control algorithm" [12] rapid prototyping. Their rapid prototyping method began with the simulation of Ardupilot hardware using software in conjunction with the commonly used flight simulator environment X-plane and algorithms developed using Simulink. The algorithms were then used to control a real-world model airplane system [12].

Similarly, a separate thesis published in 2012 explored the use of Arduino for use in a low-cost autopilot system developed for academic purposes [11]. Using Ardupilot as a base, the study identified the steps needed to fully develop a single vehicle and autopilot system from start to finish, including flight simulation using similar methods to Coombes [12]. The purpose of going through this process was to inform and develop a step-by-step process that could be applied to a learning environment [11].

Another similar study published in 2017 built upon the use of Arduino as a low-cost autopilot by adding a transistor array to a developed Arduino autopilot system [14]. This modification allowed for seamless transition between automatic and manual control, so that if the autopilot were to fail then the controls would revert to manual [14]. This system was then validated using three different vehicle types; a quadcopter, a fixed wing aircraft, and a ground vehicle [14]. Like the previous case study by B. Kissack [11], the autopilot system was developed by the researchers to be a generalized design that could be applied to different vehicle archetypes without major modification [14]. The similarities found between these studies in the use and efficacy of Arduino systems lends support to the microcomputer's use for aerospace purposes and research.

2.4 Data Collection and Analysis Methods

For many studies in the field of engineering, research results are solely supported by mathematical results and statistics. This kind of research lacks focus on human factors issues inherent to any system in which human interaction is present or inevitable, therefore, this project aims to answer the three identified research questions using responses from human participants. There are many methods of doing so, and the approach is an important part of improving user focused design.

Human factors research is an important field of study that helps engineers understand and improve how users interact with their work. From the review of articles in the human factors field performed in this study, it was found that the general research process followed these main steps:

1. Study definition: Research objectives, research questions, and target population are defined.
2. Data collection: Assessment methods are chosen and applied.
3. Data analysis: Data taken during the data collection phase of the research is analyzed, and statistical analysis is performed.
4. Conclusions: The research questions are answered using the analyzed results.

This study aims to follow this general research approach. While it is important to have a research plan in place, it is equally important to analyze research that has put these methods into practice to help inform the use of shared techniques. Table 1 summarizes five articles deemed relevant to this research that make use of methods used in this study such as the NASA-TLX and structured questionnaires:

Table 1: Summary of background study articles.

Study Title	Data Acquisition Method	Objectives	Study Methods
Assessing Workload Perception in Introductory Computer Science Projects using NASA-TLX (Naser) [10]	• NASA task load index. • One additional questionnaire obtaining responses related to the experiment.	• Assess overall workload of the computer science course under study. • Identify at risk students. • Improve the computer science course.	1. Data acquired using NASA-TLX and questionnaires. 2. Performed statistical analysis on results. 3. Interpreted results of statistical analysis. 4. Answered research questions using results from statistical analysis.
Using the Raw NASA Task Load Index Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: A Meta-Analysis Using a Mixed Model (Sadiq) [9]	• NASA raw task load index. • Demographic information obtained from institution.	• Validate new patient monitoring method. • Validate use of raw NASA-TLX for measuring MWL in clinical settings. • Analyze the effects of covariates on MWL in patient monitoring situations.	1. Obtained information from previous studies. 2. Performed statistical analysis on NASA-TLX scores and factors assumed to affect said scores. 3. Performed statistical analysis comparing old method vs new method being studied. 4. Drew conclusions from study results on validity of the raw NASA-TLX and efficacy of new method being studied.
On designing programming error messages for novices: Readability and its constituent factors (Denny) [4]	• Three separate experiments using questionnaires.	• Identify factors affecting readability of error messages encountered during coding. • Provide general guidelines for writing error messages to maximize readability.	1. Used an initial study to explore what factors affect message readability by asking participants to rate error message readability. 2. Used a second study to identify specific readability factors by asking participants what makes error messages shown to them easy or hard to read. 3. Identified four readability factors from post analysis of participant responses from the second study. 4. Used a third study to confirm the validity of the four factors found during the second study by asking participants to rank each error message from one to five for each of the four factors as well as for an additional understandability factor.

Usability Study for an Augmented Reality Content Management System (Siewert) [15]	• NASA raw task load index. • Separate questionnaire obtaining background information. • Separate questionnaire obtaining responses related to the experiment.	• Make the process of creating augmented reality content easier for educators with little to no experience in the field.	1. Study goals were defined and requirements for the system being designed as part of the study were defined. 2. Prototype application was created to handle the creation of augmented reality experiment manuals, store the manuals, and deliver the manuals to students. 3. An experiment was designed using the manual creation tool and administered to students. 4. The NASA-TLX was used to measure the mental workload of the students experience with the tool. A separate questionnaire was used to obtain background information relevant to the study. 5. The researchers drew conclusions on the ease-of-use of the software from the NASA-TLX scores and participant responses to questions related to ease-of-use.
Preliminary Study on a Novel Protocol for Improving Familiarity with a Lower-Limb Robotic Exoskeleton in Able-Bodied, First-Time Users (Lau) [16]	• System Usability Scale (SUS). • NASA Raw Task Load Index. • Two additional questionnaires to gauge user comfort level.	• Develop proper protocol and instructions to prepare users for lower-limb exoskeleton use.	1. The need for the study was established and a protocol for preparing participants for exoskeleton use was proposed. 2. One custom survey was given to participants before the study to capture comfort level with the device. 3. Participants were instructed to complete a series of tasks with the exoskeleton. To assess the effectiveness of the developed protocol, one group was given the protocol, and one group was not. 4. Participants completed the System Usability Scale (SUS), NASA-TLX, and a second custom survey to capture comfort level after the experiment.

			5. Results from SUS and the NASA-TLX were analyzed. 6. Conclusions were drawn from the analysis results.

The first takeaway from Table 1 is that the articles using the NASA-TLX as part of their data acquisition strategy used them alongside additional information obtained from the participants. This indicates that to obtain meaningful results when using the NASA-TLX method to obtain a measure of MWL, it is common practice to use additional information to make better informed conclusions based on those MWL scores. The second takeaway is that to identify guidelines, factors, or recommendations like in Naser [10], Denny [4], and Lau [16] (Table 1), it is important to obtain them using established tools, carefully targeted questions, and equally careful analysis of participant responses. The third takeaway is that research objectives can be broken across multiple studies like in Sadiq [9] and Denny [4] (Table 1), and that open ended research objectives may not be met in a single study. Therefore, it is important that a single study have a narrow focus to avoid producing convoluted results. Siewert [15] (Table 1) supports the efficacy of narrow focused research since the authors created limited and measurable objectives during the projects definition, design, and validation process. This allowed the study to reach its objectives using their outlined data acquisition methods.

Of the four studies utilizing the NASA-TLX (Table 1), Naser [10], Sadiq [9], and Lau [16] used direct scores for future statistical analysis, and Siewert [15] used relative scaling without any scoring (low or high, good or bad). Regarding the mathematical validity of statistical analysis performed on NASA-TLX scores, Bolton [17] suggests that the “level of measurement” obtained from the NASA-TLX tool is lower than assumed by those using the index for research. To elaborate, Bolton [17] breaks down level of measurement into four categories as suggested by Stevens [18];

> 1) Nominal scales represent identity or category (e.g., student identification number). 2) Ordinal scales capture order (e.g., the place one finishes in a race). 3) On interval scales (e.g., temperature in Celsius), the distances between numbers are meaningful. However, because there is no meaningful zero on an interval scale (zero does not indicate that none of the measured quantity exists), ratios between numbers are not meaningful. 4) Finally, ratio scales (e.g., length) have meaningful zeros and, thus, there is meaning between ratios of numbers on these scales. [17]

Results obtained by Bolton [17] suggest that scores obtained by NASA-TLX are at most on the interval scale or ordinal scale, depending on which of the six scores or whether the overall workload score is being analyzed. Bolton [17] uses this conclusion to claim that the NASA-TLX is "mathematically meaningless," but this is a bold statement. Looking behind the bombastic wording used in the study, the real conclusion from the study is that NASA-TLX scores should be used carefully in statistics since they are unique to an individual. This conclusion, while important in analyzing the validity of the NASA-TLX as a tool, does not impact its relevancy in human factors research. Regardless of its mathematical limitations, the NASA-TLX is a useful tool in identifying issues and trends in systems and tools that are often missed by purely mathematical results. Furthermore, the papers presented in Table 1 that used the NASA-TLX as part of the data collection and analysis process included additional data acquisition tools, indicating that the index's usefulness in measuring perceived or actual workload is dependent on additional information needed to inform the results. Therefore, though Bolton [17] does suggest that the NASA-TLX is mathematically limited, it also reinforces the importance of a human factors focus in engineering since it implies that the numbers obtained through study should be informed using additional human factors methods of analysis.

2.5 Objectives and research questions

2.5.1 First Research Objective

The first objective of this research, the issues objective, was to identify issues that novice mechatronics users encounter during their interactions with mechatronic devices. This first objective was established on the hypothesis that mechatronics has a steep learning curve, especially for novice users. Mechatronics is the combination of mechanical, electrical, and computer engineering, and since it takes years to develop proficiencies in even one of these fields, the combination of all three is uniquely difficult to learn. Any novice user needs to master topics from all three fields to become an expert in mechatronics. This hypothesis is strengthened by the complex methods used to teach the subject within university settings; reviewing course material used within mechatronics courses offered within Queen's University's Mechanical and Materials Engineering department reveals extensive use of complex, multidisciplinary, hands-on projects used to teach mechatronics concepts (see [19]). This assumption that the field has a steep learning curve is further supported by the existence of the multiple, full-term courses required to teach mechatronics (see [19] and [20]).

2.5.1.1 Research Question 1

The first research question sought to explore **what aspects of mechatronic design are consistently identified by novices after interaction with a mechatronic system** *(issues objective*, *obj. 1)*. This question was worded similarly to *obj. 1* but was phrased differently to obtain more open-ended responses; the word '*issues'* was replaced with '*aspects'* to make the question less leading for the researcher. Aspects in the context of this question refer to general ideas both positive and negative that relate to a mechatronic design or mechatronic design process. Past experiences and education can affect how people perceive information, therefore, this question sought to identify the elements of mechatronic systems that may contribute to understanding. Furthermore, the question sought to identify to what extent user experience may influence their interactions with mechatronic systems. This question was necessary to meet the research objectives and to test the hypothesis that mechatronic systems have a steep learning curve.

2.5.1.2 Research Question 3

Question three is mentioned before question two since the second research question was used to primarily meet the second research objective *(issues objective*, *obj. 1*). The third research question sought to determine **whether effort was the biggest contributor to mental workload as measured by the NASA-TLX when installing a mechatronic system** *(issues objective*, *obj 1).* During the NASA-TLX, participant effort is obtained using the question, “How hard did you have to work to accomplish your level of performance?”, and the original publication describes effort as the level of mental and physical activity required to achieve the individual’s level of perceived success with the task [6]. This research question stemmed from analysis of Lau [16], a study in which the NASA-TLX was used to assess the usability of a lower-limb exoskeleton. The analysis performed in Lau [16] is relevant to the experimental study discussed later on in this paper due to similarities with this research project; the purpose of the study was to evaluate novices on their use of a mechatronic device and the protocol used to guide the user through using device. Furthermore, this study and Lau [16] used a tutorial group and non-tutorial group to assess the efficacy of an instructional protocol. Considering the similarities between Lau [16] and the experimental study conducted as part of this research study, it was believed that the NASA-TLX scores in Lau’s [16] tutorial group would be comparable to the scores obtained later during this study. From Lau [16], the NASA-TLX results showed that effort was the highest contributor to mental workload in most steps of the lower-limb usability study, only surpassed by physical demand in some portions of the experiment. When considering the low physical demand of the experimental task proposed later in this study, it was predicted that effort would be the highest contributing NASA-TLX subsection. This question was proposed to statistically identify which aspect of mechatronic design is the most demanding from a user perspective, thus narrowing the focus on where issues may be present.

2.5.2 Second Research Objective

The second research objective was to suggest general design recommendations to intentionally design mechatronics systems that can improve the novice user's experience. This objective was established to determine whether there are ways to reduce the steepness of the mechatronics learning curve.

2.5.2.1 Research Question 2

Question 2 sought to use the NASA-TLX to determine **whether mental workload is reduced by real-time device feedback** *(recommendations objective, obj. 2)*. Modifying feedback was included within the study to create two distinct groups within the experiment with two different approaches to the same mechatronic system. By doing so, it created additional opportunities for participant success and failure, and by extension created additional opportunities for participant reflection. Real-time feedback in teaching (such as by asking a student to repeat or to synthesize knowledge as they go) provides the opportunity for better understanding, but in mechatronic systems, it is unknown what type of feedback (if any) is useful to a novice user. Therefore, the second research question was proposed to determine if there is a difference between two commonly used method of mechatronic communication, delayed device feedback and instantaneous device feedback, in terms of MWL.

2.6 Research Methods

This research sought to answer human factors questions, and an effective way to test human factors issues in man machine interfaces is to evaluate user interaction with the system. Described below are the methods used in this study to explore the research questions and research objectives established previously.

Using a relevant mechatronic system, the study needed to consist of an interactive experimental setup, an engaging scenario, and multiple opportunities for success and failure. The design of the input requirements (information to be given to the participants) and outputs (information taken from participants to answer the research questions) required several variables that could be monitored or changed. The participants needed to interact with the system and follow a laboratory-type manual that would act as an experimental instruction set. The manual was designed to provide users with all the information needed to complete the experiment while still leaving room for error. Since the aim of the study was to elicit user feedback on issues encountered while interacting with mechatronic systems, participants had to encounter situations in which human or system error might negatively contribute to their ability to complete the task.

To inform the NASA-TLX, a HTA was performed to obtain a complete breakdown of the experimental tasks. To understand where desirable and undesirable errors might occur during the experiment, the Systematic Human Error Reduction and Prediction Approach (SHERPA) was applied to the HTA output. The SHERPA method resulted in a breakdown of the task and potential error modes relevant to every task step. The experiment and instruction set were modified to remove unwanted error modes and enable desirable error modes.

The information gathered from the participants included MWL as measured by the raw NASA-TLX (referred to simply as NASA-TLX throughout the rest of this paper), the subsection data of the NASA-TLX, and responses to a list of survey questions. The questions were designed to elicit open ended responses from participants, obtain background information to better inform the NASA-TLX scores, and answer the research questions through post-experiment analysis.

Chapter 3: Hardware Design and Integration

Mechatronic systems come in all shapes, sizes, and levels of difficulty. The first step of the research plan was to identify what kind of mechatronic system would be best used to investigate the research objectives. To enable effective implementation with readily available components, it was decided to use experimental modules designed as part of the research study. These experimental modules were originally created to explore the idea of installing avionics based on orientation and other external variables, such as attached components. Each module consisted of a custom-made printed circuit board (PCB), an inertial measurement unit (IMU), and an Arduino Mega microcontroller. The custom-built PCB was developed to meet researcher defined requirements and specifications relating to the original avionics goal of the experimental modules. The experimental modules were later selectively modified for use in the participant study conducted in this research project.

3.1 Problem Definition

An auxiliary PCB for the Arduino Mega microcontroller platform was designed for use with this project and for later use as a platform for modular avionics and academic design. The form and function of the PCB was based on design criteria identified at the beginning of the project and modified iteratively to meet the needs of the experiment as it evolved. Generalized, the criteria are as follows:

1. The shield must be intuitive to use and novice friendly.
2. All components must be ‘open source’ and reasonably obtainable in terms of availability and cost.
3. Units must be able to transfer data and power to identical units when configured.
4. The units must be compatible with common peripherals used in novice level design (servo motors, sensors, etc.).

‘Intuitive’ and ‘novice friendly’ are qualitative terms used to guide design choices made during the development process. When a system is ‘intuitive’, the user can quickly infer how the system works without having interacted with it before. To make a system intuitive, the designer can draw on patterns, general knowledge, and common sense. A common example of this is the use of the colour red to indicate urgency, danger, or error like with a stop sign. Other ways include limiting the choices the user must make to operate the system, and likewise limiting the ways the user can make a mistake leading to error. ‘Novice friendly’ is like ‘intuitive’ in that by making a system intuitive, you inherently make it novice friendly because you reduce the knowledge required to operate that system. The term differs from ‘intuitive’ when considering that a system can be intuitive to those who have experience in the field associated with that system but can be confusing for those who are novices in that field. Using ‘novice friendly’ as a guiding principle means designing the system in such a way that those who have little to no experience with mechatronics or aviation can successfully set up and use the system the first time they use it.

‘Open source’ in the context of electronics and software indicates that their source material is freely available and can be used, copied, or changed by anyone. Component in this context refers to the building blocks of the electronics and software used to create the overall system. By choosing only open-source hardware (i.e., the pieces used to build the modules) components and software components (i.e., the coding libraries and functions used to build the script) for the design project, the cost and complexity of the final project was reduced. More importantly, it allowed for the design to be easily adapted and modified for future projects.

The last two criteria were hardware constraints that guided the design of the shield to enable the system to be modularized. To be able to communicate and power identical units, the PCB used must be able to electrically accommodate serial communication and direct power transfer. To be compatible with peripherals commonly used in open-source projects, the PCB needed to be able to power and communicate with multiple devices, including other similar units. These peripherals were assumed to be serial devices using serial peripheral interphase (SPI) or inter-integrated circuit (I2C) communication protocols designed for use with Arduino boards or similar microcontrollers, or were assumed to be simple electronic components such as low current servo motors.

3.2 Serial Communications Overview: SPI and I2C

Serial peripheral interphase (SPI) is a serial communication protocol used by many electronic devices for communication and control. SPI uses four separate wires for communication; peripheral in/controller out (PICO), peripheral out/controller in (POCI), serial clock line (SCL), and chip select (CS) [21] (Note: appendices and documentation uses terms MOSI for PICO and MISO for POCI). Communication between devices using SPI requires one device to be the control device and one to be the peripheral, and each peripheral device needs its own CS line.

Inter-Integrated Circuit (SPI) is a separate serial communication protocol that uses only two wires for communication, the serial data (SDA) line and the serial clock (SCL) line. Communication occurs when the control device creates a start condition, followed by an address and a bit of information, and then terminated by a stop condition [21]. Though more complex than SPI considering how information and communication are handled, the protocol is advantageous since it only uses two wires and can allow for multiple control devices in any given configuration. For example, acting as the controller one module can send information to a peripheral module. That second module can then act as the controller and send information back to the original module which would act as a peripheral module. Using I2C, the modularized configuration could consist of numerous devices acting as both controllers and peripherals depending on how the information is being sent.

3.3 Transfer of Information

One of the advantages to modularizing components within an aircraft is the potential for localized data collection and information sharing between components. The system designed within this research uses the I2C serial protocol for communication between units. However, with I2C it can be difficult to quickly share complex information between units since there is a limit to how much data can be transferred in a single packet. This problem was overcome through a software solution that used the following code:

```
//defining of structures used to send imu data as a packet
struct imuData_t {
  float yaw;
  float pitch;
  float roll;
  int address;
};
#define PACKET_SIZE sizeof(imuData_t)
union I2C_Packet_t{
 imuData_t sensor;
 byte I2CPacket[sizeof(imuData_t)];
};
I2C_Packet_t imu_data;
//end definition
```

This code creates a structure under the imuData_t variable containing yaw, pitch, roll, and address data, creates a union between that data and an empty byte of data with length defined by the size of imuData_t, and creates a packet of data under the variable imu_data. The overall significance of this is that all spatial information can be packed within imuData_t to be sent between Arduinos in a single packet of information and be easily translated back into useable information. Communication is split between I2C and SPI to allow for the serial busses to be free when needed. I2C is used for between module communication since it allows for communication back and forth between the experimental modules themselves. SPI is used for module to component communication since it is harder to communicate back and forth using SPI but many components, such as the IMU, are setup for SPI communication.

3.4 Proof of Concept: Modular System Prototype

To ensure that a system modularized using I2C communication was possible, a simple experiment using multiple Arduino microcontrollers was created. Four Arduino boards were used; three Arduino Uno boards used as peripheral devices and one Arduino Mega board used as the control device. Each Arduino board acted as a separate unit within the system, and each Uno board was connected to two LED indicators. Figure 1 represents the communication diagram for this experiment.

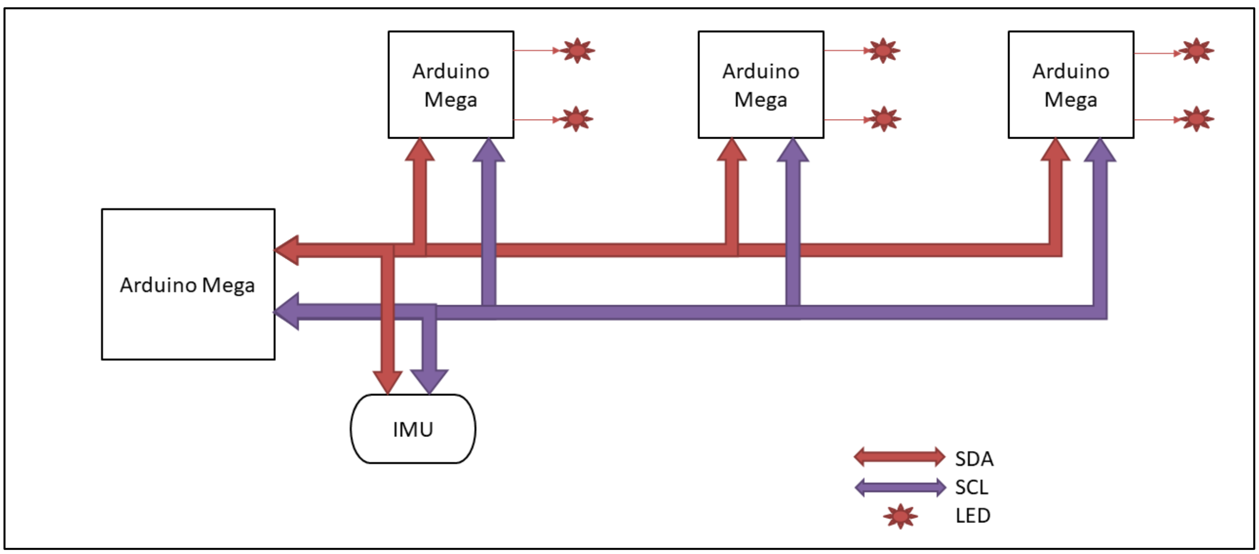


Figure 1: Flow diagram for device communication.

The control unit relayed positional information from an inertial measurement unit (IMU) positioned on the breadboard to the three Arduino Uno units. The three peripheral units were programmed to light one of two LED indicators to identify whether they were pointed left or right of their initial position. There were therefore two states that the LED indicators could be in (left or right), and the information identifying the state of each Arduino Uno was communicated back to the control unit.

This prototype proved that modularization using I2C was possible: the control unit was able to send positional information directly to the three separate Arduino Uno microcomputers using I2C, and information on the state of each Arduino was successfully communicated back to the control unit over the same protocol. There was no significant delay (below 100 milliseconds) between each step in the communication loop, indicating that the system could successfully be used in an environment where reaction time is important.

3.5 PCB Shield Overview

The final experimental module design consisted of a PCB mated with an Arduino Mega microcomputer. The PCB was designed to meet the criteria outlined in 3.1 Problem Definition. A summary of the major elements and their function can be found in Table 2.

Table 2: Breakdown of PCB components and their respective functions.

Component	Function
Adafruit BNO-085 IMU	Combination sensor that provides magnetometer, accelerometer, and gyroscope readings. Performs sensor fusion on the information obtained using on-board sensors and outputs the fused information using I2C or SPI. Which serial protocol is used is determined by the installer and is enabled via a soldered jumper. Default communication protocol used is SPI.
I2C Ports	FourI2C ports are available per board. All ports are positional and have six pins which are connected in parallel. Two pins are used for SDA and SCL, while the remaining two are power and ground. Power can be supplied directly through the I2C ports.
SPI Ports	Two SPI ports are available per board. All ports are positional. Of the six available pins, four are used for SPI communication, one for power, and one for ground. Each port has a unique chip select (CS) line while the other three SPI lines are connected in parallel.
LED Indicators	Four separate LED indicators are available for use per board. Each LED is a different colour (green, yellow, orange, red).
Servo Output	Each board has three screw terminal locations for attaching three wire servo motors. Power terminals for the servo motors are connected directly to the module’s power supply and not to the Arduino power output to provide adequate current to the motors. Each servo connector has a unique digital output for controlling the servo.
Two Pin Power Supply	Locations for two optional two pin connectors are available per board for additional power input or output options.
DCDC Converter	Pinholes are included for a standard three pin DCDC converter on each board to accommodate servo motors that have higher voltage requirements. The input and output pinhole locations are connected via a jumper during board setup if using servo motors that match the power supply used.

LCD Connector	One location for a single LCD screen output is included on each board. Of the eight available pins, four are used for SPI, one for 5-volt power, one for 3.3-volt power, one for ground, and one is left unpopulated. The 3.3-volt power is included for LCD screens with lower power requirements or for screens using 3.3-volt logic.
Adafruit BNO-055 IMU Pinhole Footprint	Included in the design are pinhole locations for a BNO055 IMU. This was included to allow the user to install a different IMU based on availability or preference.

Schematics and drawings of the board can be found in Appendix A: PCB Schematics and Drawings. The board was designed for use outside the scope of this research project and some features mentioned in Table 2 were included to improve functionality in the case it was used for academic design projects. Additionally, any researcher making use of the design can populate what components they need without compromising its functionality.

3.6 Future Design Work for the Mechatronic System

While the device developed was used to conduct the usability testing that provided answers to the research questions within the rest of this thesis, the researcher's original goal was to design the mechatronic test system to be used as an avionics system applicable to academic research. During the design process, certain obstacles that would prevent the current iteration from being useful in aviation and academia areas were identified but unaddressed because they did not interfere with its usability in this research project. Similarly, there were features identified but left unpursued or unimplemented for the same reason. Reflecting on this, the module system developed in this research would be applicable to academia and modular aerospace with further iteration that addressed the issues and features that were passed over during the design process.

3.6.1 Future Design Work: Serial Addressing

The issue that would most impact the experimental module's use in complex projects would be the I2C serial addressing. In this study, only three potential addresses were available in the software, and the potential error mode of conflicting addresses was tackled using software unique to the module with address 1. Though this solved the problem for the participant study, it didn't solve the problem of conflicting addresses. There are a finite number of I2C addresses available, and if the modules were manufactured to be complete from the supplier, there would be no guarantee that devices would have different addresses. This issue could be solved either through software or hardware, with both methods having advantages and disadvantages. Regarding software, the device could require that users connect to a computer and set their address, however, this wouldn't be feasible if the system is meant to work out of the box. In the case that the system is intended to work right out of the box, software could be used that addresses the devices based on environmental factors or peripherals unique to each module, such as was done in this study. Regarding hardware solutions, the simplest solution would be the inclusion of an encoder component that users could physically set upon installation. Unfortunately, the issue would remain that the system would need conscious intervention and would not be useable right out of the box. Which solution would be appropriate would depend on the situation, since users in the academic field would prefer the device to be modifiable, whereas users trying to use the system in modular avionics would likely prefer the device to be useable without any low-level modifications.

3.6.2 Future Design Work: EEPROM Implementation

Another issue arises with the system when considering that mechatronic devices are designed to be powered off regularly, whether intentionally (like when not in use) or unintentionally (temporary power failure), without issue. In the current iteration, the entire system resets when powered off. In the case of this research study, this was desirable since it was being used with multiple participants who each needed to set up the system. This system reset would become extremely undesirable in an aircraft; if orientation was used to address each module, then there would be no way to reliably address the devices after temporary power failure midair. This issue could be solved by saving the address determined during installation onto onboard Electrically Erasable Programmable Read-only Memory

(EEPROM), an electronic component that Arduino microcontrollers use as a tiny version of a hard drive. Though this feature would be necessary with any future iterations of the modules, there would be limitations associated with its inclusion. Firstly, there would need to be a way for the user to reliably reset the device's address, since after the installation process it would have a persistent address saved to the EEPROM. This would likely be solved using a button, or equivalent intervention method, that could be pressed during startup to override the saved address, allowing the device to readdress if needed. Secondly, there would be a limit to the number of times each address would be addressed since the EEPROM is physically limited by how many times its memory can be changed. Though the limit of the EEPROM would likely never be reached in normal operation, since it has a specified lifespan of 100 000 write/erase cycles [22], this limitation would still need to be considered.

3.6.3 Future Design Work: IMU Redundancy

Another feature that was identified but left undeveloped was IMU redundancy. In the iteration of the system used for the study, each IMU was used to address the system through orientation, but after setup only one IMU was used for yaw, pitch, and roll data. The decision to use only one IMU to represent the position of the model was made to reduce potential for error. When inspecting data from multiple IMUs, it quickly becomes apparent that no two devices are identical or perfect; all devices have different margins of error and noise that do not necessarily match. It was simpler and more accurate in the participant study to only use one IMU, since all positional information was calculated using a single origin quaternion. In future iterations, it would be useful to add software that would allow a second IMU to be used for positional information in the case that communication with the head device was lost or if the original IMU became unusable. This would be especially useful in aerospace applications where positional measurements would be more critical.

While these additional modifications to the hardware would be essential for future engagement with students, the designed mechatronic system allowed for testing of the proposed research questions in identifying usability. The experimental design process incorporating the designed software/hardware system is discussed in the next chapter.

Chapter 4: Study Design

Through personal experience, the author of this study found during their mechanical engineering education that mechatronics was a difficult to learn topic made even more so due to the way it was taught. Years later as a teaching assistant, the author found that students entering the mechanical engineering programming were still frustrated with the mechatronics courses and were having difficulty learning the subject. The original aim of this research project was to create the modular system described in the last chapter in such a way that it could be given to someone with little knowledge of avionics and allow them to install the device using only orientation and environmental factors surrounding the modular system. However, the focus of the project changed to that of identifying issues novices have with mechatronic devices and improving mechatronic device design as the project iterated. The author found that while designing the modular mechatronic system, they were encountering ways to improve the teaching of mechatronics. Therefore, the focus of the project shifted to improving usability for novices using the developing system.

The driving concept behind the participant study was to have participants identified as novices set up and use a mechatronic system in a simulated application and use responses and MWL scores to meet the research objectives. The role of the mechatronic system would be fulfilled by the experimental module described in the previous chapter, while the simulated application was chosen to be the installation of an 'autopilot' inside of an aircraft. The 'autopilot' situation was chosen since it would be easy to implement, it provided users with an adequate experimental challenge, and provided a compelling scenario to help motivate participants.

4.1 Physical Setup

The hardware setup consisted of three separate experimental modules, each consisting of an Arduino Mega microcontroller and a custom-made PCB mounted onto the Arduino. All three experimental modules were identical in terms of software and hardware. The PCBs were only populated by the components deemed necessary for the study. The following list summarizes the components populated and omitted:

- An Adafruit BNO-085 IMU was used and the location for the BNO-055 was left unpopulated.
- Only two I2C port locations were populated per experimental module, one for upstream communication and one for downstream communication.
- No SPI connection locations were populated since no additional peripheral devices were needed.
- The LCD connection location was left unpopulated since no LCD screen was needed.
- Only one servo output connection location was populated since only one servo was needed for each module.
- Only one two-pin power connection location was populated for testing purposes since power was being supplied through the I2C ports.

Each experimental module had one servo motor attached to the included screw connector location prior to the experiment. The outer casing for each module was constructed from plastic toy bricks to make it easier to repeatedly affix the modules to the experimental model. Three cables were provided to the participants; two I2C male-to-male cables for communication between modules, and one male to power cable for connecting the variable power supply provided to one of the experimental modules. A pictorial part list is included in Appendix C: Part List.

4.2 Software for Experimental Testing

The software uploaded to each of the three experimental modules used during the experiment was identical, and each device ran a unique section of the script based on its address. The software was designed to use three modules in three different configurations, and each configuration was determined using an address set during the initial portion of the script based on the orientation of the experimental modules. Address 1, Address 2, and Address 3 corresponded with roll/control module, pitch module, and yaw module respectively. The 'autopilot' portion of the script used the three modules to control the model airplane's flight surfaces to counteract movement of the model airplane. The script in its entirety can be found in Appendix D: Software Transcript, and the function of each section is summarized in Table 3.

Table 3: Software breakdown.

Section	Line # Start	Used by Address	Function
Startup	1	All	• Libraries are initialized. • Structures, global variables, and pinouts are defined. • Desired IMU outputs are defined.
Function: setReports	67	All	• Function for declaring the desired outputs from the IMU.
Setup	74	All	• LED pins, Serial output, and IMU are initialized. • Module orientation is determined. If in one of the correct orientations, the module is given an I2C address between one and three. The address will not be set if in an incorrect position. • Address 1 is the head module and roll control, Address 2 is pitch control, and Address 3 is yaw control.
Setup: Duplicate device check	194	1	• Prevents the control module from moving on until both Address 2 and Address 3 have communicated with Address 1, effectively preventing the system from running if two devices have the same address.
Loop: head module/aileron device	222	1	• Calculates yaw, pitch, and roll and sends that information every loop to the other two modules. • Changes the angle of the attached servo based on roll angle.
Loop: pitch control	269	2	• Changes the angle of the attached servo based on pitch angle.
Loop: yaw control	286	3	• Changes the angle of the attached servo based on yaw angle.
Function: receiveEvent	305	All	• Interrupt function that receives packet of information over I2C using the wire library.
Function: i2csend	316	All	• Function for sending packet of information over I2C using the wire library.

Data from the IMU is received in the form of a single rotation quaternion that is a result of sensor fusion between the accelerometer, magnetometer, and gyroscope. The direction of gravitational force is extracted from the quaternion and used in setup to determine what the device's address should be. For example, if an experimental device is oriented face up and the gravitational force vector is greater than a given threshold in the positive z-direction (using a standard x-y-z axes), the module will be given Address 1. Address 2 and Address 3 are similarly assigned using gravitational force in the x and y directions.

As mentioned earlier in Table 2, each module has four LED indicators used to communicate to the user. The meaning of each LED code is elaborated upon in Table 4.

Table 4: LED indicator communication and corresponding meaning.

One blinking LED	Device is in one of three correct orientations: • Green LED blinking corresponds to roll control (I2C address one) • Yellow LED blinking corresponds to pitch control (I2C address two) • Orange LED blinking corresponds to yaw control (I2C address three) Device must be stable for 10 seconds (blinked 20 times) to permanently set address.
The Red LED and one other LED is on	The device is oriented in one of the three correct planes but is upside down.
Only the Red LED is on	The device is oriented incorrectly and is not in one of the three correct planes. Address will be reset to zero.
All four LEDS blink once	The device's address has been set (one, two, or three).

Once the I2C address is set and the bus starts, the device begins running through the looping portion of the software. In this section of the script, which portion runs is based on the previously established address which can no longer be changed without a reset. Each module takes information extracted from the rotation quaternion and rotates their respective servo motors in a way that 'counteracts' movement of the module. Since the model is a representation of a simple aircraft, this translates to the modules moving the flight surfaces in directions which would return the airplane to an upright position parallel to the ground, heading straight north. This way the participants can determine whether the 'autopilot' they have set up is working properly. Figure 2 is a schematic of the system as it was used in the experiment, illustrating how the modules communicate with other modules using I2C and with their respective IMUs over SPI.

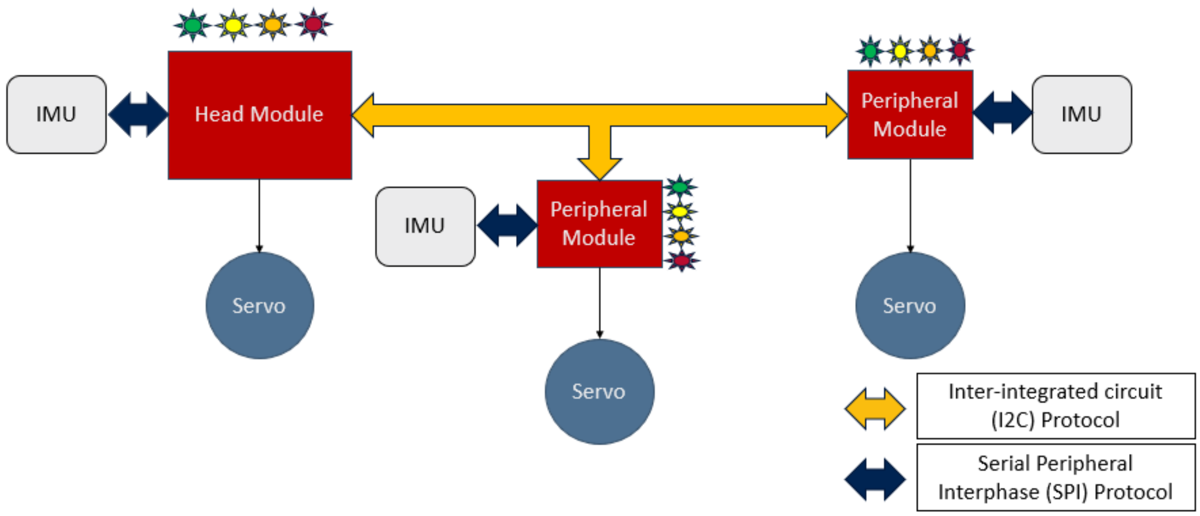


Figure 2: Schematic of the mechatronic system as it was used in the experiment.

The servo motors and LED indicators shown in Figure 2 are not controlled using serial communication and are instead control using simple digital outputs.

The pitch and yaw control code blocks differ only in what LED is activated to indicate its role, and by the formula used to translate pitch/yaw information into servo movement. The roll/control module has additional code to allow it to properly execute its role. This control device first checks whether all three devices are present by checking whether it has received information from device two and device three. If there are duplicates or if either device is not present, the red indicator light will be on along with the green indicator. Though this error condition is also present in the startup routine, the startup routine is ended once all lights blink once. Therefore, participants will be able to differentiate among the error conditions by watching whether the device exited its startup routine. The head module is also responsible for translating the quaternion into yaw, pitch, and roll information and then distributing that information to the other two modules. This information could be obtained from the IMUs of the separate devices, but only data from one IMU is used to limit the complexity of the software and potential for error (refer to 3.6.3 Future Design Work: IMU Redundancy).

4.3 Hierarchical Task Analysis

Before performing a MWL assessment like the NASA-TLX, Stanton [5] recommends that a task analysis method be used to fully understand the task being assessed. The method chosen to be used was the hierarchal task analysis (HTA) method. The HTA method is a powerful tool because it is easy to learn and apply effectively to almost any situation. This method is also used as a basis for a number of other task analysis methods [5], such as the SHERPA method applied later, and is a good starting point for any human factors assessments.

The steps used to apply HTA were taken from Stanton [5], and were used to analyze the initial experimental methods.

Step 1: Define Task under Analysis

The first step was to define the task, as well as the purpose behind the analysis. In this study, the participants were being asked to assemble and test a mechatronic system, therefore, the task analyzed was the installation of the experimental system. The purpose behind analyzing the task was to inform the data collection methods (NASA-TLX) and analysis methods used during and after the participant study.

Step 2: Data Collection Process

During this stage of the HTA, all data deemed relevant to developing the HTA was collected. Since the task had not previously been undertaken by participants, this part of the process was undertaken by the study's author. This process involved the author running through the proposed task as if they were being told to complete it.

Step 3: Determine the Overall Goal of the Task

The third step of the HTA was to define the overall goal of the task under analysis. The overall goal of the task was to correctly assemble and test the mechatronic system using the provided components and instructions.

Step 4: Determine Subgoals

The next step of the HTA was to take the overall goal defined in the third step and break it down into sub-goals which adequately describe the whole task. The overall goal of the task was to correctly assemble the mechatronic system using the provided resources, and this was further broken down into logical sub-goals. These sub goals are steps A1 to A5 in Figure 3, and steps B1 to B4 in Figure 4.

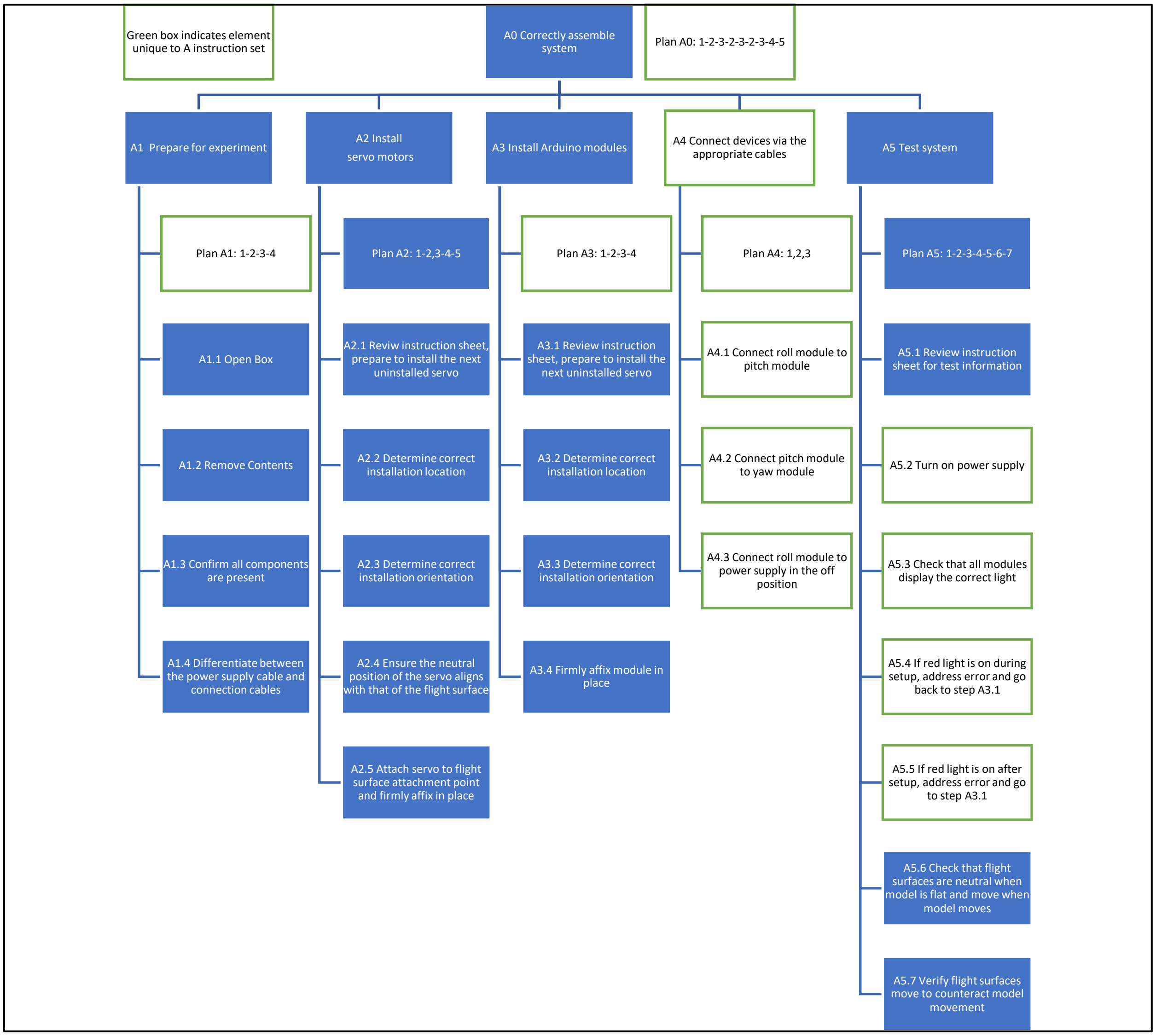


Figure 3: HTA for the first approach where system is powered after installation.

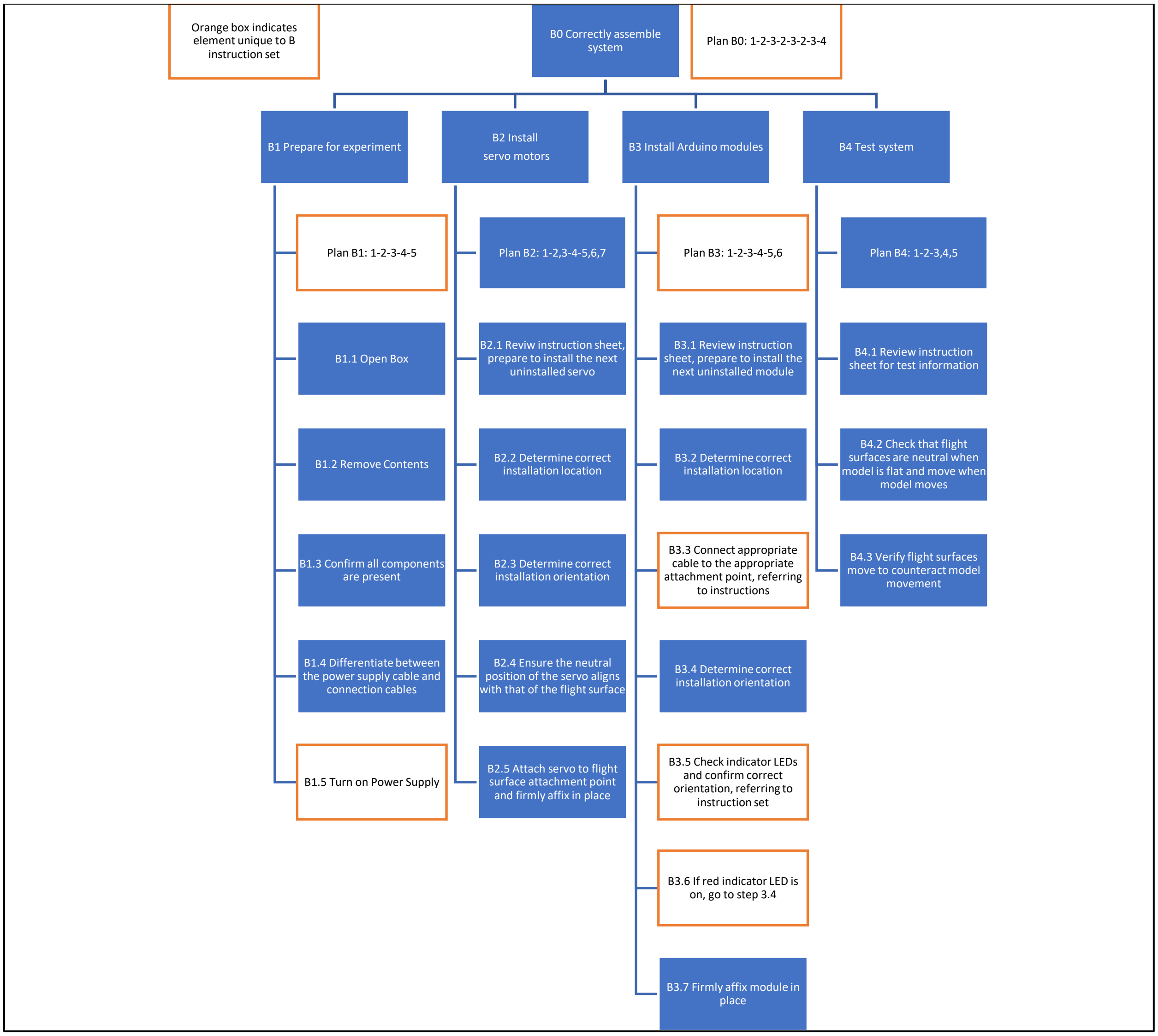


Figure 4: HTA for the second approach where system is powered before complete installation.

Step 5: Sub-goal Decomposition

Step five of the HTA involved breaking down the subgoals outlined in step 4 into their own subgoals and tasks. These decompositions are shown below steps A1 to A5 in Figure 3 for the first instruction, and below steps B1 to B4 in Figure 4 for the second instruction set.

Step 6: Analysis of Plans

Once every goal and subgoal had been identified, the plans for how the tasks would logically be achieved were determined. In the graphical HTA diagrams, a comma (“,”) denotes that the tasks separated can be done in any order, while a dash (“- “) denotes that the tasks must be done sequentially. For example, “1-2,3-4-5,6,7” reads “do 1, then 2 and 3 in any order, then do 4, then do 5, 6 and 7 in any order.”

HTA analysis was performed on the two approaches being used in the participant study. Figure 3 represents the first approach where the power supply remains off until all the components have been installed (control method), effectively providing feedback from the device after setup. Figure 4 represents the second approach where the power supply is on while the user is installing the components (experimental method), effectively providing feedback from the device as the system is being installed.

4.4 SHERPA Method Error Identification

In this experiment, one of the objectives *(obj. 1)* was to provide recommendations based on the experiences of novice participants interacting with the provided system, therefore, potential for participant error was not necessarily unwanted. The purpose of an error identification analysis in this experiment was to identify unwanted error modes that would negatively impact the experiment and provide participants with tools to address desirable error modes that could positively impact the experiment. The chosen method of analysis was the Systematic Human Error Reduction and Prediction Approach (SHERPA) as described by [5], which provides eight steps for applying the SHERPA method.

1. Conduct a human task analysis.
2. Classify each step in the HTA under one of the SHERPA behaviors taxonomies as shown in Table 5, subjectively determine what error modes are credible, and classify them using SHERPA external error mode taxonomy.

Table 5: SHERPA external error mode taxonomy [5].

Behavior Taxonomy	External Error Mode Taxonomy
Action Errors	A1. Operation too long/short A2. Operation mistimed A3. Operation in wrong direction A4. Operation too little/much A5. Misalign A6. Right operation on wrong object A7. Wrong operation on right object A8. Operation omitted A9. Operation incomplete A10. Wrong operation on wrong object
Checking Errors	C1. Check omitted C2. Check incomplete C3. Right check on wrong object C4. Wrong check on right object C5. Check mistimed C6. Wrong check on wrong object
Retrieval Errors	R1. Information not obtained R2. Wrong information obtained R3. Information retrieval incomplete
Communication Errors	I1. Information not communicated I2. Wrong information communicated I3. Information communication
Selection Errors	S1. Selection omitted S2. Wrong selection made

3. Using error taxonomy, identify and classify all credible human errors.
4. Analyze and adequately describe the consequences of each credible error.
5. Determine the potential for recovering from the error mode and identify if there is a later HTA step at which the error could be recovered.
6. Perform an ordinal probability analysis and rate the probability of each error mode occurring as low, medium, or high. In a normal analysis, low corresponds to errors that have never occurred before, medium corresponds to errors that have occurred before, and high corresponds to errors that occur frequently.
7. Perform a criticality analysis and rate the error modes based on the severity of their consequences.

8. Provide remedies for the error modes. Normally these would fall under the categories of equipment, training, procedures, or organizational. In the case of this study, the remedies are not categorized.

The complete SHERPA method output can be found in Table 12 of Appendix B: Combined SHERPA Output.

4.5 Instruction Set Design

Two user instruction sets were created using the HTA and SHERPA analyses in addition to the following design choices:

- Instruction set elements were left purposefully vague or open to interpretation to create opportunities for mistakes that participants could address and later discuss (e.g., some terms such as servo attachment point were left undefined).
- The diagrams included in the supplementary page were intentionally separated from the instruction set pages to make it more difficult for participants to draw connections between the instruction steps and their corresponding diagrams. The assumption was made that some participants may rely solely on the instruction set or the diagrams when they should rely on both.
- During the pilot study (discussed later in section 5.3 Pilot Study), it was found that overly long or complicated instruction steps (greater than two lines or including more than one required action) created unwanted confusion and a sense of being overwhelmed. This was noted under recommendations but addressed in the final iteration by breaking every instruction step into a series of one to two sentence long instruction steps.

4.5.1 Control Group Instruction Set versus Experimental Group Instruction Set

Two instruction sets were created to address the second research question investigating the effect of real-time device feedback on MWL. **By providing delayed feedback, the first instruction set acted as the control set. By providing real-time feedback, the second instruction set acted as the experimental set.** The primary difference between the two instruction sets was the order in which the devices were setup and powered. The control set instructed participants to set up each module before turning the power supply on. By powering the system after setup, the LED feedback would be delayed and occur simultaneously across all three experimental modules. The experimental set instructed participants to turn the power supply on before setting up any modules. By powering the system before setup, the LED feedback would be provided to the user as they setup the system and would allow them to use the feedback to help install each device. These instruction sets are found in Appendix E: Participant Instruction Sets. A summary of the differences between the instruction sets is listed below.

- Step 2:
  - Control Set: “Check the power supply and confirm that it is off. Power during setup will damage components.”
  - Experimental Set: “Turn on the power supply.”
- Step 6:
  - Control Set: “Connect the first module, the roll module (Figure 3), to the aircraft power supply using the power supply cable and install in orientation A (Figure 4) at the front of the aircraft (Figure 1).”
  - Experimental Set: “Connect the first module, the roll module (Figure 3), to the aircraft power supply using the power supply cable and install in orientation A (Figure 4) at the front of the aircraft (Figure 1). If in the correct position, green LED will blink (Figure 5). After ten seconds, all four LEDs will flash. The module is now set as the roll module.”

- Step 7:
  - Control Set: "Firmly affix module to aircraft."
  - Experimental Set: "Firmly affix module to aircraft. Afterwards, the red LED will remain on until all other devices are connected correctly."
- Step 11:
  - Control Set: "Connect the second module, the pitch module (Figure 3), to the roll module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1)."
  - Experimental Set: "Connect the second module, the pitch module (Figure 3), to the roll module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1). If in the correct position, the yellow LED will blink (Figure 5). After ten seconds, all four LEDs will flash. The module is now set as the pitch module."
- Step 16:
  - Control Set: "Connect the third module, the yaw module (Figure 3), to the pitch module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1)."
  - Experimental Set: "Connect the third module, the yaw module (Figure 3), to the pitch module and install in orientation C (Figure 4) near the center of the aircraft (Figure 1). If in the correct position, the orange LED will blink (Figure 5). After ten seconds, all four LEDs will flash. This module is now set as the yaw module."
- Additional Step within Control Set (Step 18):
  - "Turn power supply on and determine if each module completes setup correctly. If setup correctly, each module will blink an LED corresponding with its role for ten seconds before blinking all LEDs once, indicating its role has been set. Green identifies the roll module, yellow identifies the pitch module, and orange identifies the yaw module."

In addition, a supplementary diagram page was created to address error modes encountered during the SHERPA analysis and the pilot study. These supplementary figures are shown in Figure 5.

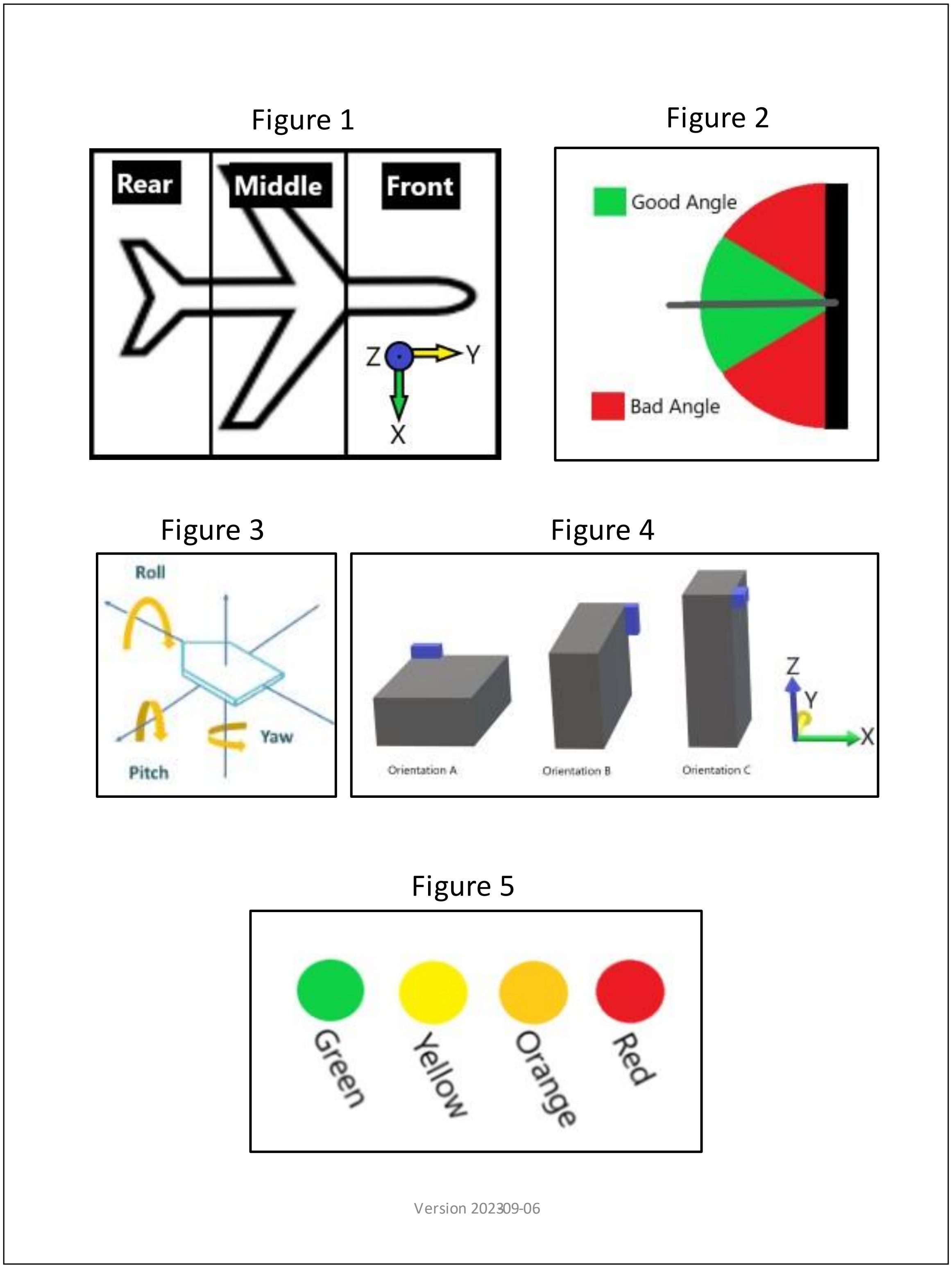


Figure 5: Supplementary diagrams presented to participants during the experiment.

4.6 Application of NASA-TLX

To evaluate the participants' interactions with the system, the NASA Task Load Index was chosen as a method of measuring the overall MWL of the task. NASA-TLX scores were used to gauge the perceived difficulty of the task and to help determine whether there were any consistent issues identified by participants (*obj. 1*). The version of the NASA-TLX used was the raw NASA-TLX (referred to throughout this text simply as NASA-TLX), meaning that the weighting factors used in the original NASA-TLX were omitted. To generate recommendations on how to better design for novices *(obj. 2)* and to help inform the NASA-TLX data, a separate survey was conducted that asked participants questions related to the experiment. A significant number of the research articles reviewed as part of this study that used the NASA-TLX also used separate questionnaires to better inform the scores obtained (refer to 2.4 Data Collection and Analysis and Table 1). Since the NASA-TLX scores only measure MWL, to answer the research questions outlined in this study an additional questionnaire was deemed necessary. The questionnaire consisted of two sections, a demographic questionnaire administered before the experiment and an interview administered after the experiment. The study and its data collection methods received ethics approval from General Research Ethics Board (GREB) Protocol #6039557.

4.6.1 Demographic Questionnaire

The pre-experiment questionnaire was developed to obtain participant background information that could better inform the results of the NASA-TLX and consisted of the following questions:

Q1. Age and Gender.

Q2. Educational background (level and faculty if applicable).

Q3. Rate your experience with microcomputers and microcontrollers.

Q4. Briefly describe your programming experience.

The experiment was open to any students from Queen's University, Kingston in undergraduate or graduate studies, and these questions were developed to investigate whether NASA-TLX scores were correlated with experience or a specific demographic. Specifically, the questions were meant to record that participants were in fact novices to mechatronics, and were meant to help investigate whether students from an engineering background had lower MWL scores than those not from an engineering background.

4.6.2 Post-Experiment Interview Questions

The second part of the questionnaire was developed to elicit feedback from the participants that could be used to answer the study's research questions. The decision was made to deliver the questionnaire orally and recorded electronically because it was theorized that participants would provide more open-ended answers to the questions if unconstrained by pen and paper. The second part of the questionnaire consisted of the following:

Q1. Describe the effectiveness of the instructions.

Q2. What aspects of the experiment aided or hindered your understanding of the instructions?

Q3. How would you describe the overall difficulty of the task presented to you?

Q4. Do you have any suggestions on improving the overall procedure?

Q5. Do you have any suggestions for improving the instructions?

Q6. In what ways do the illustrations and text complement each other?

Q7. How well do you believe you executed the instructions?

These seven questions were used to cover the most significant usability aspects controlled by the designer. While developing questions such as these, it is important to consider that the process of doing so is iterative when undertaking such experiments. In fact, the last question was added after the pilot study (refer to 5.3 Pilot Study) was administered, as it was deemed important to investigate how participant confidence correlated with NASA-TLX scores. It should also be noted for clarity that the procedure mentioned in Q4 refers to the entire installation process including the instructions and supplementary documentation.

Chapter 5: Experimental Protocol and Methods

The conceptual design for the experiment consisted of having participants setting up the provided modular system designed in this study in an 'autopilot' configuration on a model aircraft. Each participant would be given all components in a box along with either the control or experimental instruction set. By following either instruction set, the participant would be able to install each module and component onto the aircraft model and be able to test the system themselves after assembly. The system would be designed to react to the movement of the model once properly installed, providing participants with another form of feedback. Following the test protocol, an interview would occur that included additional questions described previously. The interview would be audio recorded to ensure that the researcher could revisit the answers to the questions and analyze the responses. Following the recorded interview, participants would be debriefed on the nature and purpose of the experiment.

5.1 Participants

Prior to beginning the experimental, two pilot participants completed the study. These individuals were two students who had completed their second year of mechanical engineering. The results from the pilot informed changes to the study, as listed in section 5.3 Pilot Study.

Using the updated experimental protocol, the study was conducted over a two-day period during which sixteen students from Queen's University's Kingston campus agreed to participate. As noted earlier, the study received ethics approval from GREB Protocol #6039557. Seven participants were male and nine were female with an average age of 23 +/- 5 (SD) years. Eleven of the sixteen students came from an engineering background, and of those eleven students, four had completed an undergraduate degree in engineering and were pursuing a graduate degree at the university. The remaining five students were from a non-engineering related background, and of those five students four were pursuing non-engineering related graduate degrees at the university. The students pursuing graduate degrees were split evenly across both groups, with four in the control group and four in the experimental group. Non-engineering participants also were split between the two groups, with three in the control group and two in the experimental group. All undergraduate students were in their second or third year of study. Graduate students did not all state their year of study within their graduate degree, and comparisons were difficult due to the lengths of different programs of study (i.e., students in medicine have more years of study than students in engineering). Most students reported little to no experience with microcontrollers, and those that did report having experience were not above the level of a novice user. No information was collected regarding extracurricular activities or experience before university. Future research would benefit from the inclusion of this information, such as high school shop experience or childhood environment (i.e., on a farm or in the city). No incentives were provided to the volunteers for participation.

5.2 Participant Testing Methods

The experimental setup consisted of a mounted representation of an aircraft constructed out of plastic toy bricks. The model had moving flight surfaces that represented each of the axes of rotation (i.e., yaw, pitch, and roll). The model had attachment points for each of the module's connected servos, along with a model 'cockpit' complete with wired flight surface controls that participants could interact with. The complete model is shown in Figure 6.

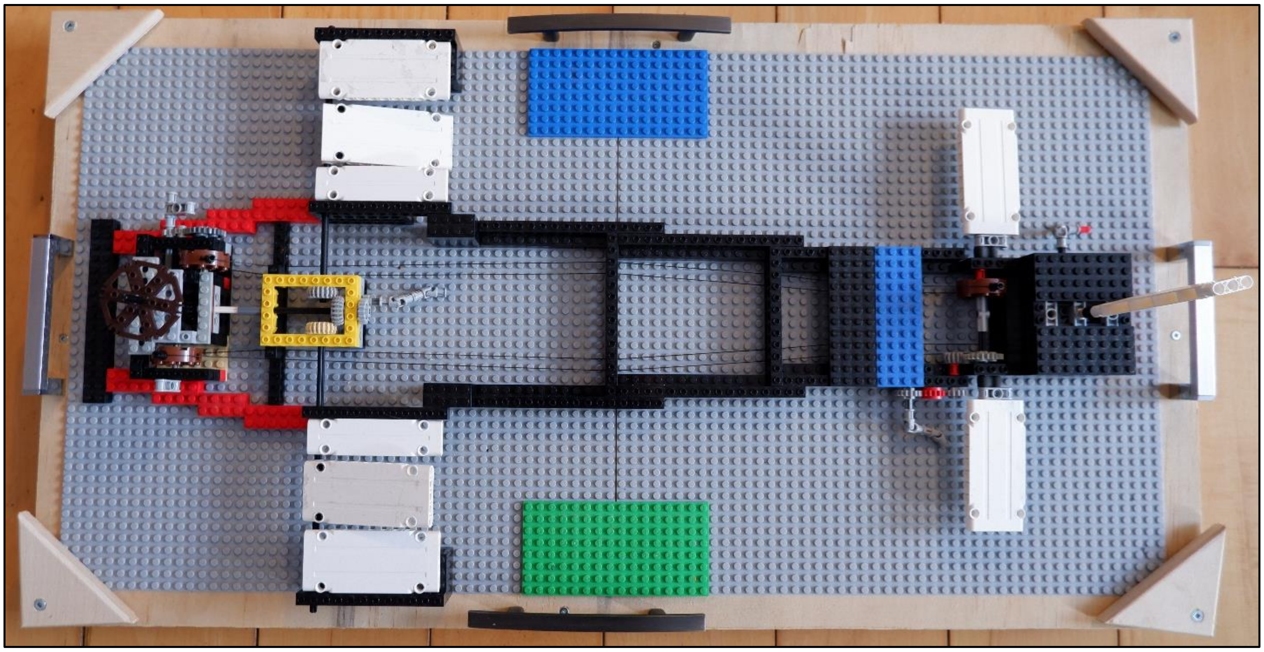


Figure 6: Experimental model used in the participant study.

Before the experiment, the study administrator ensured that all units had the correct software, and that each unit was operating properly. Before and between participants, the administrator ensured that units undamaged, and that all pieces were present and in working order.

During the experiment, each participant was provided with a box containing all required modules, cables, and accessories. Along with the box, they were provided with a model onto which they were to attach the modules and accessories, a paper copy of the NASA-TLX, a copy of the pre-experiment questionnaire, and a paper version of one of the instructions sets. Participants were not informed that there were two instruction sets. The NASA-TLX and survey sheet can be found in Appendix F: Data Collection Methods, and transcripts of the instructions are found in Appendix E: Participant Instruction Sets.

5.3 Pilot Study

A pilot study was conducted using prototype versions of the two instruction sets and a prototype experimental setup with two participants. Before the experiment, both participants were asked to fill out the first section of the questionnaire (the demographic section). The participants were then given one of the two instruction sets and were not informed of the differences in the instructions given until after the experiment. Immediately after each participant completed the experiment, the researcher interviewed the participants using question one to six from 4.6.2 Post-Experiment Interview Questions and recorded their responses. After both participants were interviewed and provided information regarding the nature of the two instruction sets, the researcher was able to discuss the participants' experiences in further detail during an unguided interview.

Based on observations made during the pilot study, the participants' questionnaire answers, and the post experiment discussion, issues with the initial experiment were identified and addressed:

- **Instruction sets:** The participants found some task steps were too lengthy and confusing (i.e., steps with more than one required action or steps covering multiple lines on the page making it difficult to read). This was addressed in the iteration used in this study by breaking up lengthy steps into smaller sub-steps. Furthermore, text identified as overly confusing by the participants was reworded. Though this feedback was useful in answering the research questions, this issue was addressed in the final experiment. Without change it was believed that issues with task step length, which was already identified as an issue in background research by Denny et al. [4]and in the pilot study, could inhibit new responses more useful for answering the research objectives in the final study by inhibiting the clarity of the instruction sets.
- **Modules:** It was found that the module casings were flimsy and needed significant change to be soundly attachable to the experimental setup. The module's second iteration, though still made of toy bricks, was made sturdier and more easily attachable to the experimental setup.
- **Servo Motors:** The initial casing for the servo motors allowed them to be attached using almost any face of the casing, however, participants found that this caused the servo motors to be

difficult to attach to the experimental setup, and when attached were flimsy. The casing was redesigned to be sturdier and to allow the motor to be installed only on a single face.

- **Experimental Setup:** In the initial design of the experimental setup, yellow squares were used to indicate where the servo motor attachment points were located (refer to Figure 7). The yellow squares were found misleading to the participants, since they believed they needed to attach the servos directly to the squares. Therefore, the yellow squares were removed. Additionally, participants made numerous mistakes in installing the servo motors onto the experimental setup and often installed them in ways that prevented the attachment axles from turning. This was addressed by the inclusion of a diagram in the supplementary diagram page that illustrated the angular limits of the attachment axle (supplementary Figure 2, Figure 5).

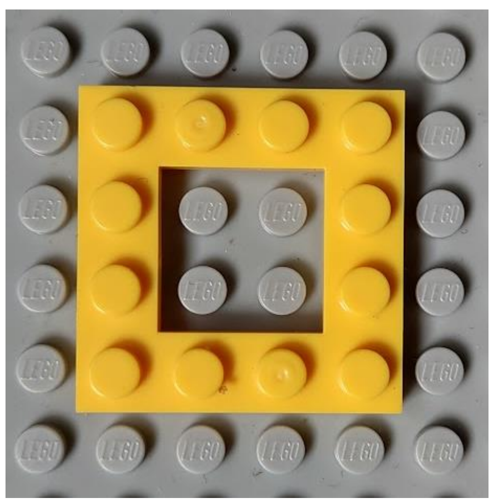


Figure 7: Yellow square removed from final iteration of the experimental model.

The pilot participants' responses to the NASA-TLX were inspected and found to equate to medium to high mental workload. Since the issues were addressed after the pilot participant experiment, the MWL scores of the final experiment were predicted to be lower.

5.4 Experimental Procedure

Once the information from the pilot participants was integrated into the instruction sets, the experimental protocol was conducted with each of the sixteen participants. Each participant was allotted a minimum of thirty minutes of time but were allowed to continue past the time if needed. The participants were divided randomly into two groups based on their participant number; odd numbered participants (1, 3, 5…) were given the control instruction set, and even numbered participants (2, 4, 6…) were given the experimental instruction set (refer to section 4.5.1 Control Group Instruction Set versus Experimental Group Instruction Set). The experiment concluded when the participant indicated that they had finished the instruction set or when the participant indicated that they could get no further with the instructions. Upon conclusion of the experiment, participants were asked to rate their MWL using the NASA-TLX. Afterwards, they were asked a series of questions (refer to 4.6.2 Post-Experiment Interview Questions) and debriefed on the nature of the experiment. The NASA-TLX MWL scores and the participant's interview responses analyzed after the experiment, and the results were used to investigate the research objectives.

Chapter 6: Experimental Results and Analysis

6.1 Mental Workload Scores

After the experiment, participants were asked to complete the NASA-TLX survey. The subsection and overall MWL scores were used in answering the research questions and by extension the research objectives. More specifically, subsection and overall MWL scores were used to investigate where usability issues might have existed with the task being measured and help guide recommendations on how to improve the usability of similar tasks. Results between the two groups with respect to the NASA-TLX subsections and overall scores are shown in Figure 8.

Figure 8: Comparison of the NASA-TLX subsection data and overall MWL scores of the control and experimental groups. Subsection and MWL scores are unitless.

Additionally, in Figure 9 the NASA-TLX subsection means for participants of engineering background and non-engineering background were compared.

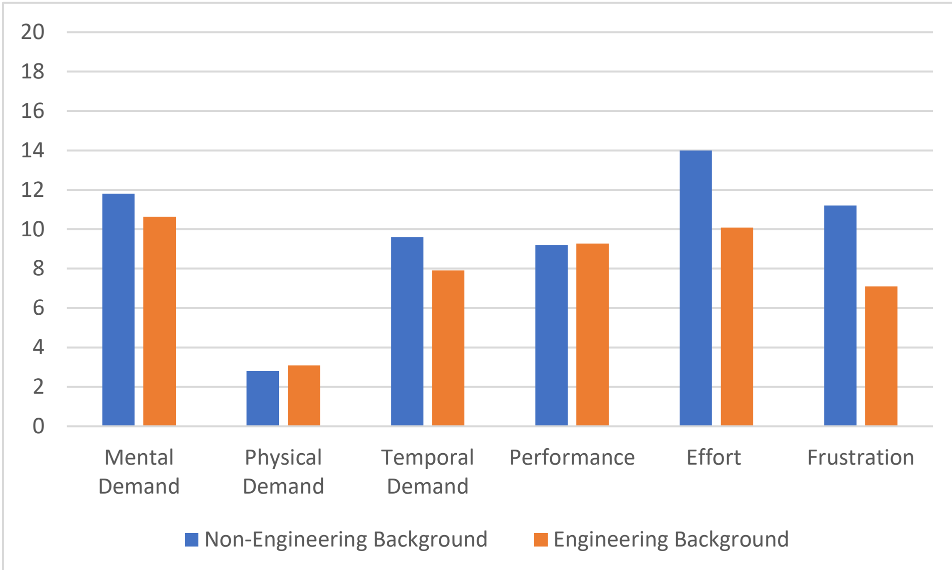


Figure 9: Comparison of NASA-TLX subsection data between participants of non-engineering background and engineering background.

Qualitatively, it can be observed that while the two demographics had similar scores in most subsections, the mean subsection scores for engineering background participants were generally lower. Statistical analysis looking at the difference between scores from the non-engineering and engineering participants was excluded due to the difference in sample sizes and the low overall population size.

6.1.1 Statistical Analysis of NASA-TLX Scores

To determine whether there was a statistical difference in the MWL between the control group versus the experimental group, single factor analysis of variance (ANOVA) was used on each individual subsection of the NASA-TLX as well as on the overall MWL score. T-tests were also considered an appropriate way to test the means and were conducted to confirm the results of the ANOVAs. Using a 95% confidence threshold for both tests, it was considered unlikely to achieve significance due to the small population size of sixteen but was used regardless due to the qualitative differences between the control and experimental groups in Figure 8. P-values obtained from the t-test were identical to the

ANOVA results and therefore were not included in Table 6. Table 6 summarizes variance sources and the p-values that resulted from each of the ANOVAs.

Table 6: ANOVA results (5% significance level) with mean and variance values shown.

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration	NASA-TLX Scores
ANOVA P-Value	0.1068	0.1502	0.3492	1.000	0.1359	0.4805	0.1878
Control Group Mean	12.8	3.6	9.4	9.3	12.9	9.5	57.4
Experimental Group Mean	9.3	2.4	7.5	9.3	9.8	7.3	45.4
Control Group Variance	15.92	4.267	11.98	31.36	18.98	45.14	338.4
Experimental group Variance	17.07	1.125	18.00	35.64	12.21	31.92	262.6
Between Group Variance	49.00	6.250	14.06	0.000	39.06	20.25	276.0
Within Group Variance	231.0	37.75	209.9	469.0	218.4	539.5	4206

Though there appeared to visually be a difference between the two groups based on the graphical representation (Figure 6), there was statistically no difference between the means of any subsections or in the overall MWL score at a significance threshold of 0.05. Given the small population size, this was not an unexpected outcome using a statistical threshold of 0.05, however, the results still are not significant using a higher statistical threshold of 0.01. Figure 10 illustrates visually how far above the significance threshold the p-values were.

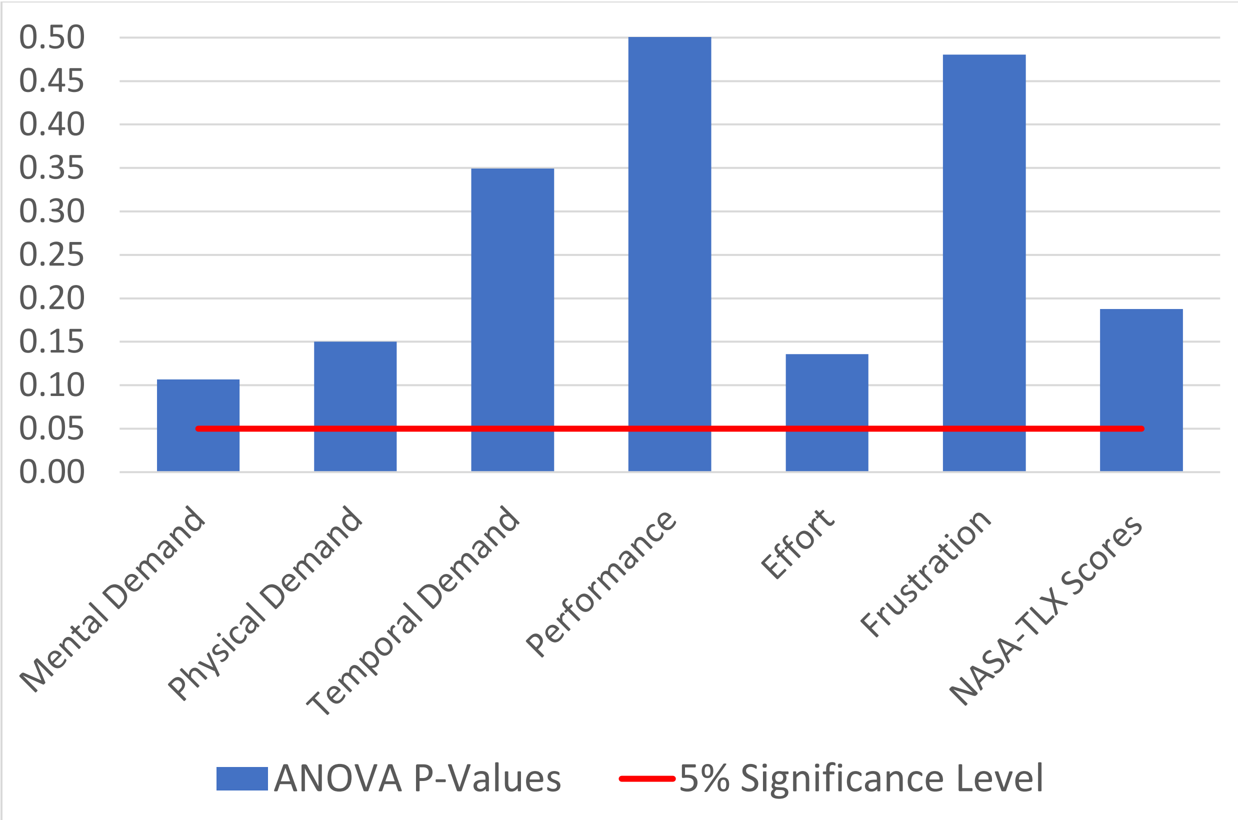


Figure 10: Summary of statistical analysis using a 0.05 significance threshold (Note: performance p-value was equal to 1.0)

6.1.2 Overall Results from the NASA-TLX MWL Survey

The study task resulted in an average MWL score of 57.4 +/- 17.2 (SD) for the control group and 45.4 +/- 15.2 (SD) for the experimental group. Effort was found to be the highest contributor to MWL, followed closely by mental demand. Since the highest possible score was 120, the averages for the two groups indicated that the overall perceived workload of the task was medium to low. When comparing the MWL averages as well as the averages of all subsections of the NASA-TLX, there appears to be a difference between the scores in every subsection except for performance. Unfortunately, these qualitative differences do not hold up under statistical analysis; for every subsection of the NASA-TLX and for the overall MWL scores, the p-values obtained from ANOVA were above the significance threshold of 0.05, therefore, there was no significant difference between the groups MWL.

6.2 Measured Experimental Success

To determine whether MWL and participant confidence had any correlation with experimental success, the experimental setup was inspected after each participant and given a score based on how successful the participant was at completing the experiment. Participants were given points based on the following:

- One point for each servo placed properly on the model.
- One point for each servo properly attached to the corresponding attachment point.
- One point for each module oriented properly.
- One point for each module placed in the proper direction per the axis in supplementary Figure 1 and supplementary Figure 4 (Figure 5) participant instructions.
- One point for each module connected properly using the given connection cables.

Individual scores are shown in Table 7.

Table 7: Scored rubric measuring experimental success (actual performance).

Participant	Servos Placement			Servo Attachment			Module Orientation			Module Direction			Module Connection			
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	Total
1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	27%
2	1	0	0	1	1	1	1	1	1	0	1	1	0	1	1	73%
3	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	87%
4	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	80%
5	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	80%
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%
7	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	93%
8	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	87%
9	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	80%
10	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	87%
11	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	87%
12	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	80%
13	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	73%
14	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	87%
15	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	87%
16	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	87%

The experimental success average was 81%, indicating a relatively high level of success from all participants. For the control group the average was 77% and for the experimental group the average was 85%. ANOVA on the two groups resulted in a p-value of 0.3119, indicating there was no statistical difference between the experimental success of the two groups. This level of success was unexpected given perceived success obtained using the NASA-TLX was considerably lower.

6.3 Participant Interview Responses

During the participant interviews, participants were asked a series of questions to elicit responses related to the experiment. Q1, Q3, and Q7 (refer to 4.6.2 Post-Experiment Interview Questions) focused on obtaining general opinions and comments regarding aspects of the experiment that could be used to inform the NASA-TLX results. The remaining questions were used to obtain more specific responses that could help meet the research objectives.

Q1 (“Describe the effectiveness of the instructions.”) was used to obtain a general response from the participants on the efficacy of the instructions. Summarized in Table 8, responses from participants indicated that there were mixed opinions on the effectiveness of the instructions. The wording of individual responses varied greatly, therefore, instead of categorizing responses along a scale, the responses were labelled as either positive or negative. A positive response indicated that the participant thought the instructions to be effective to some degree, and a negative response indicated that the participant thought the instructions to be not effective to some degree. There was a significant potential for bias in this analysis due to the subjective nature of labelling responses either positive or negative, therefore, the use of Table 8 was limited.

Table 8: Summary of overall positive and negative responses to Q1.

Effectiveness	Total	Instruction Set 1	Instruction Set 2
Negative Response	7	3	4
Positive Response	9	5	4

The quantity of positive and negative responses to Q1 were too similar to make any meaningful conclusions.

Instead of general effectiveness, the objective of Q3 ("How would you describe the overall difficulty of the task presented to you?") was to obtain general opinions on the difficulty of the experimental task, and to better inform the MWL scores obtained using the NASA-TLX. From the responses to this question, it was found that seven participants indicated that the difficulty was low, seven indicated that the task was a medium level of difficulty, and one indicated that the difficulty of the task was high. One participant was an outlier and indicated that they found the task extremely difficult at the beginning and then easy at the end of the experiment. Those participants that elaborated on the reported difficulty cited hindrances like lack of experience or familiarity with mechatronic systems, while others reiterated points made in Q1 and Q2 ("What aspects of the experiment aided or hindered your understanding of the instructions?"). There did not appear to be any correlations between those who found the task difficult and non-engineering background.

The goal of the last interview question, Q7 ("How well do you believe you executed the instructions?") was to gauge the participants' general confidence in their performance to better inform the performance scores recorded during the NASA-TLX survey. Like the performance scores, the responses to this question were mixed; two participants indicated that they did very badly, and three believed their success was poor but were positive about their results due to movement from the servos. The remaining eleven participants' confidence in their performance ranged from slightly low to slightly high.

It was found that many of those who responded with a positive level of confidence also indicated that they believe they missed a step in the instructions or setup. It was also found that those who responded

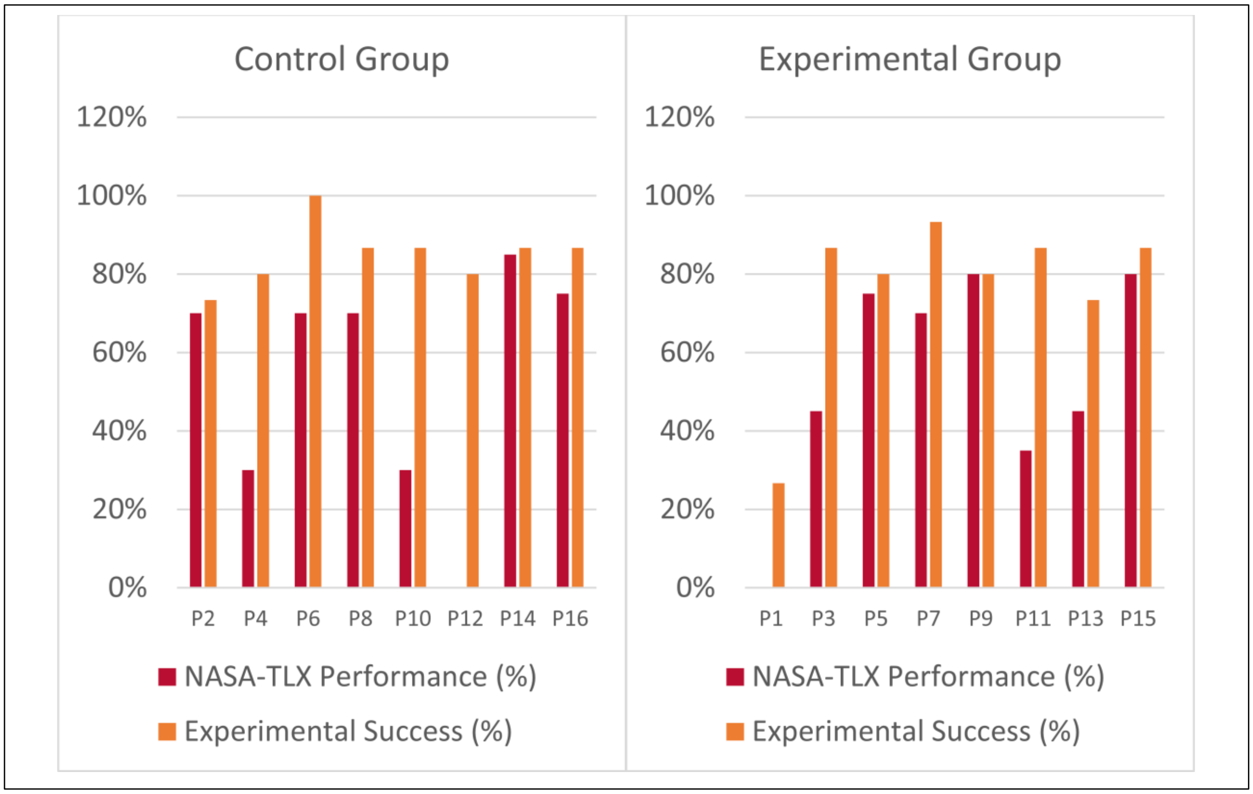


Figure 11: Perceived performance compared to experimental success.

positively also mentioned the model moving. Perceived performance and experimental success are compared in Figure 11.

The comparison in Figure 11 illustrates that perceived performance did not necessarily correlate with actual performance, since for some participants the two values were closely related while for other participants, they were extremely different. This was supported by responses to Q7 which were equally varied in terms of positivity regarding performance.

The four remaining questions (Q2, Q4, Q5, and Q6; refer to 4.6.2 Post-Experiment Interview Questions) all elicited similar responses which made it appropriate to pool the responses across all four questions. All significant comments were sorted into seven categories: exclusive to the control group, exclusive to the experimental group, engineering background exclusive, non-engineering background exclusive, mixed background group N>1, and mixed background group N=1. Results are shown in Table 9, with

A. Exclusive to Control group	1. Include more labelling (P5, P7*, P11) 2. Being able directly and physically interact with the modules and model (P5, P15) 3. Clarify where the connectors are (P1*, P13*)
B. Exclusive to Experimental group	1. Unclear what "gold mark" referred to in instructions (P4, P8*, P16) 2. Unclear whether component role (yaw/pitch/roll servo/module) was something inherent or something assigned through setup (P2, P4, P6)
C. Engineering Background Exclusive	1. Make the instructions steps shorter and more concise (P3, P4, P6)
D. Non-engineering Exclusive	1. Improve figures (generalized statement) (P1*, P8*) 2. Figure 1 arranged opposite of the model (P8*, P13*)
E. Mixed Background Group N>1	1. Language used was vague, confusing, or considered jargon (P1*, P6, P7*, P9, P13*, P14*) 2. The part list was helpful (P2, P4, P8*, P9, P13*, P14*) 3. Unclear what the blue boxes in F4 referred to (P3, P4, P6, P14*, P15) 4. Figures and instructions do not complement each other (P2, P3, P8*, P9) 5. Elaborate on the servo attachment points (P4, P6, P13*, P14*) 6. Include pictures of things already installed (P3, P4, P6, P14*) 7. Assumed colour coding resulting in confusion (P7*, P8*, P15) 8. Make instructions more in depth (P8*, P9, P12) 9. Pair instructions with pictures (P13*, P15, P16) 10. Make figures more universally understandable (P9, P14*) 11. Elaborate on terms and components (P4, P7*) 12. Make components easier to connect (P7*, P10)
F. Mixed Background Group N=1	1. Include more figures (P3) 2. Include more safety instructions (P3) 3. Provide task context (P3) 4. Explain how orientation is used in setup (P6) 5. Include more colour coding (P7*) 6. Mention that modules are chained together during setup (P10) 7. Introduce F3 later (P11) 8. Orange and red indicators were too similar (P12) 9. Use the red LED more for error communication (P12) 10. Lack of sequential steps with visual representation (P13*) 11. Provide and overall schematic of the system (P13*)

orange text representing a negative response and blue text representing a positive response.

Table 9: Summary of interview responses (orange = negative response and blue = positive response).

Participant numbers differentiated by a star ("*") represent participants from a non-engineering background. Participant responses providing general opinions on the how helpful the supplementary figures (Figure 5) were in relation to completing the experiment were removed from Table 9 and were summarized in Figure 12.

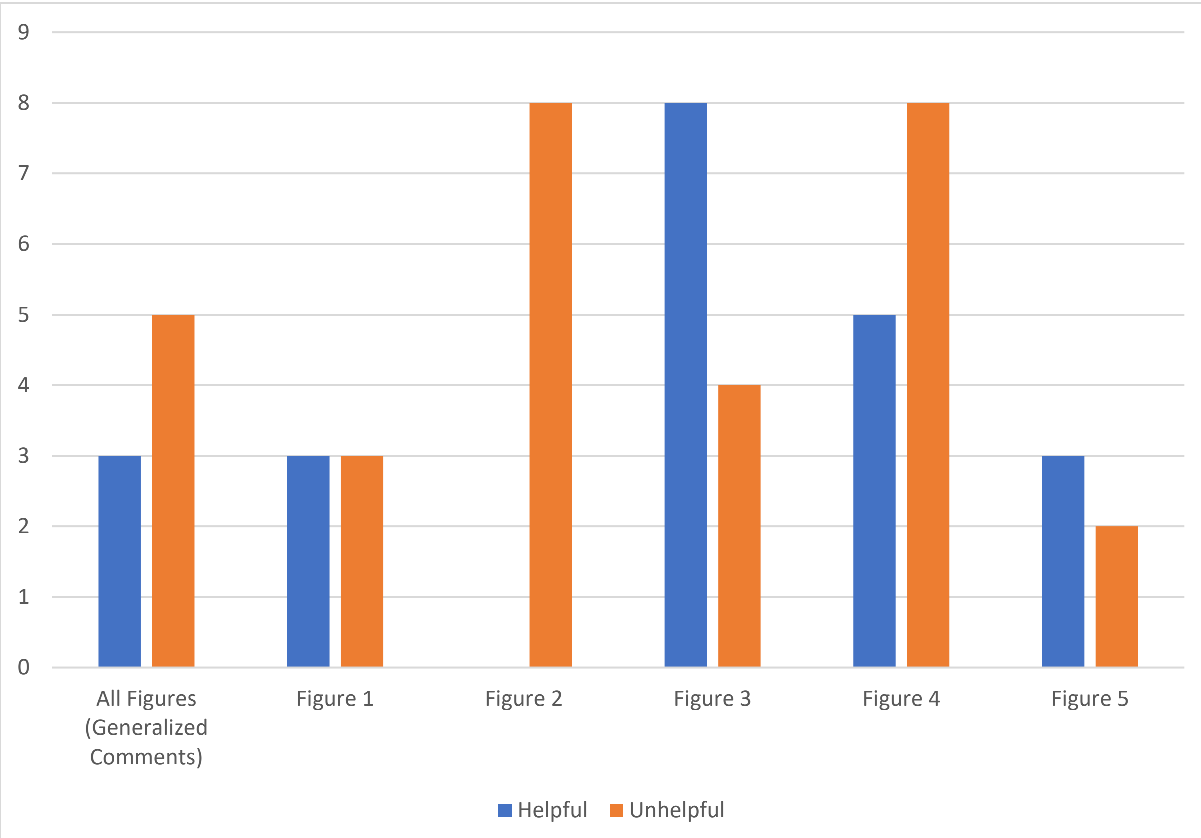


Figure 12: Positive and negative responses to supplementary diagrams given to participants summarized from interview responses.

Figure 12 shows that the helpfulness of the supplementary figures varied between participants, however, most participants found both the second and fourth figures unhelpful.

6.3.1 Attempting to Minimize Observer Bias

Extracting information from participant responses was deemed an effective way to answer the research question, but unfortunately the method is susceptible to researcher bias. The researcher in this study was the only individual reading the interview transcripts and determining whether statements from participants were relevant or significant. To try and mitigate this bias, the researcher reviewed the transcripts, indiscriminately extracted all comments that were complete ideas, tabulated all comments and responses, and then grouped like responses and comments. Upon completion, the transcripts were re-reviewed to ensure that significant responses were not missed. Though bias could still exist, this method was undertaken to try and minimize the risk of said bias.

6.4 Analysis of NASA-TLX Subsections

6.4.1 Subsection ANOVA Results

ANOVA was performed on each of the NASA-TLX subsections (refer to section 6.1 Mental Workload Scores) to determine whether there was any statistical difference between the control and experimental groups in any of the subsections. Like the overall MWL scores, there was no significant difference found between groups. Of the six subsections, mental demand, physical demand, and effort had p-values close to or below 0.15, suggesting the difference in means might be found statistically different if the experiment was performed using a larger population size or if confounding factors were removed. ANOVA results from the remaining subsections, temporal demand, performance, and frustration, were at a minimum around 0.35, far higher than the other three subsections. The following analysis investigates how certain subsections may have influenced statistical analyses performed in this study.

6.4.2 Subsection Omittance Investigation

The largest contributing factors to the lack of significance in the MWL results appeared to be the small sample size and the large variability within the NASA-TLX subsections. Though the effect of sample size could not be eliminated without further testing, the effect of subsection score variability could be explored through the removal of subsections. The practice of removing subsections is not novel, though it is usually done before the experiment is conducted. For example, Hart [7] found twelve articles in their 2006 review of NASA-TLX research studies that removed subsections from the NASA-TLX while still successfully using it to measure MWL. The study conducted by Said et al. [9] is a more recent example where subsections were removed, as physical demand was omitted from the index since the task under analysis was a computer program. In this research study, no subsections were omitted from the final NASA-TLX scores; instead, the temporary omittance of one to two categories at a time was investigated to better understand the effects of individual subsections and their corresponding variability on the final MWL scores. The effect of removing subsections directly relates to the nature of ANOVAs. ANOVA looks at the between group variance of the control and experimental groups in addition to the within group variance of each group to determine whether there is a statistical difference between the two groups' means. Small between group variance reduces the statistical difference between groups since the means are closer together. Similarly, large within group variance reduces the statistical differences between groups by increasing the overlap of the distributions. ANOVA analysis takes both variance types into account to statistically determine whether there are differences between group means. Omitting sub-groups changes both the between group and within group variances of the final score and provides insight into the effects of each group on the final scores.

This investigation was undertaken to better understand why delayed feedback had no significant impact on MWL. The statistical conclusion that real-time feedback does not significantly decrease mental workload as compared to delayed feedback is a valid conclusion, however, observations made during the experiment and the qualitative differences between the experimental and control groups suggest that there was a difference. This omittance investigation was conducted to explore the validity of this conclusion.

In this study, justification for the omittance of the physical demand, temporal demand, performance, and frustration was established, therefore, the effect omitting each subsection may have on the final MWL score was investigated. The statistical results are shown followed by the reasoning and analysis resulting from omitting these components of the full NASA-TLX.

It should be noted that for those subsections that were predicted to have low impact on the NASA-TLX before undertaking the study, such as physical demand and temporal demand, were not omitted to avoid researcher bias. Their omittance after the study was only a temporary removal to see whether the assumption that they were low impact was valid, and whether there variability positively or negatively impacted the statistical analyses performed in this study.

6.4.2.1 Statistical Results from Omittance Investigation

Subsections under investigation were removed and the overall MWL score was calculated. ANOVA was then used to determine whether there was a statistical difference between altered overall MWL scores. This process was repeated for each subsection under analysis and again with pairs of subsections removed. The results of this investigation are shown in Table 10.

Table 10: MWL scores resulting from the removal of specific subsections.

Physical Demand	Temporal Demand	Performance	Frustration	Final MWL P-Value Result
X				0.2282
	X			0.2469
		X		0.1091
			X	0.1372
X	X			0.3016
X		X		0.1357
X			X	0.1796
	X	X		0.1450
	X		X	0.1854
		X	X	0.05887

After testing whether the removal of subsections would increase the significance of the statistical analysis, it was found that removing performance and frustration from the final score resulted in a p-value of 0.05887. This p-value is much closer to the chosen 5% significance level than the original 0.1878 obtained from the complete NASA-TLX. This suggests that the high variability between individual performance and frustration scores reduced the statistical differences between the control and experimental groups. In other words, if there was less variability between frustration and performance scores, there may have been a statistically significant difference between the MWL scores of the control and experimental groups. Figure 13 illustrates the results from the ANOVA analysis graphically.

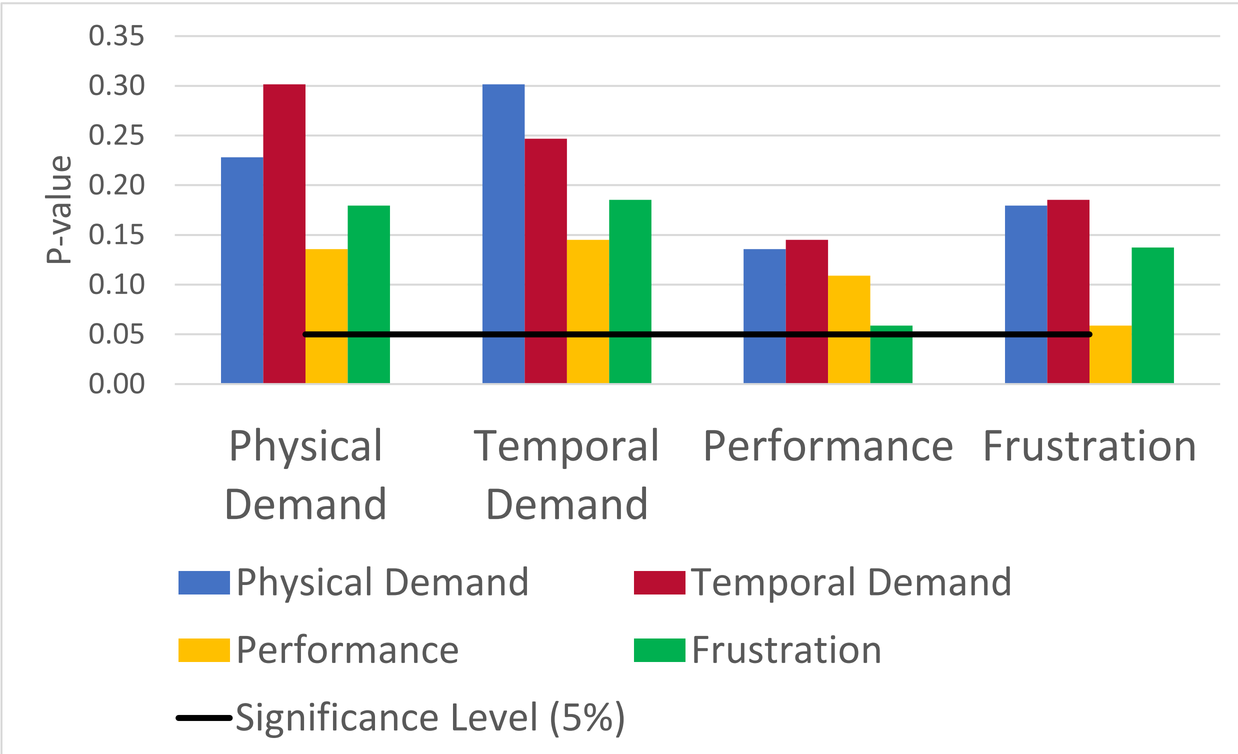


Figure 13: Summary of p-values obtained from omission ANOVAs using a significance threshold of 0.05. Category along x-axis represents primary subsection being omitted while column and colour indicate the secondary subsection being omitted.

6.4.2.2 Physical Demand

The first subsection under investigation, physical demand, had the lowest average contribution among all subsections. Physical demand was included in the NASA-TLX to address tasks where physical exertion is required and contributes to perceived workload, such as assembly tasks or operation of equipment. In those task categories, generally the primary focus of the NASA-TLX analysis is on the physical aspects of the tasks, with secondary focus on the mental aspects of the task. In this experiment, the primary focus was on the mental aspects related to interaction with the experimental system with a secondary focus on the physical consequences of said interaction. The task given to participants was designed to require little physical effort, and the subsection results supported this; physical demand had the lowest average score of all subsections. Due to the apparent insignificance of physical demand in the consideration of MWL in this scenario, its omittance was explored.

Looking at Table 6, the between group variance and within group variance of physical demand are low, especially compared to the other subsections. Omitting physical demand would therefore result in greater within group and between group variances within the summed mental workload. The results of the omittance investigation support this, since omitting physical demand had the effect of increasing the resulting ANOVA p-value performed on the final MWL value (Table 10). When omitted along with another subsection, it also increased the p-value relative to p-value obtained by only omitting the other subsection. For example, omitting temporal demand alone resulted in a p-value of 0.2469 while omitting physical and temporal demand resulted in a p-value of 0.3016 (Table 10). Given the effect of omitting physical demand, it was concluded that physical demand did not contribute to decreasing the statistical difference between the MWL of the control and experimental groups.

6.4.2.3 Temporal Demand

During the experiment, participants were provided with as much time as they desired, and they were not penalized or incentivized to finish quickly. As such, it was expected that temporal demand would be insignificant to MWL. The NASA-TLX results seem to support this assumption, since the average score in both the experimental and control groups was the third lowest average among other subsections. Additionally, there were no comments made during the interview process that indicated participants felt rushed, further supporting the assumption. Considering the apparent insignificance of temporal demand, its omittance was also explored.

Inspecting Figure 8, temporal demand had low between group variance with a medium level of within group variance. Compared to the control group, the experimental group had a larger within group variance. Like physical demand, removing temporal demand had the effect of increasing the p-value obtained from ANOVAs performed (Table 10). Therefore, it was concluded that temporal demand did not contribute to decreasing the statistical difference between the MWL of the control and experimental groups.

Earlier in this omittance investigation, the removal of physical demand and temporal before the experiment were discussed. Figure 13 illustrates how removing physical demand or temporal demand could have affected the statistical analysis of the results, showing that by doing so the statistical difference between the control group and the experimental group would have been significantly less. These results suggest that physical demand and temporal demand were likely still an important aspect of MWL for participants, though were low in overall contribution.

6.4.2.4 Performance

Of the remaining subsections, performance was considered one of the most likely to have affected the significance of the MWL scores. When averaged, both the control group and the experimental group had identical means, suggesting that the differences between instructions sets did not visibly affect perceived performance. Conversely, the variability between scores in both groups was extremely high (Table 6; control group variance: 31.36, experimental group variance: 35.64), suggesting that individual participants had wildly different opinions on how well they completed the experiment. This indicates that the equal means were likely the result of coincidence, and that external factors not measured during the study were more influential on perceived performance than the controlled variables. Since the controlled variables within the study did not appear to significantly affect perceived performance, the omittance of the performance subsection was also explored in this investigation.

Omitting performance from the overall MWL calculation resulted in lower p-values obtained from the ANOVA calculations (Table 10). Inspecting Figure 8 shows that while the between group variance was zero, the within group variance of both the control and experimental groups was extremely high. Given the effect of removing performance, it was concluded that this subsection's high within group variance negatively impacted the investigation looking into the effects of feedback on MWL.

6.4.2.5 Analysis Between Perceived Performance and Measured Performance

Looking at the NASA-TLX performance scores, the control group had an average score of 9.3, with a minimum of 3 and range of 16 (Figure 8). The experimental group also had an average score of 9.3 but had a minimum of 3 and a range of 17 (Figure 8). To put this into perspective, the NASA-TLX scale ranged from 1 to 20, where 1 would indicate the participant performed perfectly while 20 would indicate the participant failed. As stated earlier, these scores indicated that individual participants had wildly different opinions on how well they completed the experiment. Responses to Q7 ("How well do you believe you executed the instructions?") were equally varied.

To determine whether the high variability and equal averages seen in both performance scores were accurate representations of individual performance, perceived performance was compared to actual performance. Using the metric shown in Table 7, the average actual performance score was 81%. This score indicated that despite the low to average levels of perceived performance, most participants were able to complete the experiment with only minor errors. This is further illustrated in Figure 14.

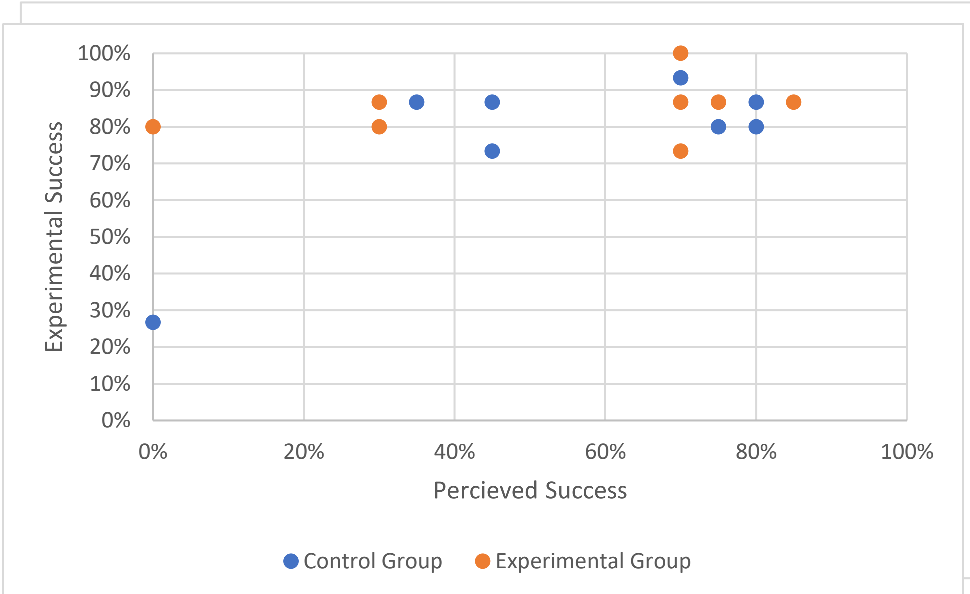


Figure 14: Scatterplot showing the relationship between experimental success and perceived success.

Excluding one outlier score of 27%, all experimental success values were above 73%, compared to the perceived success values which occurred between 0% and 85%. This discrepancy between values further suggests that there were additional factors affecting the confidence levels of participants.

From the participants' responses, two significant factors affecting participant confidence were identified: participants' lack of a reference point and a lack of understanding or familiarity regarding the instructions. Referencing Table 9, four people indicated that they wanted pictures of components already installed (E6: P3, P4, P6, P14*) and three additional participants indicated that instructions would be better if paired with pictures (E9: P13*, P15, P16). Both responses suggested that participants' performance and understanding would be improved if they had a pictorial point of reference. Furthermore, Figure 12 illustrating participants' opinions on the figures suggests that many found the supplementary figures unclear or confusing. The assumption is that a lack of understanding would have negatively impacted confidence in individual performance and lead participants to assume they made more mistakes than they did, thus explaining the discrepancy between perceived performance and experimental success.

6.4.2.6 Frustration

The last subsection under investigation for possible omittance was frustration. The frustration subsection was considered important to the final MWL calculation, however, like performance the within group variance was extremely high (Table 6; control group variance: 45.14, experimental group variance: 31.92). Unlike performance, the average scores between the groups were unequal, yet the results from ANOVA still suggested that there was no statistically significant difference. Therefore, due to the possibility that factors not measured during the study were more influential than the controlled variables, the omittance of the frustration subsection was also explored.

Omitting frustration from the ANOVA calculations had the effect of decreasing the obtained p-values. Looking again at Table 6, this result is supported by the high within group variances seen in both the control group and experimental group. Given the effect of removing frustration, it was concluded that this subsection's high within group variance negatively impacted the validity of the second research question.

Unlike performance, actual frustration was not measured, therefore, the causes of the subsection's high score variability could only be inferred from what evidence was present. The primary evidence available were the interview responses summarized in Table 9 that suggested that many participants found significant deficiencies in the instructions, figures, and experimental setup. Individual reception of these deficiencies would have likely contributed to the higher variability seen within the frustration scores. It was also hypothesized that external and difficult to measure factors were more influential in controlling frustration than factors that were measured by the researcher. The most obvious of those external factors would be personality; an individual's emotional response to obstacles and problem solving would have a far greater effect on frustration than the controlled variables within the study. Similarly, the amount of time participants spent on other activities that day (i.e., professional work, school assignments, social outings, travel, group projects, etc.) would be assumed to lower individual patience with the experimental task. Though these difficult to capture aspects of human factors were overlooked during the design of the interview questions and thus not recorded, it is still possible that they contributed to the high variability seen in the frustration scores.

6.4.2.7 Conclusion to Omittance Investigation

The omittance investigation found that though performance and frustration are important factors contributing to the MWL of novice users, the high within group variances of the subsections negatively affected the statistical analysis used to determine whether real-time feedback influenced mental workload. In conclusion, this study found no statistical difference between the control and experimental group in terms of MWL. This study does, however, recommend that additional analysis using an improved study and additional participants be performed to determine whether real-time feedback reduces MWL in relation to mechatronic devices.

Chapter 7: Discussion

In the context of this research, the mechatronic design process was considered a multifaceted process incorporating physical hardware, software, and documentation. This study included simulated scenarios in which novices undertook a task, reported on their cognitive load, and answered qualitative questions to better understand the implementation process of mechatronic systems.

7.1 Issues observed by novices (RQ1)

The first research question asked, “what issues of mechatronic design are consistently identified by novices after interaction with a mechatronic system?”. In section 2.1 Designing for Novices, background research into designing for novices was taken and summarized into three categories: simplicity, transparency, and clarity. During analysis of participant responses, it was found that the responses themselves could be connected to at least one of these concepts. For example, responses asking for elaboration on certain concepts (i.e., E5: “Elaborate on the servo attachment points.”, Table 9) could be considered requests for greater clarity. Likewise, criticism of the language complexity and length (i.e. E1: “Language used was vague, confusing, or considered jargon.”, Table 9) could be considered requests for greater simplicity, and responses looking for pictures of components already installed (i.e. E6: Include pictures of things already installed.”, Table 9) could be considered requests for greater transparency. To explore this connection, the responses summarized in Table 9 were subjectively organized under the concepts of simplicity, clarity, and transparency to determine whether they should be included in the design recommendations in some form. Responses that suggested a need or desire for more information were sorted under clarity, responses that expressed confusion pertaining to or a desire for less complexity were sorted under simplicity, and responses expressing a desire for background and deeper knowledge of the system were placed under transparency. The results of this exercise are shown in Table 11.

Table 11: Responses sorted under clarity, simplicity, and transparency (italicized comments relate to additional concepts).

Descriptor		Participant Numbers
	Clarity	
A. Exclusive to Control Group	• A1: Include more labelling • A3: Clarify where the connectors are	• (P5, P7*, P11) • (P1*, P13*)
B. Exclusive to Experimental Group	• B1: Unclear what "gold mark" referred to in instructions	• (P4, P8*, P16)
	• *B2: Unclear whether component role (yaw/pitch/roll servo/module) was something inherent or something assigned through setup*	• *(P2, P4, P6)*
E. Mixed Background Group N>1	• E1: Language used was vague, confusing, or considered jargon	• (P1*, P6, P7*, P9, P13*, P14*)
	• *E2: The part list was helpful*	• *(P2, P4, P8*, P9, P13*, P14*)*
	• *E3: Unclear what the blue boxes in Figure 4 referred to*	• *(P3, P4, P6, P14*, P15)*
	• *E4: Figures and instructions did not complement each other*	• *(P2, P3, P8*, P9)*
	• E5: Elaborate on the servo attachment points	• (P4, P6, P13*, P14*)
F. Mixed Background Group N=1	• F1: Include more figures	• (P3)
	• *F3: Provide task context*	• *(P3)*
	• F7: Introduce F3 later	• (P11)
	• F8: Orange and red indicators were too similar	• (P12)
	• F9: Use the red LED more for error communication	• (P12)
	Simplicity	
C. Engineering Background Exclusive	• C1: Make the instructions steps shorter and more concise	• (P3, P4, P6)
D. Non-engineering Background Exclusive	• *D1: Improve figures (generalized statement)*	• *(P1*, P8*)*
	• D2: Figure 1 arranged opposite of the model	• (P8*, P13*)
E. Mixed Background Group N>1	• *E7: Assumed colour coding resulting in confusion*	• *(P7*, P8*, P15)*
	• E10: Make figures more universally understandable	• (P9, P14*)
	• E12: Make components easier to connect	• (P7*, P10)
F. Mixed Background Group N=1	• *F5: Include more colour coding*	• *(P7*)*
	• F10: Lack of sequential steps with visual representation	• (P13*)
	Transparency	

A. Exclusive to Control Group	• A2: Being able to interact with the modules and model directly and physically	• (P5, P15)
E. Mixed Background Group N>1	• *E6: Include pictures of things already installed*	• *(P3, P4, P6, P14*)*
	• E8: Make instructions more in depth	• (P8*, P9, P12)
	• *E9: Pair instructions with pictures*	• *(P13*, P15, P16)*
	• E11: Elaborate on terms and components	• (P4, P7*)
F. Mixed Background Group N=1	• F2: Include more safety instructions	• (P3)
	• F4: Explain how orientation is used in setup	• (P6
	• F6: Mention that modules are chained together during setup	• (P10)
	• F11: Provide an overall schematic of the system	• (P13*)

During the exercise, it was found that some responses (*italicized lines,* Table 11) pertained more accurately to other concepts: in addition to simplicity, clarity, and transparency, the concepts of consistency and context were identified as being important areas of design for novice users. The applicability of these two additional concepts is explored further in the following sections.

7.1.1 Consistency

Consistency in design pertains to two areas. Firstly, it relates to the accuracy of system outcomes, or does operation 'A' lead to outcome 'A' every time? Secondly, consistency relates to the conformity of different parts of the designed system. Good consistency in design would mean that concepts are presented in the same format every time, and descriptions, figures, and terms accurately represent their corresponding concepts and are properly cross-referenced. From the earlier exercise shown in Table 11, the responses below were found to pertain to context more accurately.

- B2 (P2, P4, P6): "Unclear whether component role (yaw/pitch/roll servo/module) was something inherent or something assigned through setup."
    - Participants were confused by references to roll/pitch/yaw servo/module due to the instructions. Responses from participants support the conclusion that a lack of consistency between the instructions, part list, and housing construction were to blame

for this confusion. The instructions refer to each module and servo by their role, while the part list indicates that there are three identical modules and three identical servos. Furthermore, the housing construction for the modules and servos used multicoloured bricks which resulted in each housing having different colour combinations, further reinforcing the idea that each module was unique.

- D2 (P8*, P13*): "Figure 1 arranged opposite of the model."
  - Supplementary Figure 1 (section 4.5 Instruction Set Design, Figure 5) representing the plane caused confusion since the model itself was pointed in the opposite direction as the paper diagram. This response supports the conclusion that participants who struggled with this issue were confused by the figure being inconsistent with the model's direction. This issue could also be contributed to the participant's ability to transfer spatial awareness exacerbated by the inconsistency between the model and figure.
- E2 (P2, P4, P8*, P9, P13*, P14*): "The part list was helpful."
  - The part list was well received by participants, and this was made clearer during the interviews. During sorting this response didn't fit well under clarity, simplicity, or transparency. From the interviews, it was clear that participants liked the part list largely because it allowed them to identify components and concepts using pictures and labels. This pertained more closely with consistency rather than any of the original three concepts since it was the consistency between the labels, instructions, pictures, and components that allowed participants to use the parts list in a positive manner.
- E3 (P3, P4, P6, P14*, P15): "Unclear what the blue boxes in F4 referred to."
  - The blue boxes and the rectangular prisms representing the modules in supplementary Figure 4 (section 4.5 Instruction Set Design, Figure 5) were not consistent with the

modules’ housing. From the interview it was clear that the participants that commented on it were unsure what the objects in the figure represented since the representations were not consistent with the actual modules.

- E4 (P2, P3, P8*, P9): ”Figures and instructions do not complement each other.”
    - When asked how the figures and instructions complemented each other, four individuals commented that they did not complement each other without elaborating on the causes. While clarity was an issue with the figures, this response was deemed to more accurately reflect issues with consistency. Other responses, such as the previously discussed E3 (Table 9), support the conclusion that the figures and instructions were not considered complementary because the figures and instructions were not consistent with what was being described.
- E7 (P7*, P8*, P15) and F5 (P7*): ”Assumed colour coding resulting in confusion” and “Include more colour coding.”
    - Colour coding was assumed present due to the multicoloured bricks used in the construction of the model and modules. The modules themselves were inconsistent due to the construction of the casings, resulting in individuals assuming colour coding like that used in the LED indicators when there was no colour coding present in the bricks used. The same participant (P7*) later expressed a desire for additional colour coding, further reinforcing the conclusion that colour coding was assumed when none was present.

7.1.2 Context

Context in design refers to the background, motivation, or situational information that informs the use of a design. Though similar, context differs from transparency in that transparency is the need to “see through” the system and understand its inner workings, whereas context is the surrounding information that contributes to understanding the system. For novice users, context is important because it informs the use of the system, providing information on when, where, and why the system should be used. From the earlier exercise shown in Table 11, the responses below were found to pertain to context more accurately.

- E6 (P3, P4, P6, P14*): ""Include pictures of things already installed."
    - pictures of most components were included in the part list, however, the pictures showed only the components in neutral uninstalled positions. Each piece was shown alone without additional components or in any position resembling the final installed position. Each part was shown without additional context, besides that described in the instructions. This response from participants reflects how many individuals learn by example, whether it is a math professor walking students through example problems or a supervisor showing the new employee how to complete their assigned task the first time. Therefore, it was decided that response E6 supported the need for context instead of transparency, since participants were expressing a need for situational information regarding the given components.
- E9 (P13*, P15, P16): "Pair instructions with pictures."
    - Responses in E9 were categorized under context since these were requests for pictorial references connected to specific instruction steps, ergo pictorial context. The support for E9 representing a desire for context were identical to response E6 due to the similarity in content.
- F3 (P3): "Provide task context."
    - This response provided additional support that there were additional concepts (consistency and context) identified by novice users besides the initial three identified (context, transparency, and simplicity) since it did not readily fit under any of them.

Though similar, context and transparency differ enough that they are considered distinct concepts.

7.1.3 Issues Identified When Installing Mechatronic Components

The first research question was "what aspects of mechatronic design are consistently identified by novices after interaction with a mechatronic system?". Conclusions from section 7.1 suggest that the positive and negative aspects of mechatronic design that users identify after interaction with a mechatronic system generally fall under the categories of clarity, simplicity, transparency, consistency, and context. Positive aspects generally use those five concepts to enhance the usability of the mechatronic system, whereas negative aspects generally manifest as deficiencies in one or more of those five areas. Comparing these results to the suggestions from Hooper [1] and Denny et al. [4], these conclusions are consistent with the logic shown by the two articles are. The inclusion of consistency and context are also logical extensions of the original three summarized concepts (simplicity, clarity, and transparency) because they fill in gaps identified during this research.

7.2 Timing of Feedback on Mental Workload Results (RQ2)

The second research question sought to ask, "is mental workload as measured by the NASA-TLX reduced by real-time device feedback?". During the experiment, the control group and the experimental group were distinguished by the instruction sets they were given. Participants in the control group powered the system after setup while participants in the experimental group powered the system before setup. Through this slight change in instruction, the control group was provided delayed feedback and the experimental group was given real-time feedback using the same LED indicators. Statistical analysis concluded that MWL was not statistically reduced by real time feedback (Table 6). Based on qualitative evidence this was an unexpected outcome: during the experiment it appeared that those given the first instruction set (control group) had greater difficulties completing the given task than those given the second instruction set (experimental group). Based on the small sample size of sixteen, the lack of statistical evidence was not unexpected: a similar study presented in the background research omitted statistical analysis using NASA-TLX scores altogether due to the potential inaccuracies using a low population size (total population in referenced study was nine) [16]. Nevertheless, the omittance of subsections was investigated to determine whether certain subsections may have negatively affected the statistical analysis performed. The investigation found that high variability within the performance and frustration subsections reduced the statistical significance of the ANOVA results (refer to 6.4.2 Subsection Omittance Investigation). Further analysis using a larger population size and an improved study would be needed to provide confirmation of these results.

7.2.1 Differences Between Immediate and Delayed Feedback

Inspecting the data taken empirically, there does appear to be a difference between the control group and the experimental group in terms of MWL, however, statistical tests showed that the result could be accounted for by random effects. One explanation for the lack of statistical evidence comes from an interview response exclusive to the experimental, or the real time feedback group. Three individuals (Table 9, B2: P2, P4, P6) who were all from an engineering background and part of the experimental group indicated that they were confused by references to the roles of components. These participants believed that module and servo role (i.e., yaw, pitch, or roll) was something inherent to the component and not something determined during setup. The significance of this response lies in the fact that three individuals from an engineering background and all given the experimental instruction set designed to provide immediate feedback on the module’s orientation flagged the same issue. The confusion surrounding the modules could be a side effect of the active feedback itself, since the participants needed to actively use the orientation and LED indicators as they went through the instructions. For reference, the control group was instructed not to turn on the modules until all were setup, thereby providing all feedback after they made their way through the instructions. A possible explanation for this phenomenon was that by instructing participants to install all modules before powering, the instructions were unintentionally communicating that the choice of a specific experimental module chosen for each step was irrelevant since they were functionally identical. Conversely, providing continuous feedback to the experimental group unintentionally communicated that they were given a yaw module, a pitch module, and a roll module when in fact they were all functionally identical before setup. The results obtained at the end of the study showed that the MWL scores of the experimental group were lower, though not significantly compared to the control. It is possible that the phenomenon described here unintentionally reduced the MWL of the control group and increased the MWL of the experimental group. Though this likely would not have made the MWL differences between the groups significant, it is possible that it could have further confounded the statistical analysis performed.

7.2.2 Power Analysis

To determine the minimum number of participants needed for statistically significant results at the 5% significance level, power analysis was performed. The degrees of freedom value between groups were one and the within group value was 14, for a total degrees of freedom value of 15. Using [23] as reference, an effect size (f value) of 0.40 was chosen. Using these values along with a significance level of 5% and power level of 95%, power analysis revealed that a sample size of 187 would be needed to achieve a statistical difference between the control and experimental groups using the current experimental setup and instruction sets. Due to the dependence of variables used in the power analysis on the population size obtained during this study and the likelihood of further studies using a larger population size, the power analysis was not taken as a requirement but as a suggestion for future research.

7.3 Effort as a Main Contributor to Mental Workload (RQ3)

At the beginning of this research, it was predicted that effort would be the most significant contributor to MWL as measured by the NASA-TLX. This prediction stemmed from the assumption that novice users may find mechatronic systems inherently difficult due to a steep learning curve. This assumption was further supported by a similar study that used the NASA-TLX to study the usability of a lower-limb exoskeleton by first time users, which found that effort was the highest average contributor to MWL when physical demand was ignored [16] (refer to 2.4 Data Collection and Analysis).

After the study, it was found that for both groups, effort had the highest average score among the six subsections. For seven participants, effort was the highest scored subsection or was tied as the highest. Furthermore, six participants scored effort 15 or higher. These results are supported by responses obtained during Q1 ("Describe the effectiveness of the instructions.") of the interview process, where participants (P5, P6, P7*, P11, P12, P15) indicated that the instructions were effective but vague, requiring a considerable amount of effort to complete the task. Participant 1 commented, "[the instructions] were effective if you could understand them." Similarly, participant 5 described the instructions as "bad for fifth graders but fine for professionals."

The implication of these results is that systems can be improved for novice users by reducing perceived effort (*obj. 2).* Effort is harder to adjust outside of purely physical tasks, but the effort required to use a mechatronic system could be minimized in various ways, such as reducing the complexity of the design or by limiting the amount of learning required to use the system.

7.4 Limitations to the Use of the NASA-TLX in This Experimental Protocol

There are limitations to the NASA-TLX in this experimental protocol; in the conducted experiment, participants were not rushed for time, nor were they graded on the outcome of the experiment. This might not have been typical of the situation in which similar mechatronic devices would be used. If a similar experiment took place in a school or work setting, temporal demand and frustration might be greater since the novice users would be under greater pressure from both a time limit and the consequences of failing. Further study in alternative settings would be required to determine whether subsections other than effort would have a greater effect on MWL.

Another distinct limitation stemming from the subjective nature of the NASA-TLX is how each participant interpreted the subsection questions, since the way participants rated the subsections would depend on how they interpreted them. The effort subsection was especially subject to this interpretation bias, since the relevant question ("How hard did you have to work to accomplish your level of performance?") has the participant compare their subjective physical/mental/temporal demand with their subjective performance. Though this interpretation bias was considered during the creation of the NASA-TLX [6], it is worth considering when analyzing the subsections individually and together.

7.5 Relevance of Frustration and Performance Variability to Research Objectives

As established earlier in 6.4 Analysis of NASA-TLX Subsections, frustration and performance were subsections with high variability between scores. The wide ranges between the lowest and highest scores for both the control and experimental group suggest that MWL is disproportionately affected by individual comprehension of the presented information. This is not to say that the participants are to blame for their own confusion; conveying the information is the direct responsibility of the system's creator. Everyone differs in how they intake and interact with information, and it is a failure of the system's designer if they cannot universally communicate through their design.

This experiment intentionally and unintentionally made the system difficult to understand, however, the high variability within the frustration and performance subsections suggest that it was not equally confusing for everyone. For example, six individuals (Table 9, E1: P1*, P6, P7*, P9, P13*, P14*) found language confusing, five individuals (Table 9, E3: P3, P4, P6, P14*, P15) had trouble linking the blue boxes present in the orientation diagram (supplementary Figure 4, Figure 5) to the blue boxes on the modules, four (Table 9, E5: P4, P6, P13*, P14*) had trouble identifying the servo attachment points from the instructions, and four (Table 9, E6: P3, P4, P6, P14*) suggested including pictures of components already installed. While these participants listed comprehension issues and recorded high MWL scores within the frustration and performance subsections, other participants recorded low scores within the same categories. This highlights one of the primary issues that novice users encounter during interaction with mechatronic devices; communication within the presented system may unintentionally or intentionally cater towards some users but not others. In other words, the design's universal appeal is disproportionately affected by the creator's ability to convey information through their design.

7.6 Recommendations

The first research objective sought to identify what issues novices encounter when interacting with a mechatronic device. This objective was answered through RQ1 and RQ3 and therefore had a very similar answer: Issues that novices identified after interaction with a mechatronic device were generally related to deficiencies of clarity, simplicity, transparency, consistency, and/or context found in the different areas of the system. These deficiencies were found within the instructions, figures, experimental model, and in connections between those subsystems. Furthermore, since effort was the highest contributor to MWL, it was inferred that issues novices encounter could partially be attributed to the high effort requirements in mechatronic systems.

7.6.1 Primary Recommendations

The second research objective (recommendations objective) was informed by RQ2 and RQ3 and sought to provide design recommendations when designing for novices. The findings from the discussion support the conclusion that to design specifically for novice stakeholders:

- **Simplicity should be emphasized, and the system should be intuitive to use.**

- **All presented information should be clear and understandable.**
- **All presented information should be transparent and perceivable.**
- **All sub-systems should be consistent with all other sub-systems in presentation, language, and use.**
- **Information should be presented along with relevant contextual support.**

Though the five categories identified could satisfactorily be applied as design principles themselves, as a design recommendation it was believed necessary to elaborate on their applicability within mechatronic design.

As mentioned previously, the responsibility for achieving user understanding is not in fact on the user, but on the system designer, and there are multiple ways to increase the user's level of understanding.

7.6.2 Recommendation for Communication

The results of this study support several additional recommendations, the first being to **ensure that communication from the designer to the user is clear, concise, and uses universal language.** When creating any systems, it is often easy for the designers to lean on their own vocabulary and methods of communication. Unfortunately, for the end users for whom the system is designed, vocabulary and communication might not be clear among novices. An example of this was discovered in this study when some participants did not understand what a 'servo motor' was. During the design phase, the study's author did not consider this jargon. During subsequent testing and revision, the issue was overlooked since the author, editors, and testers were familiar with the concept. Some other concepts, like the 'servo attachment points', were recognizably jargon but were left in to be vague and confusing with the purpose of eliciting participants' reactions when left to make assumptions and draw connections between the instructions and model. The high variability in frustration and performance suggests that for some, the vocabulary and communication methods used were beneficial or at least understandable, whereas for others they were major stumbling blocks that discouraged further exploration. Therefore, to design specifically for novice users, it is important to use universal design thinking when communicating concepts to those users; terms should be consistent across diagrams, instructions, and labelling, thereby accommodating users who may benefit from alternative methods of learning. Furthermore, concepts should be explained using universal language that can be understood regardless of background.

7.6.3 Recommendation for Visuals

The second additional recommendation suggested by the study was to **include accurate pictorial references to enhance user visualization.** The study found that six separate participants (Table 9, E2: P2, P4, P8*, P9, P13*, P14*) found the primary accurate pictorial reference, the part list, helpful. Conversely, numerous responses (Table 9: D1, E3, E4, E6, E9, E10) indicated that the simplified non-pictorial references used in the supplementary figures (Figure 5) hindered their understanding, and these are comments outside of the general response summarized in Figure 12. Elaborating on the responses, five participants (Table 9, E3: P3, P4, P6, P14*, P15) responded that it was unclear in supplementary Figure 4 (Figure 5) what the blue boxes placed at the corner of each module represented, and four participants (Table 9, E4: P2, P3, P8*, P9) stated that the figures and instructions did not complement each other. Furthermore, when asked for suggestions for improvement, five (Table 9, E6: P3, P4, P6, P14*) suggested including pictures of components already installed and three (Table 9, E9: P13*, P14, P16) suggested pairing instructions with pictures. Lastly, two (Table 9, D1: P1*, P8*) generally indicated that the figures needed improvement, and two (Table 9, E10: P9, P14*) suggested making them more universally understandable. These examples support the conclusion that visual representation is beneficial to novice users but only when it accurately portrays what it is trying to represent. The easiest way to meet this recommendation would be to include pictures of components or highly detailed diagrams if pictures are not possible or available.

Chapter 8: Conclusions and Recommendations

8.1 Research Questions

Sixteen participants interacted with a mechatronic system and experimental model. Eight of the participants (control group) were given an instruction set that told users to power the system after setup and the remaining eight (experimental group) were given an instruction set that told users to power the system before setup, effectively providing a comparison between delayed feedback and real-time feedback. The experimental objectives were to identify issues that novice users identify after interaction with mechatronic systems and to synthesize design recommendations for systems where novices are the primary stakeholders. This was done by answering the following three research questions:

RQ1. What aspects of mechatronic design are consistently identified by novices after interaction with a mechatronic system?

Analysis in section 7.1 found that responses from participants supported five aspects of mechatronic design that were relevant to novices: clarity, simplicity, transparency, consistency, and context.

RQ2. Is mental workload as measured by the NASA-TLX reduced by real-time device feedback?

The NASA-TLX results were used to determine whether MWL was reduced by real-time device feedback. This was done through comparison of the control group and experimental group within the experiment. The control group was given the first instruction set which directed participants to power the system after setup, effectively providing delayed feedback, while the experimental group was given the second instruction set which directed participants to power the system before setup, effectively providing real-time feedback. ANOVA at the 5% significance level performed on the subsections and final MWL scores suggested that there was no statistical evidence that MWL was reduced by real-time device feedback.

RQ3. Is effort the biggest contributor to mental workload as measured by the NASA-TLX when installing a mechatronic system?

The NASA-TLX was used to measure the MWL of participants who participated in the study. The answer to this research question hypothesized that effort would be the largest contributor to MWL. This prediction assumed that users find mechatronic systems inherently confusing and difficult due to their steep learning curve. Results from the NASA-TLX showed that effort, followed closely by mental demand, was the highest contributor to MWL. This conclusion corroborated results from [16] which found effort to be the highest contributor (when physical demand was eliminated) when learning to use a mechatronic device from a structured resource, such as an instruction set.

8.2 Research Objectives

The first research objective was to identify issues encountered by novices during interaction with mechatronic devices. The issues identified during this study are summarized below:

- MWL and understanding were disproportionately affected by the designer's ability to convey information through their design.
- Effort is usually the greatest contributor to MWL for novice users when interacting with mechatronic systems.
- Most issues directly identified by participants related to deficiencies in clarity, simplicity, transparency, consistency, and/or clarity.

The second research objective was to synthesize design recommendations relevant to designing specifically for novice users. The recommendations synthesized throughout the study are summarized below:

- Simplicity should be emphasized, and the system should be intuitive to use.
- All presented information should be clear and understandable.
- All presented information should be transparent and perceivable.
- All systems should be consistent with all other systems in presentation, language, and use.
- Information should be presented along with relevant contextual support.
- Ensure that communication from the designer to the user is clear, concise, and uses universal language.
- Include accurate pictorial references to enhance user understanding and visualization.

The novice focused design recommendations presented by this study are not established rules, but conclusions reached by the author based upon the experimental results obtained. Further exploration of novice focused design should be conducted to support and elaborate upon the recommendations put forward here.

Chapter 9: Limitations and Future Research

9.1 Limitations of the Results

9.1.1 Population Size

The experimental study performed in this research was subject to several limitations, the most significant of which was limited participation. Over two days, the study was only able to achieve a total of 16 participants, despite active recruitment efforts. Power analysis (7.2.2 Power Analysis) revealed that the suggested number of participants needed for conclusive statistical results was 187, far higher than the achieved 16. With this low number of participants, there is the possibility that the results could have been skewed by a small number of participants in a way that may not accurately reflect values or responses typical for the target population. Furthermore, the low study population was likely not enough to normalize the NASA-TLX values or provide significant support to the repeated points that arose during the interview process.

9.1.2 Demographic Distribution

Related to the previous limiting factor, the small population size, the demographic sizes within the tested population were skewed significantly. Though both the engineering and non-engineering participants obtained for the study were considered novice mechatronic users, only five of the sixteen participants were from a non-engineering background. The consequence of these skewed demographic sizes was a significant engineering bias. With this bias, the average NASA-TLX scores could have been affected, and issues significant to non-engineering participants could have been missed during the interview process.

Another consideration is that having mixed background could have increased subsection variability. Figure 9 suggests that the subsection means for participants from non-engineering backgrounds were slightly higher than those from engineering background. Though the demographic population sizes and study population size were too small to support these observations or make conclusions regarding the overall populations themselves, there is the possibility that there is a statistically significant difference in the MWL of the two groups when they interact with mechatronic devices. Therefore, by having five participants from a population that might be statistically distinct could have increased the variability of the NASA-TLX results. Further testing would be needed to remove this bias from consideration.

9.1.3 Personal Factors

The results of this research study were also limited by personal factors not measured by the experiment. The NASA-TLX by nature is a subjective tool, and the scoring of each subsection is affected by internal and external personal factors. For an internal example, two individuals with different problem-solving approaches could rate the perceived amount of effort needed completely differently for the same task. Another internal example would be the prior experiences that may have contributed to reducing the MWL of the individual during interaction with the device, such as childhood experience with the toy bricks used in the experiment. For an external example, an individual who completed the experiment early in their day before any errands were complete and an individual who completed the experiment after a day of hard work could rate the frustration of the task completely differently, assuming that their patience with the task would be affected by other events during the day. Similarly, external factors like amount of sleep or time of day could potentially increase stress in individuals, thereby increasing MWL. Therefore, the results of this study are limited by the unknown effects of these personal factors since they were not measured over the course of the experiment.

9.1.4 Environmental Factors

This research study was further limited by the environment in which the experiment was conducted. The experiment took place over the course of two days, with participants given access to the experiment one at a time. During the experiment, participant’s questions were not answered unless they were deemed not to affect the validity of the experiment, and participants were given as much time as possible. While this environment was conducive to an experimental setting, it did not accurately reflect the kind of scenario the experiment itself was portraying. In a real-world setting, there would likely be constraints and consequences associated with the scenario. Conversely, there would likely be additional resources for the user to access online or in person. Further testing in a real-world environment would be necessary to determine their effects on user MWL.

9.2 Future Research

The experiment conducted in this research resulted in a significant amount of information relevant to designing for novice users, however, further research would be necessary to reassess and validate the recommendations made. Power analysis suggested that for future studies the population should consist of a minimum of 187 participants, however, due to factors used in the power calculations, this is only taken as a suggestion. The better approach to improving future studies would be to consider lessons learned during this study and apply them in future experiments. This research could be thought of as an evaluation of the experimental protocol considering the initial pilot study conducted and the size of the experimental population. Taking this into account, there were factors identified during this research that could be applied to future iterations.

9.2.1 Improvement of Data Collection Procedures

During the pre-experiment questionnaire, participants were asked their age, gender, and experience with mechatronics and programming. Excluded from this questionnaire were questions pertaining to extracurricular activities, education before university, and general life experiences that might be relevant to using mechatronic devices. Future studies would benefit from additional exploration of participant background to determine whether it affects the MWL of individuals using mechatronic devices.

Like the pre-experiment questionnaire, the interview conducted after the experiment could be improved to better elicit non-biased responses from participants. For example, instead of asking "how difficult was the task," the question could be "how simple was the task" to obtain a different perspective. Similarly, other questions within the interview could be made less leading by excluding biased words like "effectiveness". Additional questions could also be included to better understand NASA-TLX subsection variability by investigating individual interpretation of NASA-TLX questions. Alternatively, Likert methods (i.e., rating an answer very bad, bad, neutral, good, or very good) could be incorporated into the post-experiment data collection process to allow for mixed method data analysis. Furthermore, when analyzing participant response and comments, additional team members should be recruited to extract comments to avoid the bias discussed in 6.3.1 Attempting to Minimize Observer Bias.

9.2.2 Improvement and Modifications to Experimental Setup

The experimental apparatus used in this study consisted of an airplane model, three experimental modules, three servo motors, and the various cables needed to connect everything. During the experiment, it was found that there were several areas where improvements could be made.

During the experiment, an assumption was made by several participants that the devices were colour coded based on the toy bricks used in module casing, servo casings, and the model itself. This led to confusion: individuals who assumed colour coding was present tried to extract meaning from the brick colour when there was no meaning to be had. If toy bricks are used in the construction of future experimental apparatuses, then the brick colour chosen should be uniform. Alternatively, more detailed models and casing could be constructed using 3D printing approaches which would eliminate colour coding confusion and look more professional.

Another area that could be improved is the experimental module hardware. Participants found that while cables could easily be attached to the experimental module ports, detaching the modules was much more difficult. The connectors chosen during the design phase of the project were intended to be difficult to detach during operation to ensure that communication and power were not interrupted. However, participants needed to detach the connectors from the ports during the experiment when they believed they made a mistake, and this was made difficult by the connectors. For future iterations of this experiment, alternative ports that are easier to disconnect should be used.

This research could also be improved in future iterations by using modified or more complex version of the model to which participants installed the mechatronic system. More advanced versions could even go so far as to use working airplane models to apply components and systems to. At the minimum, future iterations should use models that more closely resemble the situational application, since some individuals in this experiment had trouble recognizing the model as representative of an airplane.

9.2.3 Improvement of Documentation and Instruction

The instruction sets were intended to be confusing to a degree; however, in the future they should be rewritten to incorporate changes to the experiment and to remove some areas of confusion. The problem with attempting to intentionally create opportunities for error is that it introduces harder to control variables into the experiment. This, in turn, can cause the results to be convoluted because of the additional uncertainty. Therefore, future iterations of this research should create instruction sets that are more focused onto one or two variables instead of trying for general vagueness.

The supplementary figures introduced into this experiment were widely disliked by participants mainly due to the lack of consistency and clarity. Using the recommendations made during this research, future iterations should ensure that figures used have a clear purpose for inclusion (such as indicating orientation) and clearly represent what they are attempting to represent (such as using pictures of modules instead of rectangular prisms designed to represent them).

The parts list included in the experiment was widely liked by participants, therefore, any future iterations of this research should include a part list in the experimental setup. In this study, several components attached to the model were left out of the part list, causing confusion. Future research should ensure that all components described in other portions of the documentation be included to allow participants to have a visual reference.

9.2.4 General Recommendations for Future Research

The primary goal of future research stemming from this study should be to further explore and validate the conclusions made here. The objectives of this study were to identify issues novices encounter with mechatronic devices (*obj. 1*) and to provide recommendations on how to design mechatronic devices with novice users in mind (*obj. 2*). The most significant outcome from these objectives were five conceptual areas where deficiencies occur and improvements can be made (i.e., simplicity, clarity, transparency, consistency, and context). Future research in this area should explore the applicability of those five concepts and weight them in terms of importance to novice users within the fields of mechatronics and mechanical engineering. Furthermore, if determined to be applicable, future research should consider developing a general system design approach that incorporates those five concepts.

Future research could take many forms; however, this study recommends that it should be conducted in an educational environment, specifically in university classrooms. A similar experiment to what was conducted in this research could be applied within mechatronics classrooms (i.e., where novice users learn to use mechatronic devices) to help validate and explore the results obtained from this study. The classroom environment would be advantageous since it would provide larger sample sizes and better demographic distributions. Furthermore, lessons learned from this study could directly be applied to improve the novice user experience with these devices. The mechatronic system developed during this research could also readily be used, reducing the requirement for experiment redesign.

Alternatively, the experimental protocol could be modified to reduce variability using several methods. The first would be using a paired experiment where participants conduct two separate tasks that are similar but different, such as performing the task used in this experiment using the control instruction set and then the experimental instruction set. The difficulty would lie in eliminating the learning effect, which might only be accomplished through the creation of two distinct tasks. Making those tasks relevant to the study would be difficult but would be an excellent area for further research. The second way to improve the study and hopefully remove variability would be to focus the task used in the experiment. This study attempted to balance the complexity of the instruction sets with the complexity of the task, but further research could adjust those scales to focus either on the difficulty of the instruction set or difficulty of the task. For example, the task itself could be made extremely simple but the instruction set could be made difficult to use. Alternatively, the instruction set for a complex task could be created through multiple iterations to be as simple as possible. The challenge inherent to this modification would be adjusting the difficulty of the instruction set with the difficulty of the task itself, which would be a challenge because they directly correlate. However, if focus was achieved on one of the two, it might eliminate some of the variability found during this study.

The other topic that would need to be considered is the effect of feedback on MWL. This study determined statistically that MWL is not reduced by real-time feedback, however, there was doubt as to the validity of this conclusion. If possible, this topic should be explored using a larger study population and a modified experimental procedure designed specifically to explore the effects of feedback on MWL. This could be incorporated directly into the classroom setting mentioned previously.

References

[1] D. A. Hooper, “Are we Designing Products for Experts or Just Really Good Beginners?,” *Procedia Manufacturing*, vol. 3, pp. 166–72, 2015, doi: 10.1016/j.promfg.2015.07.122.

[2] J. S. PhD, “Do Novices or Experts Uncover More Usability Issues? – MeasuringU.” Accessed: Sep. 12, 2023. [Online]. Available: https://measuringu.com/novice-expert-issues/

[3] J. Rasmussen, “Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 3, pp. 257–266, May 1983, doi: 10.1109/TSMC.1983.6313160.

[4] P. Denny *et al.*, “On designing programming error messages for novices: Readability and its constituent factors,” in *CHI ’21: CHI Conference on Human Factors in Computing Systems, 8-13 May 2021*, in CHI ’21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, May 2021, p. 55 (15 pp.). doi: 10.1145/3411764.3445696.

[5] N. A. Stanton, *Human factors methods: a practical guide for engineering and design*. Burlington, VT: Ashgate Publishing Company, 2013.

[6] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” in *Advances in Psychology*, vol. 52, Elsevier, 1988, pp. 139–183. doi: 10.1016/S0166-4115(08)62386-9.

[7] S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006, doi: 10.1177/154193120605000909.

[8] “TLX @ NASA Ames - NASA TLX Paper/Pencil Version.” Accessed: Apr. 17, 2023. [Online]. Available: https://humansystems.arc.nasa.gov/groups/tlx/tlxpaperpencil.php

[9] S. Said *et al.*, *Using the Raw NASA Task Load Index Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: A Meta-Analysis Using a Mixed Model. (Preprint)*. 2020. doi: 10.2196/preprints.19472.

[10] N. Al Madi, S. Peng, and T. Rogers, “AssessingWorkload Perception in Introductory Computer Science Projects using NASA-TLX,” in *53rd Annual ACM Technical Symposium on Computer Science Education, SIGCSE 2022, March 3, 2022 - March 5, 2022*, in SIGCSE 2022 - Proceedings of the 53rd ACM Technical Symposium on Computer Science Education, vol. 1. Virtual, Online, United states: Association for Computing Machinery, Inc, 2022, pp. 668–674. doi: 10.1145/3478431.3499406.

[11] B. J. Kissack, “EXPLORATION OF A LOW-COST AUTOPILOT SYSTEM FOR USE IN ACADEME,” p. 91.

[12] M. Coombes, O. McAree, W.-H. Chen, and P. Render, “Development of an autopilot system for rapid prototyping of high level control algorithms,” in *Proceedings of 2012 UKACC International Conference on Control*, Sep. 2012, pp. 292–297. doi: 10.1109/CONTROL.2012.6334645.

[13] ArduPilot, “ArduPilot,” ArduPilot.org. Accessed: Nov. 24, 2022. [Online]. Available: https://ardupilot.org

[14] M. Wojciechowski, M. Swanepoel, B. Troub, R. Garrido, J. Magnin, and C. Montalvo, “Low-Cost, Multi-Purpose Autopilot for Ground and Aerial Vehicles using an Arduino MEGA with Transistor Array Safety Circuit,” in *AIAA Flight Testing Conference, 5-9 June 2017*, in AIAA Flight Testing Conference. Reston, VA, USA: AIAA - American Institute of Aeronautics and Astronautics, 2017, p. 25 pp.

[15] J. L. Siewert, M. Wolf, B. Böhm, and S. Thienhaus, “Usability Study for an Augmented Reality Content Management System,” in *Cross Reality and Data Science in Engineering*, M. E. Auer and D. May, Eds., in Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2021, pp. 274–287. doi: 10.1007/978-3-030-52575-0_22.

[16] J. C. L. Lau and K. Mombaur, “Preliminary Study on a Novel Protocol for Improving Familiarity with a Lower-Limb Robotic Exoskeleton in Able-Bodied, First-Time Users,” *Frontiers in Robotics and AI*, vol. 8, 2022, Accessed: Dec. 22, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2021.785251

[17] M. L. Bolton, E. Biltekoff, and L. Humphrey, “The Mathematical Meaninglessness of the NASA Task Load Index: A Level of Measurement Analysis,” *IEEE Transactions on Human-Machine Systems*, pp. 590–9, 2023, doi: 10.1109/THMS.2023.3263482.

[18] S. S. Stevens, “On the Theory of Scales of Measurement,” *Science*, vol. 103, no. 2684, pp. 677–680, Jun. 1946, doi: 10.1126/science.103.2684.677.

[19] “Mechanical and Materials Engineering < Queen’s University.” Accessed: Jan. 31, 2024. [Online]. Available: https://www.queensu.ca/academic-calendar/engineering-applied-sciences/academic-plans/mechanical-engineering/#coursestext

[20] “Mechatronics Robotics Engineering, B.A.Sc. (Class of 2025) < Queen’s University.” Accessed: Jan. 31, 2024. [Online]. Available: https://www.queensu.ca/academic-calendar/engineering-applied-sciences/academic-plans/mechatronics-robotics-engineering/mechatronics-robotics-engineering-class-of-2025/

[21] “How do you choose between SPI, I2C, and UART for embedded communication?” Accessed: May 31, 2023. [Online]. Available: https://www.linkedin.com/advice/0/how-do-you-choose-between-spi-i2c-uart-embedded

[22] “EEPROM Library | Arduino Documentation.” Accessed: Jan. 23, 2024. [Online]. Available: https://docs.arduino.cc/learn/built-in-libraries/eeprom/

[23] “FAQ How is effect size used in power analysis?” Accessed: Dec. 11, 2023. [Online]. Available: https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/effect-size-power/faqhow-is-effect-size-used-in-power-analysis/

Appendix A: PCB Schematics and Drawings

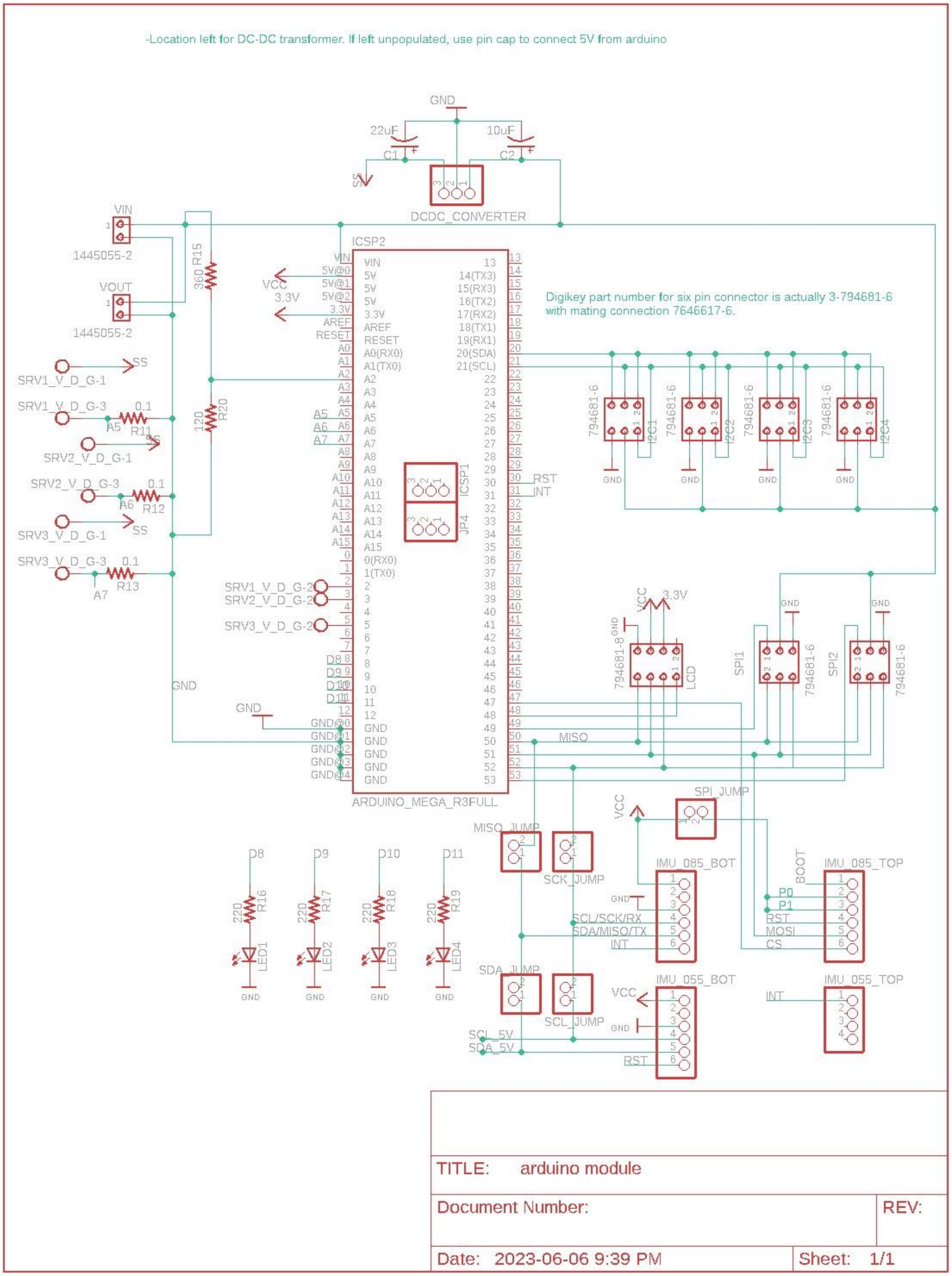


Figure 15: Arduino PCB schematic.

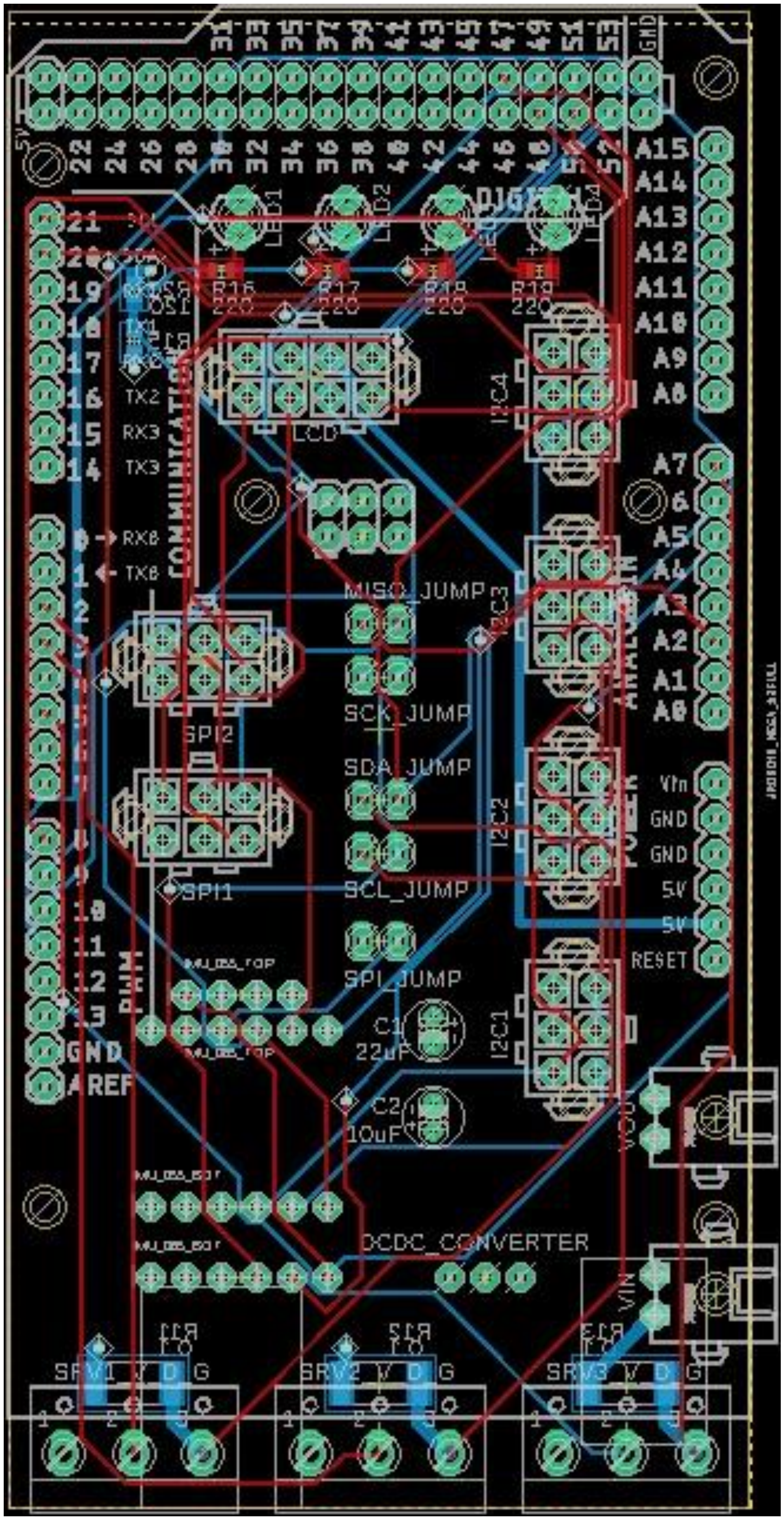


Figure 16: Arduino PCB diagram.

Appendix B: Combined SHERPA Output

Table 12: Combined SHERPA output from Figure 3 and Figure 4. Empty rows indicate error mode was insignificant.

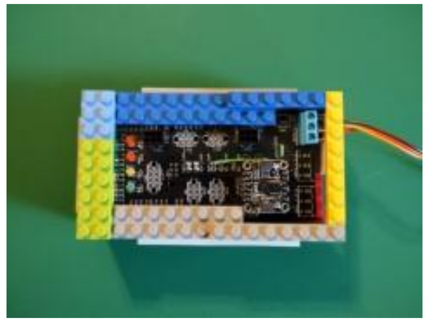
Task Step	Error Mode	Error Description	Consequence	Recovery	P	C	Remedial Strategy
A(B)1. Prepare for Experiment							
Plan A1: Do 1 to 4 in order							
Plan B1: Do 1 to 5 in order							
A(B)1.1 open box							
A(B)1.2 remove contents							
A(B)1.3 Confirm all components are present	C1	Failed to check all components are present.	Missing Components.	Components replaced upon discovery.	L	L	Administrator checks box between participants for missing components.
A(B)1.4 Differentiate between the power supply cable and connection cables	R2	Failed to differentiate between the two cable types.	Uses wrong cable.	A4	L	L	Provide a part list with images.
B1.5 Turn on Power Supply	A8	Failed to turn on power supply.	Modules are unpowered	B3.5	L	L	Lack of LED feedback in step B3.5 will alert user that power is off.
A(B)2. Install servo motors							
Plan A(B)2: do 1, then 2 and 3 in any order, then 4, and then 5.							
A(B)2.1 Review instruction sheet	I2	Instructions misread.	Incorrectly installed component	Step 5.3 and 5.4	M	M	Iterate instructions using external feedback to ensure understanding.
A(B)2.2 determine correct installation location	A5	Incorrect servo location.	Misaligned servo motor and damage to model	None	L	H	Design model so that there can only be one viable position for each servo.
A(B)2.3 determine correct installation orientation	A5	Incorrect servo orientation.	Misaligned servo motor and damage to model	None	L	H	Design model so that there can only be one viable position for each servo
A(B)2.4 Ensure the neutral position of the servo aligns with that of the flight surface	A5	Misaligned servo.	Flight surface will constantly be in wrong position	Step 5.6	M	M	Provide indicators for neutral position on flight surfaces and servo.
A(B)2.5 Attach servo to flight surface attachment point and firmly affix in place	A4	Servo not firmly placed.	Movement during testing.	Immediate	L	L	Administrator to ensure that each module can be easily and firmly fastened to the model.
A(B)3. Install Arduino modules							
Plan A3: Do 1 to 4 in order							
Plan B3: Do 1 to 7 in order							
A(B)3.1 review instruction sheet, prepare to install the next uninstalled servo	I2	Instructions misread.	Incorrectly installed component.	Step 5.3 and 5.4	M	M	Iterate instructions using external feedback.
A(B)3.2 determine correct installation location	A7	Incorrect location.	Wiring won't reach.	Relocate during step A4/B3.3.	M	L	Error mode unaddressed, desired

							user problem solving area.
B3.3 Connect appropriate cable to the appropriate attachment point, referring to instructions	A6	Incorrect attachment point used.	Devices won't communicate.	Immediate	L	H	Devices designed so that attachment point doesn't matter.
A3.3/B3.4 determine correct installation orientation	A7	Incorrect orientation.	Wrong module address.	Reorientate during step 5.4 or 5.5	M	H	Error indicator is integrated into code to communicate when this occurs.
B3.5 Check indicator LEDs and confirm correct orientation, referring to instruction set							
B3.6 If red indicator LED is on, go to step 3.3							
A3.4/B3.7 firmly affix module in place	A4	Module not firmly placed.	Movement during testing.	Immediate	L	L	Administrator to ensure that each module can be easily and firmly fastened to the model.
A4 Connect devices via the appropriate cables							
Plan A4: do 1 to 3 in any order							
A4.1 Connect the roll module to pitch module	A4	Cable isn't fully connected.	No communication between modules.	Step 5.3	L	M	Ensure that cables can easily connected but not easily disconnected so that participants will be able to fully connect cables while not disconnecting them during operation.
A4.2 connect pitch module to yaw module	A4	Cable isn't fully connected.	No communication between modules.	Step 5.3	L	M	Refer to step A4.1.
A4.3 Connect roll module to power supply in off position	A4	Cable isn't fully connected	No power to modules	Step 5.2	L	M	
A5 Test system							
Plan A5: Do 1 to 7 in order							
Plan B4: Do 1 to 3 in order							
A5.1/B4.1 Review instruction sheet for test information	R2	Instructions misread.	User is unsure how to test if modules are operational.	Immediate	L	L	Researcher to add handles model base to ensure that testing is easy and obvious for the user.
A5.2 Turn on power supply	A8	Failed to turn on power.	Model won't react to movement.	Immediate	L	L	Lack of LED indicators will alert the user to the lack of power.
A5.3 Check that all modules display the correct light	C2	Failed to check all modules.	Users moves on with incorrectly installed module.	Step 5.4	M	L	Users are instructed to check the red LED indicator in the next step to determine whether this failure mode is present.
A5.4 If red light is on, go back to step A3.1 and address the error with the power on	R2	User has misinterpreted the error LED.	Addresses the wrong issue.	Step 3.1	M	L	Devices are coded to not move into operation mode until there are no red LED indicators present.

A5.5 If red light is on after setup, turn power off, address error and go to step A3.1	R2	User believes that device is still in setup mode.	Two modules are oriented the same way.	A5.6	M	L	All LEDs are coded to flash between setup and operation to signal that setup has ceased. Flight surfaces will not move if there is a red LED indicator present.
A5.6/B4.2 Check that flight surfaces are neutral when model is flat and move when model moves	C1	User failed to check that all flight surfaces are moving.	User will be unaware if flight surfaces are misaligned or not moving.	No recovery.	L	L	This step is for the researcher to examine whether the user setup the device properly. Experiment is designed so that users shouldn't need to go back if they reach this step.
A5.7/B4.3 Verify flight surfaces move to counteract model movement	C2	User failed to check that flight surfaces move to counteract movement of the model.	User will be unaware if flight surfaces are misaligned or not moving.	No recovery.	L	L	Refer to step A5.6/B4.2.

Appendix C: Part List

Part List

Control Module		x3
Servo Motor		x3
Connection Cables		x2
Power Cable		x1

Appendix D: Software Transcript

```
//Startup
#include <Wire.h>
#include <Adafruit_Sensor.h>
#include <Adafruit_BNO08x.h>
#include <Servo.h>

Servo servo;
int address = 0;

//defining of structures used to send imu data as a packet
struct imuData_t {
  float yaw;
  float pitch;
  float roll;
  int address;
};
#define PACKET_SIZE sizeof(imuData_t)
union I2C_Packet_t{
 imuData_t sensor;
 byte I2CPacket[sizeof(imuData_t)];
};
I2C_Packet_t imu_data;
//end definition

//define global variables
//quaternion variables
float theta = 0;
float q0=0;
float q1=0;
float q2=0;
float q3=0;
//gravity calculation variables
float gx=0;
float gy=0;
float gz=0;
//misc calculation variables
int servoang = 0;
bool flag1;
bool flag2;
//define pinouts
int servo1 = 2;
int servo2 = 3;
```

```
int servo3 = 5;
int iservo1 = A5;
int iservo2 = A6;
int iservo3 = A7;
int green = 8; //LED1
int yellow = 9; //LED2
int orange = 10; //LED3
int red = 11;  //LED4
int CSLCD = 48;
int CSSPI_1 = 49;
int CSSPI_2 = 53;
int deviceList[128];
int deviceAddress[128];
int Hys = 3;

//define specific pins for IMU
#define BNO08X_RESET 30
#define BNO08X_INT 31
#define BNO08X_CS 47

Adafruit_BNO08x  bno08x(BNO08X_RESET);
sh2_SensorValue_t sensorValue;

//Definition of IMU outputs
void setReports(void) {
  Serial.println("Setting desired reports");
  if (!bno08x.enableReport(SH2_ROTATION_VECTOR)) {
    Serial.println("Could not enable rotation vector");
  }
}

void setup() {
  //initialize digital pins as outputs or inputs
  pinMode(green,OUTPUT);
  pinMode(yellow,OUTPUT);
  pinMode(orange,OUTPUT);
  pinMode(red,OUTPUT);

  //begin serial
  Serial.begin(115200);
  // pause until serial console opens
  while (!Serial) delay(10);

  Serial.println("Adafruit BNO08x test!");
  if (!bno08x.begin_SPI(BNO08X_CS, BNO08X_INT)) {
```

```
    Serial.println("Failed to find BNO08x chip");
    while (1) { delay(10); }
  }
  Serial.println("BNO08x Found!");
  setReports();
  Serial.println("Reading events");
  if (bno08x.wasReset()) {
    Serial.print("sensor was reset ");
    setReports();
  }

  while (!bno08x.getSensorEvent(&sensorValue)) {
  }
  //local variables used for while loop
  int adrcheck;
  int counter = 0;
  //while loop determines device address, device needs to stay still for 10
seconds to
  //move forward in the code (obtain the same non-zero address 20 times)
  while(true){
    //need this code in every loop to ensure that the IMU is being read
    if (bno08x.wasReset()) {
      Serial.print("sensor was reset ");
      setReports();
    }
        if (!bno08x.getSensorEvent(&sensorValue)) {
      return;
    }
    q0 = sensorValue.un.rotationVector.real;
    q1 = sensorValue.un.rotationVector.i;
    q2 = sensorValue.un.rotationVector.j;
    q3 = sensorValue.un.rotationVector.k;
    gx = 2 * (q1*q3 - q0*q2);
    gy = 2 * (q0*q1 + q2*q3);
    gz = q0*q0 - q1*q1 - q2*q2 + q3*q3;
    Serial.print(" gx ");
    Serial.print(gx);
    Serial.print(" gy ");
    Serial.print(gy);
    Serial.print(" gz ");
    Serial.println(gz);

    //Positive gravity in the z direction indicates roll/head module (green LED)
    if(gz>0.75){
      address = 1;
```

```
      digitalWrite(green,HIGH);
    //negative gravity in the z direction indicates correct plane but upside down
(red and green LED)
    }else if(gz<-0.75){
      address = 0;
      digitalWrite(green,HIGH);
      digitalWrite(red,HIGH);
    //positive or negative gravity (module on its side) indicates pitch module
(yellow LED)
    }else if(gx>0.75){
      address = 2;
      digitalWrite(yellow,HIGH);
    }else if(gx<-0.75){
      address = 2;
      digitalWrite(yellow,HIGH);
    //negative gravity in the y direction indicates rudder control (orange LED)
    }else if(gy<-0.75){
      address = 3;
      digitalWrite(orange,HIGH);
    //positive gravity in the y direction indicates correct plane but upside down
(red and orange LED)
    }else if(gy>0.75){
      address = 0;
      digitalWrite(orange,HIGH);
      digitalWrite(red,HIGH);
    //all other initial positions give an error (red LED)
    }else{
      address = 0;
      digitalWrite(red,HIGH);
    }
    //check if address is same as the last (is the device steady)
    if(address > 0 && adrcheck == address){
      counter = counter + 1;
    }else{
      counter = 0;
    }
    adrcheck = address;
    Serial.println(counter);
    delay(250);
    //turn off LED (causes a blinking effect)
    digitalWrite(green,LOW);
    digitalWrite(yellow,LOW);
    digitalWrite(orange,LOW);
    digitalWrite(red,LOW);
    delay(250);
```

```
    //code continues after obtaining the same address 20 times
    if(counter == 20){
      break;
    }
  }
  Serial.println(address);
  //blink all LEDS to indicate address has been set
  digitalWrite(green,HIGH);
  digitalWrite(yellow,HIGH);
  digitalWrite(orange,HIGH);
  digitalWrite(red,HIGH);
  delay(1000);
  digitalWrite(green,LOW);
  digitalWrite(yellow,LOW);
  digitalWrite(orange,LOW);
  digitalWrite(red,LOW);
  delay(1000);
  //begin I2C bus with set address
  Wire.begin(address);
  Wire.onReceive(receiveEvent);
  //gives error until both yaw and pitch control communicate
  flag1 =0;
  flag2 = 0;
  if(address == 1){
    digitalWrite(green,HIGH);
    while(true){

      if(imu_data.sensor.address == 2){
        flag1 = 1;
      }
      if(imu_data.sensor.address == 3){
        flag2 = 1;
      }
      Serial.print("flag 1: ");
      Serial.print(flag1);
      Serial.print("  flag 2: ");
      Serial.println(flag2);
      if(flag1 ==1 && flag2 ==1){
        digitalWrite(red,LOW);
        Serial.println("Yaw and pitch controllers connected.");
        break;
      }else{
        digitalWrite(red,HIGH);
      }
    }
```

```
  }
}

void loop() {
  delay(10);
  if (bno08x.wasReset()) {
    Serial.print("sensor was reset ");
    setReports();
  }

  if (!bno08x.getSensorEvent(&sensorValue)) {
    return;
  }

//switch case changes what code is run based on address set in setup
switch (address){
  case 1:
    servo.attach(servo3);
    Serial.print("Head module and aileron control. ");
    digitalWrite(green,HIGH);
    digitalWrite(yellow,LOW);
    digitalWrite(orange,LOW);
    digitalWrite(red,LOW);

    //calculate yaw, pitch, and roll from rotation quaternion
    q0 = sensorValue.un.rotationVector.real;
    q1 = sensorValue.un.rotationVector.i;
    q2 = sensorValue.un.rotationVector.j;
    q3 = sensorValue.un.rotationVector.k;
    imu_data.sensor.pitch = atan2(2.0*(q2*q3 + q0*q1), q0*q0 - q1*q1 - q2*q2 +
q3*q3); //0 to 1.5, go down. 0 to -1.5 go up
    imu_data.sensor.roll = asin(-2.0*(q1*q3 - q0*q2)); //0 to 1.5, roll to left.
0 to -1.5 roll right
    imu_data.sensor.yaw = atan2(2.0*(q1*q2 + q0*q3), q0*q0 + q1*q1 - q2*q2 -
q3*q3); //0 to 3.1, turn right. 0 to -3.1 turn left
    Serial.print(" yaw ");
    Serial.print(imu_data.sensor.yaw);
    Serial.print(" pitch ");
    Serial.print(imu_data.sensor.pitch);
    Serial.print(" roll ");
    Serial.print(imu_data.sensor.roll);
    //send yaw/pitch/roll data to peripherals
    i2csend(imu_data,2);
    i2csend(imu_data,3);
```

```
    //equation of a line with rise being servo range, run being roll range, and
90 being neutral point of servo.
    //opposes change in orientation
    servoang = ((180)/(1.5+1.5))*imu_data.sensor.roll*2+90;
    Serial.print(" servoang ");
    Serial.println(servoang);
    servo.write(servoang);
    delay(300);
    break;

  case 2:
    servo.attach(servo3);
    Serial.println("pitch control");
    digitalWrite(green,LOW);
    digitalWrite(yellow,HIGH);
    digitalWrite(orange,LOW);
    digitalWrite(red,LOW);
    imu_data.sensor.address = address;
    i2csend(imu_data,1);
    //equation of a line with rise being servo range, run being pitch range, and
90 being neutral point of servo
    //opposes change in orientation
    servoang = ((180)/(1.5+1.5))*imu_data.sensor.pitch*2+90;
    //hysteresis loop to prevent servo burnout
    servo.write(servoang);
    delay(300);
  break;

  case 3:
    servo.attach(servo3);
    Serial.println("yaw control");
    digitalWrite(green,LOW);
    digitalWrite(yellow,LOW);
    digitalWrite(orange,HIGH);
    digitalWrite(red,LOW);
    imu_data.sensor.address = address;
    i2csend(imu_data,1);
    //equation of a line with rise being servo range, run being pitch range, and
90 being neutral point of servo
    //opposes change in orientation
    servoang = ((180)/(3.1+3.1))*imu_data.sensor.yaw*2+90;
    //hysteresis loop to prevent servo burnout
    servo.write(servoang);
    delay(300);
  break;
```

```
}
}
//interrupt function for I2C communication
void receiveEvent(int howMany)
{
  byte byteArray[PACKET_SIZE];
  int i=0;
  while(Wire.available()){
    imu_data.I2CPacket[i] = Wire.read();
    i++;
  }
}

//function for sending packets of information over I2C
void i2csend(I2C_Packet_t x, int ID)
{
  Wire.beginTransmission(ID); // transmit to device ID
  Wire.write(x.I2CPacket, PACKET_SIZE);
  Wire.endTransmission();     // stop transmitting
}
```

Appendix E: Participant Instruction Sets

Set A: Control Set (Powered After Setup)

1. Remove components from the box and confirm all parts are present (refer to part list).
2. Check the power supply and confirm that it is off. Power during setup will damage components.
3. To connect the roll servo to the flight surface attachment point at the front of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.
4. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel to the ground).
5. Connect the servo, ensuring that the connector can safely turn (Figure 2).
6. Connect the first module, the roll module (Figure 3), to the aircraft power supply using the power supply cable and install in orientation A (Figure 4) at the front of the aircraft (Figure 1).
7. Firmly affix module to aircraft.
8. To connect the pitch servo to the flight surface attachment at the rear of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.
9. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel to the ground).
10. Connect the servo, ensuring that the connector can safely turn (Figure 2).
11. Connect the second module, the pitch module (Figure 3), to the roll module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1).
12. Firmly affix module to aircraft.

13. To connect yaw servo to the flight surface attachment point at the rear of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.
14. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel with the length of the aircraft, perpendicular to the ground).
15. Connect the servo, ensuring that the connector can safely turn (Figure 2).
16. Connect the third module, the yaw module (Figure 3), to the pitch module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1).
17. Firmly affix module to aircraft.
18. Turn power supply on and determine if each module completes setup correctly. If setup correctly, each module will blink an LED corresponding with its role for ten seconds before blinking all LEDs once, indicating its role has been set. Green identifies the roll module, yellow identifies the pitch module, and orange identifies the yaw module.
19. Check the roll module to determine whether the Red LED is still on after setup. If red LED is on, return to step 4 and check whether instructions have been followed properly.
20. Test whether the stabilizing autopilot is working by tilting the model in each of the primary axes (yaw, pitch, and roll) and determine whether the flight surfaces are moving to counteract the change in orientation.

Set B: Powered at Start of Setup

1. Remove components from the box and confirm all parts are present (refer to part list).
2. Turn on the power supply.
3. To connect the roll servo to the flight surface attachment point at the front of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.
4. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel to the ground).
5. Connect the servo, ensuring that the connector can safely turn (Figure 2).
6. Connect the first module, the roll module (Figure 3), to the aircraft power supply using the power supply cable and install in orientation A (Figure 4) at the front of the aircraft (Figure 1). If in the correct position, green LED will blink (Figure 5). After ten seconds, all four LEDs will flash. The module is now set as the roll module.
7. Firmly affix module to aircraft. Afterwards, the red LED will remain on until all other devices are connected correctly.
8. To connect the pitch servo to the flight surface attachment point at the rear of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.
9. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel to the ground).
10. Connect the servo, ensuring that the connector can safely turn (Figure 2).
11. Connect the second module, the pitch module (Figure 3), to the roll module and install in orientation B (Figure 4) near the center of the aircraft (Figure 1). If in the correct position, the

yellow LED will blink (Figure 5). After ten seconds, all four LEDs will flash. The module is now set as the pitch module.

12. Firmly affix module to aircraft.

13. To connect yaw servo to the flight surface attachment point at the rear of the aircraft (Figure 1), first ensure that the servo is in its neutral position as indicated by a gold mark.

14. Adjust the flight surface position using the cockpit controls until they are also in their neutral position (parallel with the length of the aircraft, perpendicular with ground).

15. Connect the servo, ensuring that the connector can safely turn (Figure 2).

16. Connect the third module, the yaw module (Figure 3), to the pitch module and install in orientation C (Figure 4) near the center of the aircraft (Figure 1). If in the correct position, the orange LED will blink (Figure 5). After ten seconds, all four LEDs will flash. This module is now set as the yaw module.

17. Firmly affix module to aircraft.

18. Check the roll module to determine whether the Red LED is still on after setup. If red LED is on, return to step 4 and check whether instructions have been followed properly.

19. Test whether the stabilizing autopilot is working by tilting the model in each of the primary axes (yaw, pitch, and roll) and determine whether the flight surfaces are moving to counteract the change in orientation.

Appendix F: Data Collection Methods

Pre-experiment Questionnaire

Complete this section before completing the experiment.

1. *Age and Gender:* ____

2. *Educational background (level and faculty if applicable):*

   ____

3. *Rate your experience with microcomputers and microcontrollers:*

   ____

4. *Briefly describe your programming experience:*

   ____

NASA-TLX Survey

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Post-experiment Interview

Before you are the instructions that you used during the experiment, I will be referring to them during this interview.

1. Describe the effectiveness of the instructions.
2. What aspects of the experiment aided or hindered your understanding of the instructions?
3. How would you describe the overall difficulty of the task presented to you?
4. Do you have any suggestions on improving the overall procedure?
5. Do you have any suggestions on improving the instructions?
6. In what ways do the illustrations and text complement each other?
7. How well do you believe you executed the instructions?

Appendix G: GREB Letter of Intent and Consent Form

<ins>**Letter of Information/Consent**</ins>

Improving Usability for Novices in the Design of Mechatronic Devices: A Study Using Novel Arduino Modules

DEPARTMENT OF MECHANICAL AND MATERIALS ENGINEERING

McLaughlin Hall, 130 Stuart Street
Queen's University
Kingston, Ontario, Canada K7L 3N6
Tel 613 533-2575
Fax 613 533-6489
http://me.queensu.ca/

Principal Investigator: Levi Huyer
Supervisors: Claire Davies and Shane Pinder

Introduction:

This study is being conducted within the department of Mechanical and Materials Engineering at Queens University, Kingston. This study has been reviewed for ethical compliance by the Queen's University General Research Ethics Board (GREB).

This study aims to provide recommendations to improve usability and understandability of mechatronic devices. As part of this process, devices based on Arduino microcontrollers were developed for future use as novice friendly modular aircraft electronics. You will be asked to answer a questionnaire with basic demographic information. You will then be asked to interact with these devices in an experiment simulating what it would be like to install them in an aircraft. After the experiment, the researcher will ask you questions regarding your experience in the experiment. Audio will be recorded during this interview using Zoom with video disabled and transcription enabled, so we will seek your consent to being recorded. No identifying information will be collected and only the transcript will be used for the study. Likewise, only the written transcript will be retained, and all audio information will be deleted within 7 days of your participation. The experiment and interview are expected to take a total of **30 minutes**. The risks include being potentially overwhelmed by the complexity of the task and anxiety linked to being audio recorded or taking surveys. There are no direct benefits to participants. Participation is voluntary and participants can decline to the participate in the research or any aspect of the research at any time without penalty.

Confidentiality:

All data taken will be confidential, and only the principal investigator and project supervisors will have access to data during and after collection. The Queen's General Research Ethics Board (GREB) may request access to study data and/or all other study materials used in this research to ensure that we (the research team) have or are meeting our ethical obligations in conducting this research.

Research Data:

Information that will collected includes age, gender, educational background (level and faculty), experience with microcontrollers/microcomputers, and programming experience. All data will be collected confidentially, de-identified, and retained for a minimum of five years per the Queen's University retention policy by the supervisors on the Queen's University encrypted OneDrive server. No identifying information will be stored. At any time during or within four weeks following the experiment, you may contact the researcher to request to have your information withdrawn by providing your reference number. De-identified research results will be published as part of the completion of the principal investigator's master's degree and potentially within journal or conference publications.

Question about this Study:

If you have any ethics concerns, please contact the General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or email chair.GREB@queensu.ca. If participants have any questions about the research study, they can contact the following:

Levi Huyer (Principal Investigator)
Department of Mechanical Engineering
McLauglin Hall, Room 319
Queen's University, 130 Stuart Street
Kingston, Ontario, Canada K7L 3N6

Claire Davies (Supervisor)
Department of Mechanical Engineering
McLaughlin Hall, Room 203
Queen's University, 130 Stuart Street
Kingston, Ontario, Canada, K7L 3N6
Email: claire.davies@queensu.ca
Phone: 613-533-6000 ext. 78743

Shane Pinder (Supervisor)
Department of Mechanical Engineering
Queen's University, 130 Stuart Street
Kingston, Ontario, Canada, K7L 3N6
Email: shane.pinder@queensu.ca

Consent:

By signing this form, the participant

- Consents to the use of their de-identified information for the purposes of this study.
- Acknowledges that all their questions have been answered.
- Has been provided a copy of this consent form and returned a copy to the researcher.

I Consent to audio recording (Initial) ____________

I Consent to use of quotes (Initial) ____________ You have not waived any legal rights by consenting to participate in this study. **Please retain this letter of information/consent in case you have questions or need more information about the study.**

Participant:

Name (Please Print):__________________ Signature:__________________ Date:__________________

Person Performing Informed Consent Discussion:

Name (Please Print):__________________ Signature:__________________ Date:__________________

Appendix H: GREB Approval

Queen's University General Research Ethics Board (GREB)

GREB Initial Ethics Clearance

October 02, 2023

Huyer
Department of Mechanical and Materials Engineering
Queen's University

TRAQ #: 6039557
Study Title: "Improving Usability for Novices in the Design of Mechatronic Devices: A Study Using Novel Arduino Modules"
Date Ethics Clearance Issued: October 02, 2023
Ethics Clearance Expiry Date: October 02, 2024

Dear Huyer:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled **"Improving Usability for Novices in the Design of Mechatronic Devices: A Study Using Novel Arduino Modules"** for ethical compliance with the Tri-Council Guidelines (TCPS 2) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405), your project has been cleared for one year.

Renewals: An annual renewal event form or a study closure event form must be submitted annually as per the TCPS 2 Article 6.14. As a courtesy, the Office of Research Ethics Compliance may send reminders 30 days in advance of the ethics clearance expiry date. All lapses in ethics clearance will be documented on the annual renewal clearance letter. A Suspension letter may be issued for lapses in ethics clearances, with subsequent termination and closure of the ethics file for lapses greater than 10 business days.

Completion/Termination: The GREB must be notified of the completion or termination of this study through the submission of a study closure event form in TRAQ. This should be submitted at the time of completion; there is no need to wait until the annual renewal due date.

Amendments: No deviation from, or changes to the protocol, informed consent form and conduct of study should be initiated without prior written clearance or an appropriate amendment event from the GREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the study. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/traq/signon.html; click on "Events;" under "Create New Event" click on "General Research Ethics Board Request for the Amendment of Approved Studies." Once submitted, these changes will automatically be sent to the Ethics Coordinator, GREB, at University Research Services for further review and clearance by GREB or the Chair, GREB.

Adverse Events: You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one-year period (access this form at http://www.queensu.ca/traq/signon.html/; click on "Events;" under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Sincerely,

Chair, General Research Ethics Board (GREB)
Associate Professor and Distinguished Faculty Fellow of Marketing
Master of Digital Product Management
Smith School of Business
Queen's University
chair.GREB@queensu.ca