Augmented vs. Traditional Instruction in Manufacturing Assembly: An Affordance-Based, Multi-Modal Assessment of Learning, Recall, and Retention

by

Danny J. O'Leary

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 3, 2024

Keywords:
Augmented Reality, Mixed Reality, Augmented Instruction, Learning & Retention Metrics
Augmented & Mixed Reality Assessment Methods, Manufacturing Assembly Training

Approved by

Dr. John Evans, Co-chair, Professor, Industrial & Systems Engineering
Dr. Gregory Harris, Co-chair, Professor & Chair, Industrial & Systems Engineering
Dr. Richard Sesek, Professor, Industrial & Systems Engineering
Dr. Konstantinos Mykoniatis, Assistant Professor, Industrial & Systems Engineering

Abstract

This research rigorously assesses the effectiveness of augmented and mixed reality (AR/MR) instructional techniques in manufacturing training. By investigating how augmented methods designed to leverage specific affordances impact operator learning, recall, and retention in an authentic assembly context, it addresses critical gaps in understanding the relative efficacy of these technologies.

A cohort of 54 participants without prior exposure to similar tasks or AR/MR technology underwent simulated assembly training at the Tiger Motors Lean Education Lab. This between-groups study introduced participants to one of four instructional methods: Paper Work Instructions (PWI), Projected AR (PAR), head-mounted AR, and head-mounted MR. The same task was used across treatments, with minimal changes to instructional design. MR was differentiated from AR by allowing participants to freely manipulate the workpiece while maintaining aligned augmentations.

Performance metrics including task duration and error rates / types were carefully analyzed to assess learning progression and task proficiency. Measures of instructional reliance were also captured during recall, better informing the learning efficacy of each method. Several weeks later, retention was assessed on a volunteer subset to compare the durability of learning by treatment. The NASA Task Load Index and System Usability Scale were also administered to assess perceived workload and user experience. These data, coupled with demographics and qualitative feedback from open-ended exit interviews, contributes to a comprehensive expression of each participant's experience.

Key findings revealed a nuanced trade-off between speed and accuracy across methods. Augmented technologies generally led to fewer errors but slower initial performance, while traditional methods allowed for faster execution but higher error rates. Recall and retention results were mixed, suggesting complex relationships between instruction method and long-term learning outcomes. Surprisingly, no significant differences in perceived workload were found across treatments, despite varying technological complexities.

This study's implications extend beyond manufacturing to analogous tasks involving installation, repair, and safety training. Through its rigorous experimental design, ecological validity, and systematic operationalization of affordances, this research provides actionable insights to optimize AR/MR technology implementation and learning outcomes in industrial settings. The affordance-based framework introduced here offers a novel approach for evaluating and designing AR/MR training systems across various domains.

Acknowledgements

Research is what I'm doing when I don't know what I'm doing.

I've lived this definition of research, commonly attributed to Wernher von Braun, on and off for the better part of four years now. That would not have been possible without the support and encouragement of many people, all of whom I greatly appreciate. There are several that deserve special acknowledgement at the beginning of the end of this journey.

First, my dissertation committee: Drs. John Evans, Gregory Harris, Richard Sesek, and Konstantinos Mykoniatis, and outside reader, Darren Olsen, JD. Particular thanks to Dr. Evans for first welcoming me into the ISE program, to Dr. Harris for further opportunities, and to Dr. Sesek for his shameless advocacy. Sincere thanks to all for your generous support.

Thanks also to the faculty, staff, and students of Industrial and Systems Engineering at Auburn University. I would not have done it without you. A special thanks go to my instructors in the analytics certificate, Drs. Jeff Smith, Jorge Valenzuela, and Alice Smith, who rekindled my fascination with programming and modeling, and introduced me many of the tools that made this possible.

My research would have been impossible without the support and assistance of the Tiger Motors Lean Education Center, its Director, Dr. Tom Devall, and staff, especially Diego Caputo. Thanks for sharing this amazing facility and for the great work it does.

I must also give thanks to my research assistants, Alex Barras, David "Brown" Teague, Carson Tillery, collaborators, Drs. Victoria Ballard and Md Monir Hossain, and protocol assistant, Kralyn Thomas. I'm forever grateful for your contributions.

Beyond the faculty, staff, and students of Auburn, its College of Engineering, and the Department of Industrial and Systems Engineering, I owe immense gratitude to my family.

To my parents, Mike and Sherry, for nurturing in me a spirit of overwhelming curiosity and prudent skepticism and for showing me what it means to be an engineer and educator. I can never repay your love, but do my best to pass it on each and every day. To my brother, Scott, for being my partner in resilience. Thank you for your strength and determination. To my eldest son, Patrick, daughter in-law Kirsten, grandson Teagan, and youngest son, Timothy. Thank you for putting up with an eccentric old man and sharing your lives with him. You teach me every day.

But most of all, to my wife, Cheryl. You are the best partner I could have ever hoped for. I cannot express the gratitude I have for your tireless support and encouragement. This particular part of our 38-year journey has challenged me in new ways, but it was likely harder for you in others. Thank you for being a part of who I am. *Always remember there was nothing worth sharing, like the love that let us share our name.*[1]

[1] The Avett Brothers, "Murder in the City"

Table of Contents

List of Tables

List of Figures

1 Introduction

Challenged to meet the demands of reduced cost, global competition, sustainability, shorter product life cycles, and product complexity, the manufacturing industry is in the midst of a digital transformation. Industry 4.0 and the underlying shift from mass production to mass customization places new demands on the workforce. Today's manufacturing operators must manage a wider range of responsibilities and increased information flow amidst decreasing margins for error, changing methods, and new technologies. These trends show no signs of abatement (Danielsson et al., 2020). Extended Reality (XR) devices, in their various forms, are expected to aid this problem through operator training and support.

Augmented Reality (AR) systems "combine real and virtual, are interactive in real time, and are registered in 3-D" (Azuma, 1997). By realistically integrating informative and/or interactive virtual objects in our view of the world, AR aims to enhance the users' interaction with and perception of it. Its essential affordance is the direct and natural manipulation of virtual objects in everyday surroundings. Relative to metaphorical digital interfaces, this is thought to improve the uptake of knowledge by reducing the overall cognitive load and better distributing it across multiple sensory pathways (Shelton & Hedley, 2003). AR-assisted learners demonstrate improved perception, performance, and understanding of spatial concepts, with outcomes correlated to the amount of physical engagement involved (Chen et al., 2019). As a result, AR is thought to be well-suited for task-related learning. Using untethered, hands-free devices with optical see-through head-mounted displays, AR can continuously enhance the user's actions in the real world (Leonard & Fitzgerald, 2018). These benefits have broad industrial applications.

In manufacturing, operator support has been a common application of AR research and development since the early 1990s (Azuma & Bishop, 1994). It is also seen as a source of innovative operator training methods required to meet rapidly increasing demand for skilled labor due to high retirement rates, global expansion, and increasing specialization (Kress, 2020). Manufacturing support, training, and related applications have been identified in the areas of assembly, maintenance, operations, quality control, safety, design, visualization, logistics, and marketing (Oztemel & Gursev, 2020).

Despite great potential, the adoption of AR is slowed by technical, market, and other important social and legal obstacles (Azuma, 2019). XR technologies remain relatively immature and will face new challenges as their development moves from research labs to the shop floor. There, priorities shift from building technology to delivering solutions. No longer "proofs of concept," these systems will be evaluated on the basis of their return on investment and other key performance indicators essential to the business case (Masood & Egger, 2019, 2020). That performance will be assessed in the context of the entire operation, where technical considerations are balanced by organizational and environmental factors. The long-term success of XR initiatives ultimately rests on how those results compare with other investment alternatives.

But AR remains a highly fragmented market, including a diverse selection of screen-based, projected, and head-mounted technologies (Kress, 2020). Studies show that the efficacy of these systems varies with the task type, technology used, application design, and other factors (Kaplan et al., 2021). Research in this area is young but accelerating. Most of it focuses on efficiency (task time) and accuracy (error count). These are relevant but incomplete measures for assessing training outcomes, where the learning rate and transfer effectiveness must also be considered (Büttner et al., 2020).

This dissertation addresses critical gaps in understanding the relative effectiveness of different AR/MR technologies for enhancing learning outcomes in manufacturing assembly training. Specifically, it explores how diverse augmented instructional methods, designed to leverage specific affordances, impact operator learning, recall, and retention in a simulated manufacturing context. The research employs an innovative affordance-based framework to evaluate and compare these technologies, providing a new theoretical lens for understanding their effectiveness.

The primary research objectives are to compare the effectiveness of various augmented instruction with traditional methods, evaluate the effect of those methods on the quality and durability of learning, investigate the role of specific affordances in those results, and explore how perceived workload, usability, and performance are related.

This study employs a rigorous, multi-phase experimental design conducted in the Tiger Motors Lean Education Lab, an environment designed to simulate modern automotive manufacturing. By combining quantitative performance metrics with qualitative user feedback, the research provides a comprehensive assessment of each instructional method's effectiveness.

The findings of this study contribute to both theory and practice in the field of AR/MR-assisted training. Theoretically, it advances our understanding of how technological affordances translate into learning outcomes in practical settings. Methodologically, it demonstrates the value of a comprehensive, multi-phase assessment approach in capturing the full impact of these technologies on skill acquisition and retention. Practically, it offers insights to guide the implementation of AR/MR training systems in manufacturing contexts, helping to optimize both immediate performance and long-term skill development.

The remainder of this dissertation is structured as follows:

- Chapter 2 provides a comprehensive literature review, situating this research within the broader context of AR/MR applications in manufacturing and relevant learning theories.
- Chapter 3 articulates the specific problem statement, research questions, and hypotheses that guide this study.
- Chapter 4 details the experimental methods, including participant recruitment, data collection procedures, and analytical approaches.
- Chapter 5 presents the results of the study, organized by research question and hypothesis.
- Chapter 6 discusses the implications of these findings, acknowledges limitations, and suggests directions for future research.

This work aims to advance the understanding of AR/MR technologies in manufacturing training, ultimately contributing to more effective workforce development in an increasingly complex industrial landscape.

2 Literature Review

2.1 Review Methodology

This review employs a hybrid approach, combining traditional systematic approaches (Kitchenham, 2004) with an emerging class of modern tools. Systematic methods were used to identify interesting references for each search, based on relevance, prominence (citation count), and debate (supporting and contrasting citations). This phase of the search primarily leveraged meta-databases including Web of Science,[1] Scopus,[2] Semantic Scholar,[3] and Google Scholar.[4] The specific search parameters and criteria for inclusion varied with each use.

The resulting set of publications was used to seed a secondary search using a combination of graph and AI-based tools, including scite_,[5] Inciteful,[6] ResearchRabbit,[7] Connected Papers,[8] and Litmaps.[9] At the time of this writing, this category of tools was experiencing rapid growth and change. No single "best" tool or approach had yet emerged, but their collective benefits provided a valuable complement to the systematic approach. The tools and methodology described here were influenced by the work of Mushtaq Bilal (2023) and Ilya Shabanov (2024).

Broadly speaking, these tools link papers based on citation trees, bibliographic coupling, analysis of the citation statement, and other sophisticated methods. From their original findings, users can interactively traverse connected papers in graph and/or timeline view, focus on specific authors or collaborators, and otherwise refine the search. Abstracts and links to the papers are available throughout the process to guide exploration.

Integrating traditional and modern approaches in this iterative and exploratory fashion teased out unexpected connections, incorporated a wider range of sources, and facilitated the author's understanding of relevant discourse across multiple dimensions, including time, application context, and research domain. This iterative process was repeated for

[1] Web of Science: https://www.webofscience.com/
[2] Scopus: https://www.scopus.com/
[3] Semantic Scholar: https://www.semanticscholar.org/
[4] Google Scholar: https://scholar.google.com/
[5] scite_: https://scite.ai/
[6] Inciteful: https://inciteful.xyz/
[7] ResearchRabbit: https://www.researchrabbit.ai/
[8] Connected Papers: https://www.connectedpapers.com/
[9] Litmaps: https://www.litmaps.com/

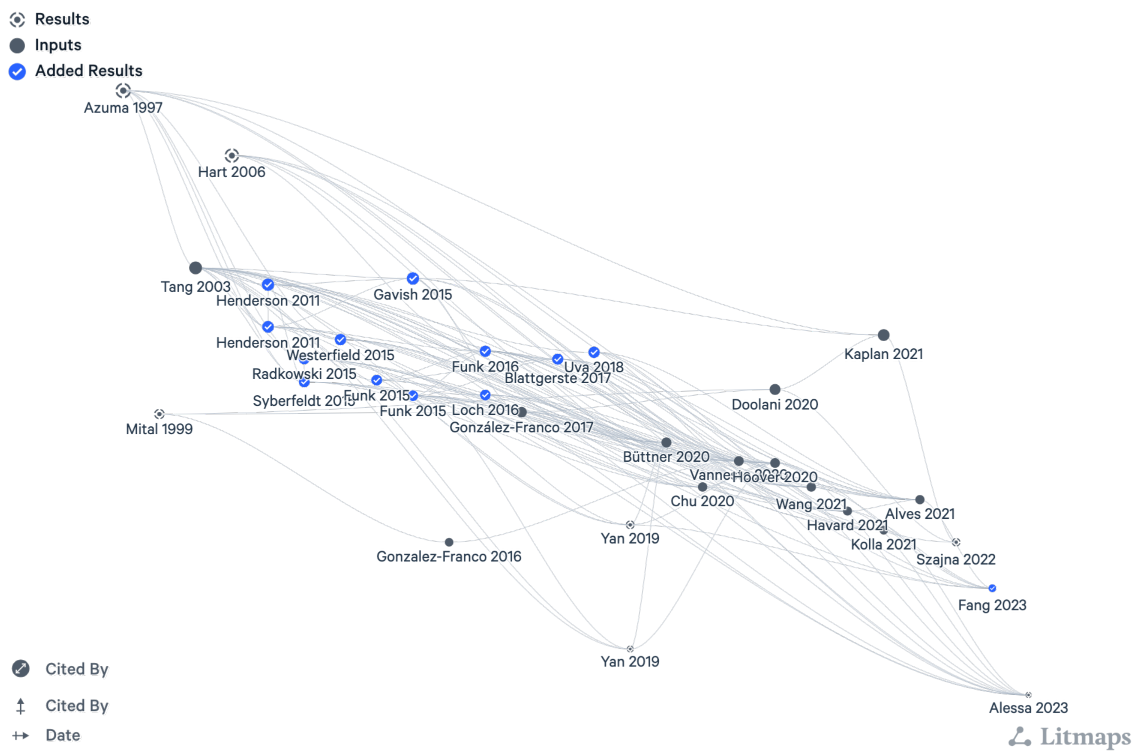


Figure 2.1: Discovering Relevant Sources in Litmaps

each question and topic area. The resulting reference collections were imported into Zotero, which managed the bibliographic data and related PDFs. Notes made while reading these sources were imported to Obsidian for review and synthesis.

This approach involves several sources and tools, the implementation of which creates technical challenges that may dissuade many researchers. Over time, as the benefits are better understood and more integrated workflows emerge, it seems likely that it will become widely adopted.

2.2 Chapter Overview

The adoption of augmented and mixed reality (AR/MR) technologies for manufacturing training has shown promise, yet faces significant barriers that hinder widespread implementation. This literature review provides a comprehensive examination of this technology and related research. It begins by describing the challenges faced by the manufacturing industry in the I4.0 era, highlighting the need for effective and efficient training methods to address the growing skills gap. That provides the context and motivation for this study, and introduces extended reality devices as essential components of the I4.0 technology stack.

The chapter continues with a thorough review of XR technologies, including their history, applications, and potential for training and support in the evolving landscape of manufacturing. It details the human and technical requirements of XR and the trade-offs required for its successful adoption in manufacturing settings. This knowledge provides a foundation for assessing the value of XR in manufacturing applications and designing the study that follows.

The review next covers preliminary results highlighting the potential benefits of AR/MR in manufacturing, before exploring the theoretical basis for these benefits. This focuses on differentiating AR from VR and examining relevant theories of learning and cognition. The review then addresses the barriers to widespread adoption of AR/MR in manufacturing, including technical limitations, market considerations, and social/legal issues. A number of empirical case studies are reviewed to better understand the quantifiable benefits of AR/MR in this domain. Finally, existing tools and frameworks for the development and assessment of AR/MR systems in manufacturing training are examined. The review concludes by summarizing the findings, describing the proposed framework, and enumerating important considerations for future research.

2.3 Advanced Manufacturing

Ongoing changes in manufacturing are expected to have profound impact on the people, businesses, and governments of the world. The so-called *Fourth Industrial Revolution* (4IR) follows prior revolutions of mechanization, mass production, and digitization, and is the first to be predicted in advance, not observed after the fact (Drath & Horch, 2014). First described by German economist and founder of the World Economic Forum, Klaus Schwab (2015), 4IR is driven by today's rapidly evolving and converging digital technologies. Brynjolfsson and McAfee (2014) note that these *Second Machine Age* advances uniquely exhibit sustained exponential rates of improvement while being easily combined and efficiently distributed. The innovative fusion of these cross-disciplinary technologies is transforming our physical, digital, and biological worlds in unprecedented ways.

2.3.1 Industry 4.0

Four years before Schwab's 4IR keynote, European manufacturing leaders had already imagined the potential benefits of digital convergence. In January of 2011 Germany's BMBF (the Federal Ministry of Education and Research) announced a new initiative. "Industrie 4.0" (I4.0) was introduced as the digital transformation of manufacturing, a paradigm shift intended to protect and expand Germany's influence as a world leader in the sector (Kagermann et al., 2011). Since then, I4.0 has become a prominent trend in Advanced Manufacturing. Its adoption is driven by a combination of application-pull (social, economic, and political change) and technology-push (automation, digitalization, communication, and miniaturization) market factors (Lasi et al., 2014).

I4.0 is a data-driven approach to manufacturing, where product specifications direct aspects of production. This is accomplished with connected, automated, autonomous components that respond in real-time to variable requirements (Negri et al., 2017). I4.0 is therefore advocated as the means by which manufacturing operations can meet modern organizational and societal demands for increased decentralization, flexibility, and resilience (Tao & Zhang, 2017). Time and cost to market and productivity are also expected to improve, along with sustainability measures, including energy cost and emissions. There is widespread optimism for these outcomes and their positive overall effect on global economic growth (Kagermann, 2013).

That optimism has encouraged the adoption of I4.0 methods worldwide. The Industrial Internet Consortium (IIC), founded by AT&T, Cisco, General Electric, IBM, and Intel, is the most prominent of several I4.0-related alliances in the United States (Hardy, 2014). As of 2021, the IIC (now known as the Industry IoT Consortium) boasts more than 150 member companies. Other major initiatives are underway in the UK, Taiwan, Japan, South Korea, France, Turkey, and more (Oztemel & Gursev, 2020). As of 2015, China was reportedly investing over $200B / year to related research and development. This bid to move from imitator to innovator is a clear signal of the returns that China expects from new markets and efficiencies unlocked by its I4.0 transformation (Woetzel et al., 2015).

Though a crisp definition of I4.0 might be expected given the support it has received, the literature is sorely lacking. It seems that "Industry 4.0" simply emerged as the most popular of several names given the technology-driven manufacturing renaissance that was commonly expected to result from its digital transformation (Culot et al., 2020). The integration of adjacent schools of thought, including "Industrial Internet" (Evans & Annunziata, 2013) and “Smart Manufacturing” (Radziwon et al., 2014), partially explains the lack of a standard definition for I4.0. Rapid divergent development by academics and practitioners and overzealous marketing have also contributed to the diffusion of this idea.

In fact, the literature suggests that I4.0 is best understood as a general concept, philosophy, or vision of manufacturing characterized by a group of functionalities, including process integration, real-time information transparency, virtualization, and autonomy, and their enabling technologies (Culot et al., 2020).

I4.0 has been linked to over 1200 technological components, from 30 disciplines (Chiarello et al., 2018). To provide a useful definition of I4.0 in terms of the technologies involved, some abstraction is essential. In their review of over 100 relevant and credible sources, Culot, et al. (Culot et al., 2020) identified 13 categories of technology. Each was assessed along two continua: software-hardware technology and local-global connectivity, as seen in Figure 2.2.

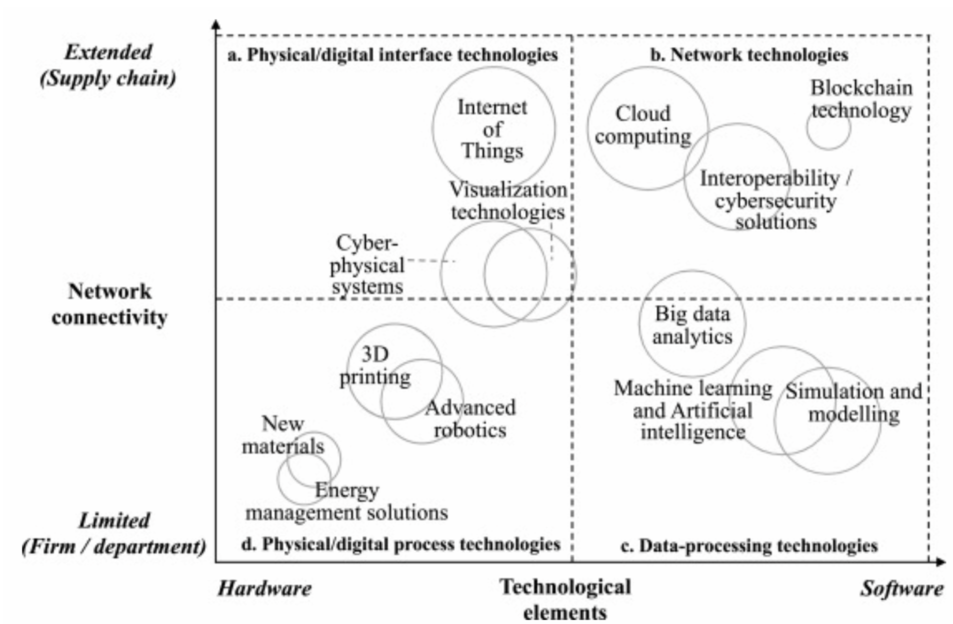


Figure 2.2: Enabling technology categories of Industry 4.0. Culot et al. (2020)

Four technology quadrants emerge in this figure: (a) physical-digital interfaces, (b) networking, (c) data-processing, and (d) physical-digital processes. The sensing, connecting, and analyzing activities of the first three quadrants are what differentiate I4.0 from advanced manufacturing.

As described in the next section, the specific technologies and the manner in which they are integrated and applied define an I4.0 system. The permutation of possible outcomes, each a different embodiment of the I4.0 concept, is the ultimate source of definitional ambiguity in this field.

2.3.2 Cyber-Physical Systems

Cyber-Physical Systems (CPSs) is an emerging cross-disciplinary field engaged in the design of new models and methods for problems at the intersection of physical and digital engineering traditions (E. A. Lee, 2015). It was simply defined in E. Lee's seminal paper (E. A. Lee, 2006) as the "integrations of computation with physical processes." CPS promotes the novel evolution of classic embedded systems through their interconnection and integration with computation and control mechanisms. This enables the real-time autonomous control of large engineering systems (E. A. Lee, 2006; Pascual et al., 2019). Though commonly associated with I4.0, CPS is independent of specific applications or implementations, e.g., I4.0 and IoT.

CPS enables I4.0 by integrating the previously identified sensing, connecting, and analyzing capabilities to "monitor and control physical processes, usually with feedback loops where physical processes affect computations and vice versa" (E. A. Lee, 2006). An I4.0 CPS is comprised of physical objects, networked data models of those objects, and services based on that data (Drath & Horch, 2014). Their technical building blocks are summarized below (Bottani et al., 2017):

- Internet of Things (IOT) - sensored networked devices
- Machine-to-Machine (M2M) - interconnected, interoperable systems
- Digital Twin (DT) - mirroring of physical and virtual objects
- Cloud Computing - distributed computing services
- Big Data - large scale data capture, storage, and analysis
- Modeling - data or physics driven methods for descriptive, diagnostic, predictive, and prescriptive analysis
- Extended Reality (XR) - virtual, augmented, or mixed reality visualization and interaction
- Advanced Manufacturing - including additive methods, automation, and robotics

To understand their roles in an I4.0 CPS, J. Lee's 5C Architecture is instructive (J. Lee et al., 2015). This popular framework identifies five implementation activities in step-wise fashion: get data from sensors, convert data to information, analyze information, present data, and provide control feedback. These activities correspond to the 5Cs of Connect, Convert, Cyber, Cognition, and Configuration, as depicted in Figure 2.3, with related attributes.

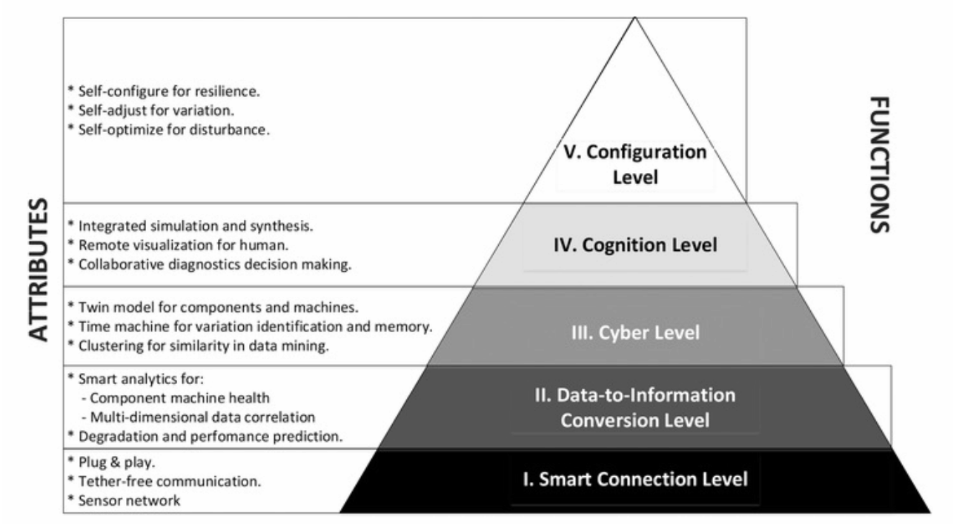


Figure 2.3: 5C architecture to implement I4.0 CPS. J. Lee et al. (2015), as adapted by Pascual et al. (2019)

In this framework IOT and M2M enable smart Connections between sensored devices. In the Conversion level Big Data collects and contextualizes the data. Virtual representations of the physical components are created by Digital Twins in the Cyber step. Extended Reality devices aid visualization and Cognition. Various Modeling methods are employed throughout to support manual and automated decision-making. The Cloud Computing architecture integrates it all and facilitates feedback in Cognition. The resulting closed loop system drives Advanced Manufacturing processes in real-time.

Ideal I4.0 CPS systems are fully integrated within the enterprise: horizontally, vertically, and across the system life-cycle. Horizontal integration occurs across the value chain, from supplier and production to end customer. Vertical integration covers the manufacturing hierarchy, from the shop floor to enterprise planning (Pascual et al., 2019). Fully realized systems are driven by individual product specifications, maximizing flexibility and resiliency, along with their attendant social, market, and sustainability benefits. Taken to the limit, such systems are capable of operating with a batch size of one, the ultimate Lean Manufacturing benchmark and the key to unlocking mass personalization and customization (Culot et al., 2020; Kagermann et al., 2011; Lasi et al., 2014).

In the following section we focus on the heart of Lee's 5C architecture, the Digital Twin.

2.4 Digital Twins

The Digital Twin is the mechanism by which I4.0 synchronizes the virtual and physical system states. It consists of a virtual replication of the system that is coupled to its physical counterpart via a bi-directional flow of sensor and control data. DTs enable a data-driven approach to life-cycle management that can employ optimal methods and practices for each environment. The continuous, bi-directional data flow and synchronization of an idealized DT differentiates it from traditional modeling and simulation methods which typically operate as off-line, asynchronous processes (Jones et al., 2020).

The DT concept was introduced by Michael Grieves in late 2002, partly inspired by dynamic CAD modeling methods that were then emerging. He originally promoted it as a tool for distributed, collaborative problem solving in product life-cycle management (PLM) (Grieves & Vickers, 2017). Grieves developed the idea under different names until 2011, when he first used the phrase Digital Twin to describe it (Grieves, 2011). Therein he credits collaborator John Vickers of NASA with coining the term, which also appeared in NASA's draft strategy for Simulation-Based Systems Engineering in 2010 (Shafto et al., 2012).

Following similar growth in adjacent fields, interest in the DT concept has accelerated rapidly since 2016. While most research activity remains focused on Industry 4.0 applications, progress in academia and industry has led to some divergence in both interpretation and application of the concept (Ante, 2021). A 2017 survey of manufacturing literature found that no less than 16 unique definitions had been proposed for Digital Twin since 2011 (Negri et al., 2017). Despite the literature offering no common understanding of the term, the DT concept is recognized as a key enabler for I4.0 (Kritzinger et al., 2018).

2.4.1 The Synchronization Process

Jones described the DT synchronization process, also known as "twinning," as a cycle of measuring and reflecting changes in the parameters of interest (Jones et al., 2020). During metrology, changes to one system state are measured. In the realization phase those changes are reflected in the other system. This process operates bi-directionally between physical and virtual entities, creating a system that is capable of continuous adaption. See Figure 2.4.

In Jones' model the term parameter refers to the values synchronized by the DT. Common parameters are related to form, functionality, process, and performance. Examples include

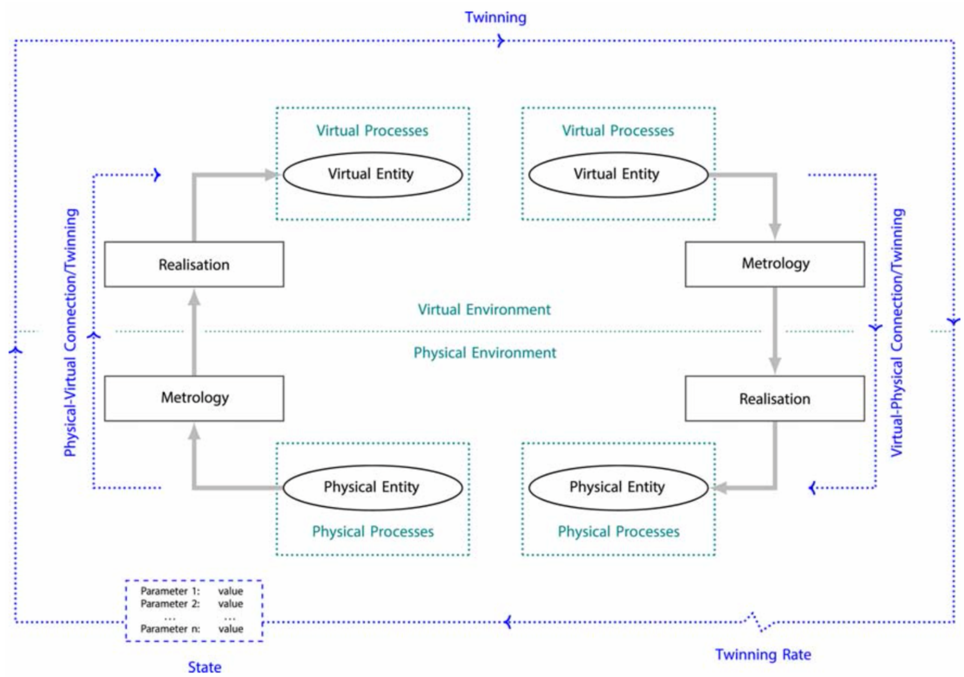


Figure 2.4: The Digital Twinning Process. Jones et al. (2020)

part tolerance, assembly time, and machine health. Parameters can be measured, computed, observed, or otherwise derived. The overall system state is described by the current value of all parameters. The fidelity of a DT is a measure of the number of parameters, their accuracy, and the level of abstraction involved (Jones et al., 2020).

The DT concept is not entirely new. Elements of it are evident in other fields, including Computer-Integrated Manufacturing and Virtual Manufacturing Systems, both of which predate Grieves' work. Of those, only Model-Based Predictive Control, Advanced Control Systems, and Building Information Modeling (BIM) share DT's approach to closed-loop control (Jones et al., 2020).

2.4.2 Life-cycle Considerations

This method is most valuable for objects that are changing over time, and when measurement data that can be correlated with this change can be captured (Wright & Davidson, 2020). To account for this, Grieves describes two manifestations of the Digital Twin: Prototype (DTP) and Instance (DTI). The DTP models a prototypical physical object, providing an idealized, immutable reference for that thing, including the means to produce physical instances of it. The DTI is the virtual reflection of a unique, as-built thing in the world. Multiple DTIs are maintained, each synchronized with a single instance of the physical object for the duration of its life-cycle (Grieves & Vickers, 2017).

The DT model is dynamic. In each phase of the system's lifecycle (creation, production, operations, disposal) the directionality of metrology and reflection changes. Modeling tools are first used to develop and test the DTP in the creation phase. Physical instances are derived from the DTP in production, when their as-built specifications are captured and reflected in corresponding DTIs. During the operations phase the real-virtual link becomes bi-directional, synchronizing the system states and enabling continuous adaption. Finally, information about the system is used to properly dispose of it, before being archived for the benefit of future designs (Grieves & Vickers, 2017).

Data collected throughout this process is used by various modeling methods that support the Conversion, Cognition, and Configuration levels of Lee's 5C architecture. The conversion layer primarily relies on descriptive and diagnostic approaches to interrogate and analyze system status. The cognition layer utilizes predictive methods to aid human understanding. Prescriptive methods that recommend specific actions are employed in the configuration level to drive continuous adaption through parameter optimization or policy selection (Bottani et al., 2017).

2.4.3 State of the Art

In 2017, Grieves set the lofty goal for models that "fully [describe] a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level" (Grieves & Vickers, 2017). His position is representative of a bias towards fidelity that is commonly expressed in the literature, despite the absence of any example using more than a subset of the known parameters (Jones et al., 2020). Digital Twins that perfectly replicate the reality of complex systems in real-time may never be practical. Tradeoffs must be made between fidelity, accuracy, available compute, and update rate. Models need only be sufficiently physics-based, accurate, and quick to meet the system requirements in a trustworthy manner. This depends on properly managing model verification and validation, uncertainty, model selection, and associated metadata (Wright & Davidson, 2020).

We are still far from the idealized DT described above. Though many perceived benefits have been identified, few papers include quantitative analysis to validate those claims (Jones et al., 2020). Most research in the area is concept oriented. Of the few published case studies, most systems are uni-directional, with low fidelity and/or little integration

(Kritzinger et al., 2018). Implementation relies on connections between physical and digital systems that are often difficult to implement without human involvement, and current modeling tools fall well short of understanding and replicating the physical world (Grieves & Vickers, 2017). Limited collaboration and a lack of technical standards are also commonly noted (Ante, 2021). Together, these shortcomings hinder development and slow adoption of the DT concept.

Though the research area remains immature, a number of additional frameworks have recently emerged in response to issues with standards, validation, fidelity, and interoperability. Grieves' Tests of Virtuality (GTV) were proposed as a means to evaluate the fidelity and validity of a DT. Performance is assessed by comparing the look, behavior, and synchronization of a physical system and its virtual counterpart (Grieves & Vickers, 2017). Tao's seminal paper describes the DT of a shop floor in terms of its architecture and technology. Architecturally, he identifies four integrated layers: geometry, physics, behaviors, and rules. The many necessary technologies are grouped into the five areas of interconnection and interaction, modeling and verification, construction and management, operation and evolution, and smart production services (Tao & Zhang, 2017).

At least two maturity models have been proposed. Kritzinger's model is based only on the level of physical-virtual integration, as expressed by the Digital Model (DM), Shadow (DS), and Twin (DT) classification scheme. A DM has no connection or uses manual methods of data exchange. One-way flow of data characterizes the DS, while bi-directional flow is the hallmark of a DT (Kritzinger et al., 2018). Hyre's model also considers how capability and complexity increase with a DT's level of integration. Her 4Rs (Representation, Replication, Reality, and Relational) provide a framework for the incremental development of a DT that incorporates verification and validation of the system (Hyre et al., 2022).

2.4.4 DTs for the Development and Testing of Complex Systems

The physical-virtual synchronization of Digital Twins enables the operational benefits of an I4.0 CPS, as previously described. That twinning process requires a trustworthy virtual replication of the system, which offers many additional benefits for the development and testing of these complex systems.

A complex system is defined as one in which connections between the components are unfamiliar, unplanned, unexpected, and/or invisible, making it difficult to predict system states

(INCOSE, 2015). Such systems are prone to "Normal Accidents," in which cascading failures escalate suddenly and often catastrophically. Human inconsistency (following rules, processes, and procedures) and poor sensemaking (understanding what is perceived) often play a role in those accidents, especially in high stakes situations when good decision making is most critical (Perrow, 1999).

Complex systems are the domain of Systems Engineering, where traditional methods rely on the verification and validation of physical objects. This approach, exemplified by the commonly used Waterfall, Spiral, and Vee models, is expensive, centralized, and sequential. As a consequence, it focuses the scope of investigation on areas where undesirable effects are predicted. The most dangerous category of system behavior, that which leads to unpredicted and undesirable outcomes, is often first encountered when the system is deployed, creating the risk of catastrophic failure and harm to the users (Grieves & Vickers, 2017).

Digital methods are, by contrast, low cost, composable, and easily distributed (Brynjolfsson & McAfee, 2014). Trustworthy virtual systems can be tested more thoroughly than the physical equivalent, with less risk. Increased test coverage helps identify and mitigate unpredicted, undesirable outcomes. Reduced risk permits the evaluation of circumstances that traditional methods would not allow. Thus, DTs can test more broadly, including conditions that are uncommon or hazardous and/or involve interaction with a diversity of personnel. This directly addresses the leading causes of those "Normal Accidents" that we seek to avoid, and is a primary intended benefit of the Digital Twin (Grieves & Vickers, 2017).

2.4.5 DTs for Visualization

Though DTs are widely embraced as the synchronizing mechanism in an I4.0 CPS, and for the development and testing of complex systems, they offer another important benefit. As previously mentioned, the concept was first promoted as a tool for collaborative problem solving; a way for stakeholders to understand and visualize the current system state.

A Digital Twin improves problem solving and innovation by aiding the human processes of conceptualization, comparison, and collaboration. Effective visualization simplifies the cognitive steps involved in translating symbolic information, facilitating conceptualization. Overlaying the physical and virtual allows for direct comparison, which is ideal for human perception and analysis. Collaboration is enabled by digitally replicating and distributing the experience to an audience of stakeholders (Grieves, 2015).

Visualization is an essential outcome of the Digital Twin concept. High fidelity interactive visualizations of virtual systems can be shared globally in real-time using modern technology, allowing the direct, side-by-side visual comparison of the physical and virtual product. Today, the tools and technologies best suited to deliver on this promise are found in the area of Extended Reality.

2.5 Extended Reality

Extended Reality (XR) is the umbrella term for a range of technologies where human-machine interactions occur in environments that blend real and simulated stimulus (UL, 2022). XR covers the entire Virtuality Continuum (VC), as famously described by Milgram and Kishino (1994), and pictured in Figure 2.5.

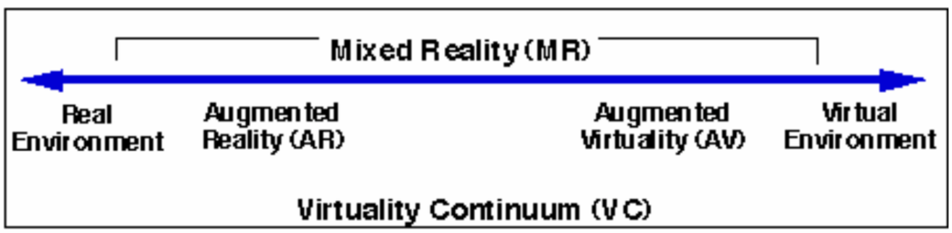


Figure 2.5: The Virtuality Continuum. Milgram and Kishino (1994)

This continuum spans the complete range of real to synthetic experiences. Though typically associated with adding or replacing visual stimulus, the VC also includes technologies that are subtractive in nature and/or affect other senses. For example, noise cancellation headphones can be considered a form of "diminished reality" audio AR device (Kress, 2020).

2.5.1 Origins of XR

Many precursors to XR can be identified in the 1800s and early 1900s, culminating in Morton Heilig's patented head-mounted display (HMD) in 1960, which boasted 140° field of view, stereo earphones, and air / scent discharge nozzles (Heilig & States, 1960). As seen in Figure 2.6, images from the 60 year old filing are surprising in their familiarity. Soon thereafter, engineers at the Philco Corporation created the first such device that tracked the wearer's head motion and updated the display accordingly (Jerald, 2016).

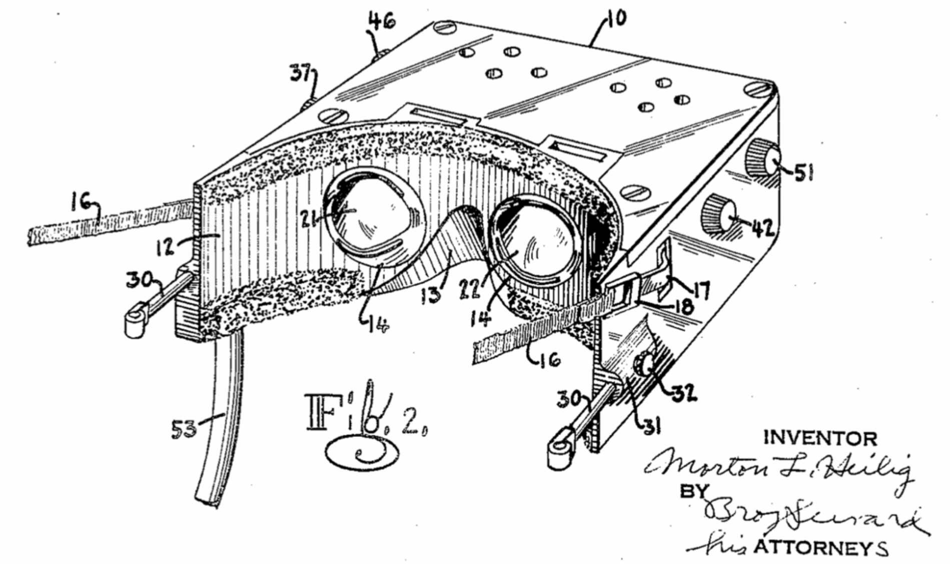


Figure 2.6: Heilig's Stereoscopic Television Apparatus. Heilig and States (1960)

In 1965 Ivan Sutherland[10] published *The Ultimate Display*, which described his vision for a "kinesthetic display" at a time when "the ability to draw simple curves would be useful" (Sutherland, 1965). In it, he commented:

> *A display connected to a digital computer gives us a chance to gain familiarity with concepts not realizable in the physical world. It is a looking glass into a mathematical wonderland.*

Three years later, Sutherland and his students at the University of Utah were first to demonstrate a HMD that combined tracking and computer generated imagery. The device, known as the Sword of Damocles, is the original prototype for all modern VR technology. Its name was in reference to the story of King Damocles, owing to the precarious position the device maintained over a user's head (Kiyokawa, 2015). It took nearly 30 years for its AR equivalent to emerge.

In 1994 Ronald Azuma presented the first AR system capable of accurately maintaining the spatial registration of real and virtual objects based on changes to the user's viewpoint. Key

[10] Ivan Sutherland is a distinguished computer scientist, known for pioneering work in computer graphics and interactive computing. During his tenure at the University of Utah, he co-founded real-time graphics pioneer Evans & Sutherland, and fostered a generation of computer graphics experts. Sutherland is credited with creating the first graphical user interface and fundamentally changing computer-aided design. He has received several prestigious awards for his lifelong contributions, including the ACM Turing Award (1988) and the Kyoto Prize (2012).

contributions of that open-loop system included custom hardware, calibration, and head pose prediction methods (Azuma & Bishop, 1994).

Commercial interest in XR has since experienced alternating periods of boom and bust, fueled by promises that exceeded the technologies of the time. Through it all, research in the corporate, government, academic, and military sectors continued. Capitalizing on the runaway success of the smartphone industry following the 2007 iPhone launch, the current wave of XR began to emerge in 2012. This generation of hardware leveraged newly available components, including displays,[11] processors, batteries, cameras, and sensors, along with the maturing software infrastructure, to offer products that were more sophisticated and compelling in all sectors (Kress, 2020).

Emblematic of that shift is Field of View To Go (FOV2GO), an experimental, untethered, DIY HMD developed in the Mixed Reality Lab at the University of Southern California's Institute for Creative Technologies, and first shown at the IEEE VR conference in 2012 (Olson et al., 2011-03-19/2011-03-23). Their design utilized two iPhone 4's as displays with an off the shelf lens assembly and tracking system, all mounted on a cardboard body. Software was powered by the Unity game engine and a Python script. Their conference poster is pictured in Figure 2.7.

FOV2GO team members founded Oculus VR soon thereafter and demonstrated a prototype of their Rift VR HMD in June of that year. The Rift Kickstarter campaign launched in August, meeting its $250,000 funding target in less than four hours and securing over $2.4m in total. Oculus subsequently raised over $90m in venture capital before being acquired by Facebook for $2b in March of 2014 (Jerald, 2016).

XR has experienced tremendous growth and development in the last 10 years. Many declinations of XR have been identified, including Virtual, Augmented, Mixed, Blended, and Merged Reality. The literature identifies significant overlap and some disagreement in their interpretation. Of those, virtual and augmented reality are the most agreed upon terms.

2.5.2 Virtual and Augmented Reality

Virtual Reality (VR) is a synthetic, multi-sensory experience that imitates real-world interactions. VR is a very concrete concept in which purely synthetic environments are experienced through opaque HMDs, via interactions that are primarily controller-based. This

[11] In this context the term *display* can apply to devices that present information for any human sense. For example, a speaker is an audio display, and haptic devices are displays for the senses related to touch.

Figure 2.7: FOV2GO, IEEE VR Conference 2012 Poster. Olson et al. (2011-03-19/2011-03-23)

combination of familiar features has been experienced by many, thanks to the availability and maturity of consumer devices like the Meta Quest. VR is widely understood as a way to provide immersive experiences that lead to the sensation of presence.

Immersion is the degree to which an XR experience provides consistent, believable inputs with corresponding outputs. It is a function of the range and congruence of the sensory modalities involved, the quality and spatial cohesion of the displays used, and the simulation's responsiveness to user interaction (Slater & Wilbur, 1997). Vividness and interactivity are often cited as the functional mechanisms underlying the efficacy of XR (Jiang & Benbasat, 2007; Steuer, 1992). A study by Yim et al. (2017), involving over 800 US college students found that immersion plays a mediating role in that relationship. That is, vividness and interactivity promote immersion, which promotes presence.

Yim's study defined vividness as the ability of the technology to display high fidelity stimuli over multiple sensory channels. Interactivity was described as a function of both the underlying technology, including responsiveness, interface, and overall level of interaction supported, and the quality of the experience's design and implementation. Together, technology and design enable and engender interaction (Yim et al., 2017). The depth of immersion is a characteristic of the hardware and software involved, and its effects are subjective. The way different users experience immersion is known as presence.

Presence refers to a psychological state that can result from immersion, and is commonly defined as "a sense of being there" (Cummings & Bailenson, 2016). Presence is associated with an "illusion of nonmediation," where users fail to perceive or acknowledge the existence of the interfacing technology and act as if it were not there (Lombard & Ditton, 1997). A strong sense of presence leads to experiences that are perceived as real, generating cognitive, psychological, and behavioral effects that are similar and long-lasting (Bailenson, 2018). While presence can also occur in AR, other mechanisms of the medium have a stronger, more valuable effect.

Augmented Reality (AR) is a more abstract and nuanced concept which has so far refused to converge on a single implementation. As originally described in Azuma's highly cited first survey of the field, Augmented Reality (AR) systems "combine real and virtual, are interactive in real time, and are registered in 3-D" (Azuma, 1997). The value of AR comes from its ability to enhance a user's natural interaction with and perception of the real world.

Azuma's definition demands real-time interaction with a spatially coherent mix of real and virtual objects. This new interface paradigm is based on concepts that would become known as Spatial Computing, which Greenwold defined as “human interaction with a machine in which the machine retains and manipulates referents to real objects and spaces” (Greenwold, 2003). In this way, AR proposes to replace metaphorical input devices like the keyboard and mouse with sensor-based interfaces that directly measure and interpret the world and our actions in it.

In a general sense, AR systems can enhance perception by mapping any sensor input to any mix of displays, allowing users to see, hear, feel, etc. in ways not normally possible. Sensor inputs can refer to either raw data from a single measurable phenomenon or “fused” data developed from multiple sources. Interaction also benefits from the user’s improved understanding (Azuma, 1997).

Traditional AR and VR devices integrate computation, sensors, and displays into a HMD, which may suggest they offer a similar experience and benefits. Both offer novel forms of visualization and interaction, but the essential characteristics of each are entirely different. In the study of interaction design and related fields these characteristics are referred to as *affordances*, the quality or property of an object that defines its possible uses or makes clear how it can or should be used (Norman, 2013). For example, a button affords pushing and a handle affords pulling.

VR is a new medium that immerses the senses in a virtual replacement for reality and, through the psychological phenomena of presence, mimics the effects of as-lived events (Bailenson, 2018). AR is a new model of computing that augments our perception of reality and, through a natural, spatially connected interface, enhances our understanding of and interactions with the real world (Azuma, 2019). Where VR is an extension of games and film, AR is seen as the most likely next step on the path towards ubiquitous computing.

2.5.3 Ubiquitous and Wearable Computing

Ubiquitous computing is the idea, first proposed by Weiser at the Xerox Parc research lab in 1988, that technology should or will be completely assimilated, disappearing into the woodwork of our lives (Weiser, 2002). The steady march of miniaturization began with the invention of the transistor and has continued ever since. Today this trend presses the limits of human physiology, where human interfaces, not computational considerations, constrain the size of machines. Ubiquitous computing requires the replacement of physical interfaces with more natural mechanisms (Greenwold, 2003).

In the field of wearable computing, the assimilation of technology is the goal. A wearable computer is any worn or body-borne computer that is designed to provide useful services while the user is performing other tasks. Their on-the-go use and background operation are the primary characteristics that distinguish wearables from other computing devices. This is accomplished through interfaces designed to be unobtrusive and unencumbering, if not entirely hands-free (Starner, 2015). From the beginning, research in the field has been egocentric, i.e., focused on the user and their interaction with the world. Devices that supplement the user's memory and data retrieval, or augment their view have been demonstrated since the late 1990s (Billinghurst et al., 2015). Wearables are always-on devices that rely on sensor-based interactions with and between the user and their environment (Barfield, 2015).

The potential benefits of such a device have been recognized by industry since the 1990s, when AR R&D was already exploring the areas of medical visualization and training, manufacturing and repair, annotation and visualization, robot path planning, entertainment, and military aircraft navigation and targeting (Azuma, 1997).

2.5.4 XR Devices

While VR has converged on a singular form, Azuma's definition of AR is not constrained to any particular display type or "mix" of real and virtual. As such, XR includes a diverse range of possible devices, each best suited for different use cases. This is summarized in Figure 2.8 from Bernard Kress'[12] 2020 book, *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets* (Kress, 2020). Kress divides the range of XR HMDs into four classes: smart eyewear, VR, AR, and Mixed Reality. In his taxonomy, Mixed Reality refers to AR devices with the precise world tracking capabilities and other advanced spatial features.

From this chart it can be inferred that HMD physical configurations vary by:

- form factor: overall size, shape, and balance
- displays integrated: visual, audio, haptic etc.
- visual display type: opaque or optical / video see-through

[12] Dr. Kress was principal optical architect on the Google Glass project before joining Microsoft in a similar role for their first and second generation HoloLens devices. He has since returned to Google as their Director for XR Engineering. He serves as Vice President of the International Society for Optics and Photonics (SPIE). Dr. Kress' publications are heavily leveraged throughout this section. SPIE Profile: https://spie.org/profile/Bernard.Kress-16356

			Consumer		Enterprise			Medical		Defense	
		Product examples	Day-long usage	Occasional indoor usage	Factory floor (shifts)	Heavy outdoor industry	R&D	Non-surgical	Surgical	Training	Battlefield
Smart glass	Audio only smart eyewear w prescription correction	Bose Frames Huawei/Gentle Monster, Amazon Echo Frames	+++	+++	++	+++	+	++	++	+	+
	Rugged Smart Glasses, monocular, opaque	RealWear HMT-1 Vuzix m300	+	+	++	+++	++	+	---	---	++
	Smart Glasses, monocular, see-through	Vuzix Blade Digilens Mono HUD Optinvent ORA	+	++	++	++	++	+++	+	+	++
	Smart eyewear w display and prescription correction	Google Glass North Focals **Bosch Frames** Lumus DK32	+++	+	+	---	---	+++	+	+	---
VR	Standalone VR without video see-through (3DOF)	Oculus GO Google DayDream VR Samsung Gear VR	---	+++	---	---	+	---	---	+++	---
	Standalone VR with video see-through	Oculus Quest NTC Vive Focus 2.0 Pico Neo	---	+++	+++	+	+++	+	+++	+++	---
	PC tethered VR with inside-out sensors (6DOF)	Oculus Rift "s" HTC Vive Pro Windows MR 3rd party	---	+++	-	---		---	+	+++	---
	Large FOV PC tethered VR headsets	Varjo VR Foveated PiMax 8k Acer Star VR	---	+	-	---	+++	---	+	+++	---
AR and entry - level MR	Tethered AR headsets to PC	Meta 2 DreamWorld Glasses	-	+	++	---	++	--	+++	++	---
	Standalone AR headsets	Epson Moverio Lumus DK50 / Vision Digilens Cristal	-	+++	++	++	++	++	+	+	++
	Standalone AR headsets w 6DOF and gesture sensing	ODG R9 nReal AR glasses Daqri, Atheer Labs,	--	++	+++	+++	++	+++	+	++	+++
High-end MR	High end see through untethered MR	HoloLens V1 / V2	---	+	+++	++	+++	+	+++	+++	+++
	Pod-tethered high end see through MR	Magic Leap One Lenovo ThinkReality	---	++	++	+	+	+	++	+	+

Figure 2.8: Current product offerings by device class and market. Kress (2020)

- visual display ocularity: monocular, binocular, or stereo
- visual display location: centered or offset in the user's field of view
- tracking: none, three, or six degrees of freedom
- input modalities: controllers and/or gestures
- tethered or standalone
- integrated vision correction

World-fixed and hand-held alternatives to HMD XR must also be considered. World-fixed solutions use projectors or flat panel displays to surround the observer / participant with imagery. This is typified by the Cave Automatic Virtual Environment (CAVE[13]) invented in the Chicago Electronic Visualization Lab at the University of Illinois (Cruz-Neira et al., 1992). Hand-held XR implementations are common on smartphone and tablet devices, where integrated cameras, displays, and sensors enable screen-based AR that is device-centric (i.e., motion and display are relative to the device, not the user's head and eyes) (Jerald, 2016).

XR devices, particularly AR HMDs, are not "one size fits all." In addition to their physical configuration, key specifications strongly dictate the intended purpose of a device and its suitability for specific tasks. Technology limitations and the diverse requirements found in different application domains force trade-offs in system design and selection (Kiyokawa, 2015). Subsequent sections will discuss each of those considerations in greater detail.

2.5.5 XR HMD Requirements

All modern XR HMDs are complex devices comprised of display, sensing, compute, and power management systems. Optical see-through (OST) devices require additional components to project and combine the image in the user's field of view. Figure 2.9 depicts the major sub-systems of an OST HMD (Kress et al., 2020). The peak complexity of an idealized OST AR HMD provides a comprehensive case study in the tradeoffs and benefits of XR. Lessons learned from state of the art requirements and architecture apply, in limited fashion, to devices with a reduced feature set.

Mixed Reality (MR) is the label given by Kress to advanced AR devices with the precise head tracking, gesture sensing, and depth mapping capabilities required to support spatially synchronized interactions, providing an elevated and differentiated user experience (Kress & Cummings, 2017). He measures the ultimate quality of that experience in two dimensions:

[13] CAVE is a recursive acronym and reference to the *allegory of the Cave* from Plato's *Republic*, in which a philosopher contemplates perception, reality, and illusion.
en.wikipedia.org/wiki/Cave_automatic_virtual_environment

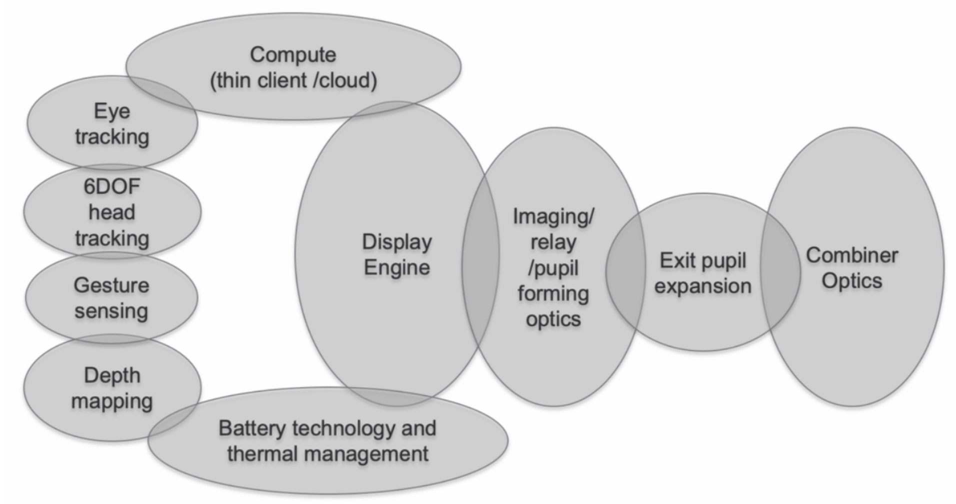


Figure 2.9: Functional building blocks of an OST XR HMD. Kress et al. (2020)

comfort, including wearable, vestibular, visual, and social components; and immersion, a function of all sensory input and output. Given the goals of comfort and immersion, an extensive list of design requirements can be derived for idealized MR devices. In Figure 2.10, dark grey shading indicates features that are reliant on fast, accurate, universal eye tracking, a critical enabling technology for idealized MR HMDs.

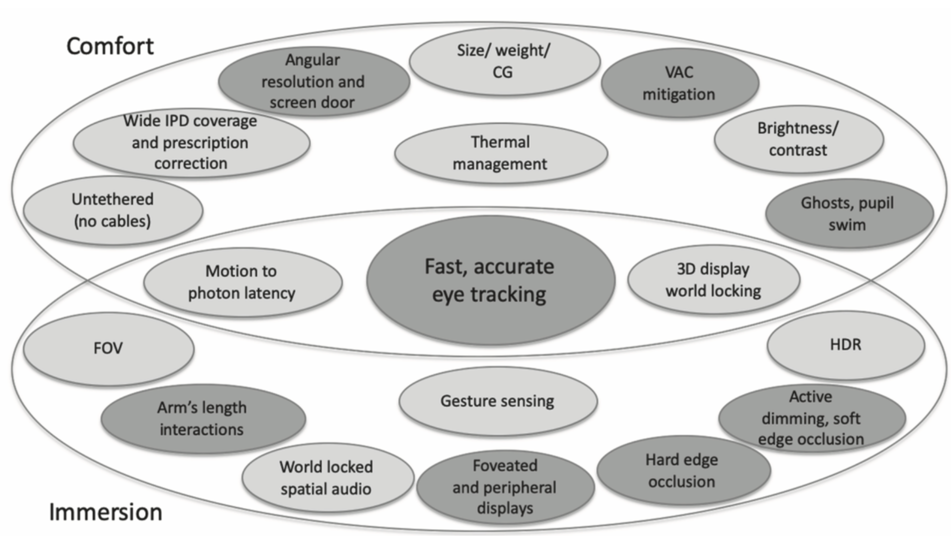


Figure 2.10: Comfort and immersion requirements for an ideal MR experience. Kress (2020)

This summary reflects other findings in the literature which identify requirements related to precise tracking, form factor, brightness / contrast, field of view, latency, resolution, occlusion, frame rate, depth of field, and visual discontinuity (Azuma, 2017; Fischer, 2015;

Gay-Bellile et al., 2015; Jerald, 2016; Kiyokawa, 2015; UL, 2022).

Visual comfort is a function of both the display features and the overall speed and accuracy of the integrated sensor output. Sensor fusion refers to that integration process and the hardware / software system that accomplishes it. Figure 2.11 depicts the inputs and processing flow for a typical system. The demands of sensor fusion have led companies like Microsoft to design custom processors to provide the best user experience (Kress, 2020).

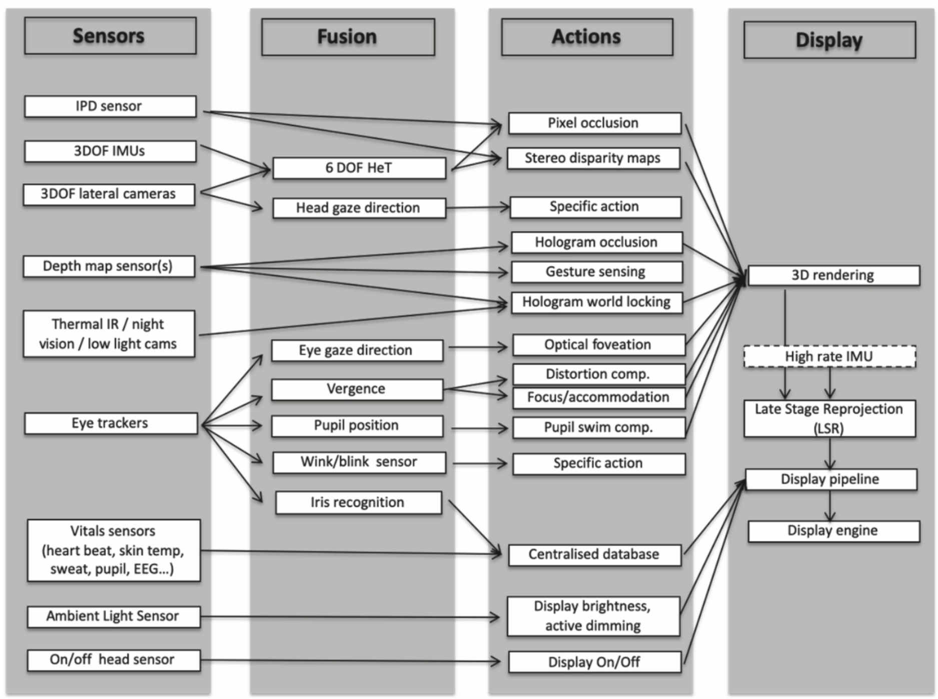


Figure 2.11: Sensor fusion flow in typical MR systems. Kress (2020)

High-level considerations in the design of HMD systems include tradeoffs between real world visibility and pictorial consistency, FOV and angular resolution, near and far accommodation, and the importance of perceived depth, which is influenced by occlusion and ocularity (Kiyokawa, 2015). Directly conflicting requirements are common in OST HMD design, where the tight interdependencies of these sub-systems and ambitious overall requirements necessitate a global optimization approach to design (Kress et al., 2014). Knowledge of the human factors involved can aid the process.

2.6 Human Factors

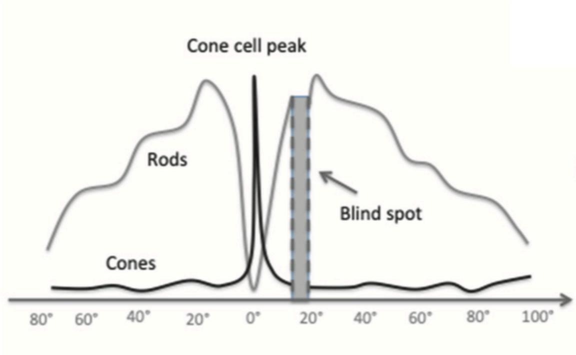
A human-centered approach to HMD development allows designers to tailor requirements to human needs rather than absolute measures of performance, reducing system complexity without impact to the immersiveness or comfort of the experience. The following sections provide a brief overview of human factors related to vision, balance, and motion. The senses involved are critical to both immersion and comfort.

2.6.1 The Visual System

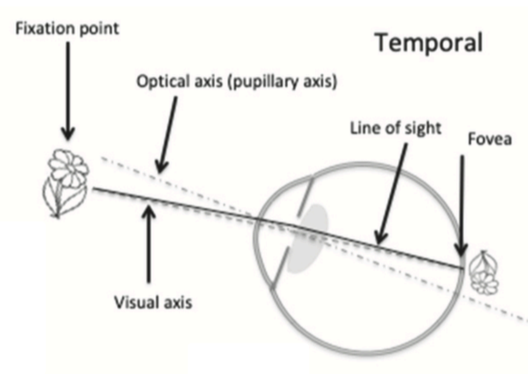
Optical components of the eye, including cornea, iris, pupil, and lens, coordinate to focus an image on the surface of the retina, where photosensitive cone and rod cells translate it into signals sent to the brain via the optic nerve. Cones are adapted to provide detailed color vision in high illumination. They are concentrated in the fovea, near the center of the retina, maximizing the eye's resolving power around the line of sight. Conversely, rods are concentrated in the visual periphery. They perform well in low light and are optimized to detect fast motion or flicker. The resulting signals follow different visual pathways in the brain, where they are strongly influenced by other sensory systems and cognitive processes, forming our subjective, conscious perception of the experience.

Visual Acuity

Visual acuity refers to a group of measures for human visual performance, including separation and recognition acuity. Separation acuity is the ability to resolve fine details at a distance. Specifically, it is the smallest angular separation that can be resolved between neighboring black stripes on a white background. One arc minute (1/60th of a degree) is the lower limit for "normal" separation acuity, corresponding to a gap of just over 1/16" (1.75mm) when viewed from 20' (6m). This attribute of human vision is rarely measured directly. Instead, recognition acuity tests like the Snellen eye chart are designed to assess separation acuity via the discernment of shapes or symbols. The results are given as a ratio expressing the acuity of the subject relative to someone with "normal" (20/20) vision. For example, "20/40" indicates half the normal acuity. Visual acuity is influenced by the entire optical-neural path, but is primarily a function of the cones and varies with their distribution in the field of view. These concepts are illustrated in Figure 2.12.



(a) Code and Cone Cell Density on the Retina



(b) Optical Axis and Line of Sight

Figure 2.12: Visual Acuity Varies with Rod and Cone Density in the Field of View

Field of View

Field of view (FOV) is the angular measure of the environment that is visible at any instant. As shown in Figure 2.13, the horizontal FOV is approximately 160 deg for each eye, and 200-220 deg combined. Vertical FOV is slightly smaller, with a slight downward bias. Overlapping monocular vision creates a central binocular range of 120 deg with vertical asymmetries caused by the facial profile.

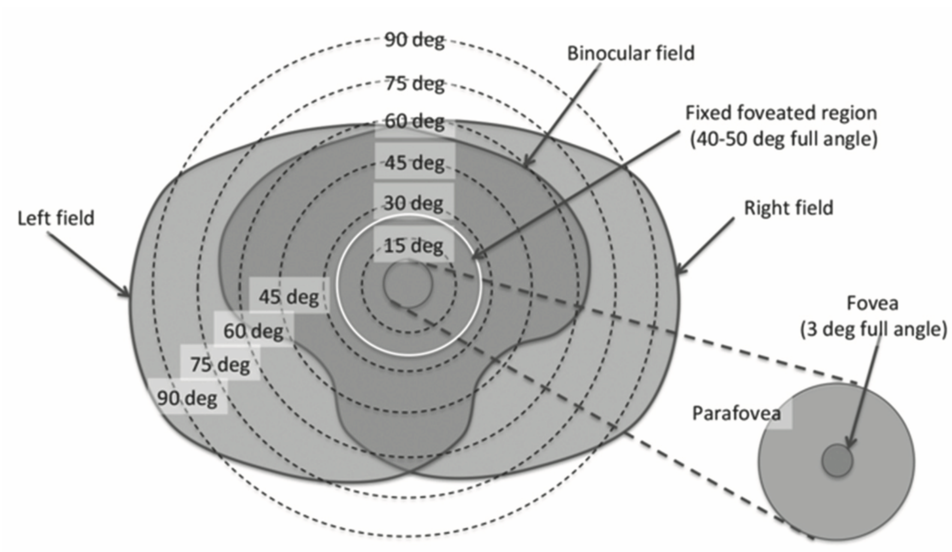


Figure 2.13: Binocular field of view. Kress (2020)

Though depicted in static terms, the FOV is dynamic due to continuous voluntary and involuntary eye motions that balance our directed attention with general awareness while accounting for motion of the head, body, and environment.

Stereopsis and Depth Perception

Due to the separation of binocular vision, a slightly different view of the world is observed in each eye. In a process called stereopsis, the brain processes these disparities to form a single percept with a sense of depth and three-dimensional structure. In a related process called vergence, a variety of depth cues trigger the inward (convergence) or outward (divergence) rotation of the eyes to effectively regulate binocular vision. When vergence occurs, it triggers the natural focusing reflex known as accommodation.

Other than the binocular disparities described above, the strongest triggers for the vergence-accommodation reflex are occlusion and motion parallax. Occlusion occurs when nearby opaque objects naturally obscure more distant objects. Motion parallax is the phenomena where an object in motion appears to move at different rates based on its depth in the scene.

2.6.2 The Somatosensory System

The somatosensory system is a part of the sensory nervous system responsible for the perception of touch, temperature, body position, balance, and pain. It is a network of sensory receptors and neurons spread throughout the body and brain. Within this system, proprioception and balance, which enable our awareness of the body's dynamic and kinematic state, are most relevant to the design and use of HMDs.

Proprioception

Proprioception is the egocentric sense of movement, force, and body position. Through largely subconscious processes it provides the feedback mechanism necessary for effective coordination, refinement, and regulation of body motions. Specialized neurons distributed throughout the musculoskeletal system sense joint extension and limb position, velocity, and resistance. Signals from those proprioceptors are integrated with information from the visual and vestibular systems to create a sense of the body's overall state, enabling fast and unconscious execution of planned and reflexive behaviors. Proprioception is essential to both voluntary and involuntary motor control activities. It drives the continuous adjustment of body posture required to maintain balance and is a critical contributor to the process of learning and perfecting motor skills.

Balance

Equilibrioception is the sense of balance and spatial orientation. It is the integrated perception of stimuli from the visual, proprioceptive, and vestibular systems. Two organs of the inner ear comprise the vestibular system: semicircular canals and otolith organs. Three semicircular canals located in the labyrinth of each ear sense rotation around their orthogonal axes. Movement of fluid in the canals is sensed as pressure changes, which are signaled to the brain. In the otolith organs, signals from hair cells are triggered by head motion. Those signals are interpreted by the brain to distinguish head tilt from body motion and sense the lateral and vertical components of acceleration.

Rotational and translational stimuli from the vestibular system are used to control posture, as described above, and eye movement, via the vestibulo-ocular reflex (VOR). VOR helps stabilize gaze direction as the head moves by directing opposing eye movement to compensate. This limits retinal image slip by maintaining the visual point of interest in the center of the field of view.

2.7 Enabling Immersion

The immersiveness of an XR experience is limited by the ability of the hardware and software systems involved to create an illusion that is cohesive and undistracted. Understanding the human factors involved, as described above, can help achieve that. The following sections will explore the technical underpinnings of vividness and interactivity, the primary components of immersion.

2.7.1 Resolution and FOV

Resolution and FOV are key measures of the fidelity for visual display devices. For near-to-eye (NTE) displays found in HMDs, resolution is typically expressed in dots per degree (DPD), rather than dots per inch (DPI) or raw pixel counts, as in conventional displays. An angular resolution of 50 DPD (1.2 arc minute) roughly corresponds to the resolving power of 20/20 vision (Kiyokawa, 2015).

The FOV of an HMD includes the aided region, where real and virtual images are overlaid; the peripheral region, outside the aided region; and the occluded region, where vision is obscured by the device (Kiyokawa, 2015). FOV specification in HMD design must identify the angular span, aspect ratio, and location of the aided region within the user view. These decisions are interrelated and driven by task and market requirements (Kress, 2015). Figure 2.14 depicts the range of implementations found in state of the art XR HMDs, overlaid on the binocular FOV and the fixed foveated display region (Kress, 2020).

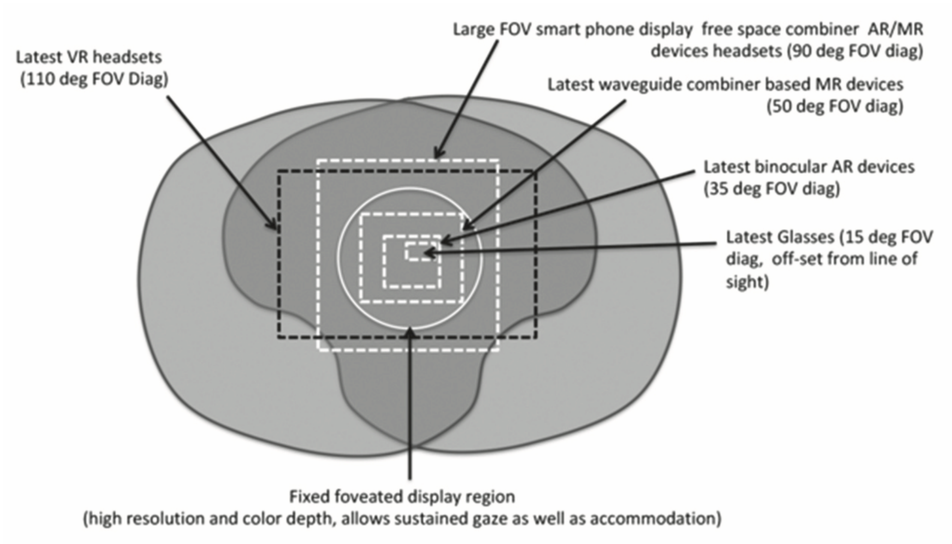


Figure 2.14: Typical FOVs for SOTA XR HMDs. Kress (2020)

Very high pixel counts are required for ideal resolution in wide FOV devices. For example, a 16:9 display with 50 DPD angular resolution and 160 deg horizontal FOV per eye, would require 8,000 x 4,500 pixels. Two such displays (one per eye) would have more than eight times pixel count of a modern 4k monitor (3,840 x 2,160).

Such devices will not soon be practical. Meanwhile, pixel doubling and other mitigating techniques can improve perceived resolution. Foveated displays offer an alternative that exploits the bi-modal nature of human vision. This emerging technique renders a high resolution region, positioned either statically, central to the field of view, or dynamically, based on eye tracking. This image is combined with a lower resolution peripheral display using digital or optical methods (Kress, 2020). AI-based methods also show promise (Kaplanyan et al., 2019).

2.7.2 Frame Rate and Latency

Frame rate is the number of times the rendered scene is updated per second. It can be different from the system update rate, which is the rate at which the display updates. Both are typically on the order of 30-120Hz, with most modern XR devices operating at 60-90Hz.

High frame rates increase the smoothness of motion, approaching the continuous nature of real world visuals. Update rate is a fixed property of the display hardware, but frame rate depends on the scene complexity and visual fidelity, along with hardware and software performance. The inverse of frame rate, rendering time, contributes to overall system latency. Tradeoffs must be made in the design and implementation of XR experiences to achieve the desired visual performance and limit system lag (Jerald, 2016).

Latency is the lag between head motion and update of the rendered scene, resulting in discrepancies between the user's visual and vestibular senses. In optical see-through systems this results in registration error, which leads to confusion, disorientation, and motion sickness. To compensate, head motion prediction and other methods are used (Kiyokawa, 2015). Specifically, motion-to-photon (MTP) latency of no more than 20ms, and ideally less than 10ms is recommended in the literature (Albert et al., 2017). Because MTP latency greater than 20ms is a key factor in motion sickness, this is a foundational requirement of HMD design (UL, 2022). Approaching this goal compels optimization of the entire pipeline, including custom silicon designs for the sensor fusion process.

2.7.3 Pictorial Consistency and Visual Quality

Visual quality is an assessment of the visible stimuli produced by an XR device. It is a qualitative measure of vividness, also described in the literature as realness, realism, or richness (Yim et al., 2017). Key contributors, including geometric resolution, scene complexity, and the quality of lighting and shading are limited by the frame rate and latency related considerations previously described. Visual quality is a critical performance measure for VR devices. In OST and VST AR/MR devices it is only one component of pictorial consistency.

Pictorial consistency refers to the degree with which virtual objects match their real world counterparts in an AR/MR display. Visual discontinuities introduced throughout the imaging pipeline reduce immersion and its attendant benefits in OST devices. The limited visual quality of virtual objects is further diminished by an incomplete understanding of scene depth and environmental conditions. When rendered, this creates additional lighting, shading, and depth related discontinuities in the real-world view (Fischer, 2015). Limitations in display and optical combiner technologies, particularly in their ability to mimic the brightness, contrast, and dynamic range of the real world compound this problem (Kress, 2020).

VST devices trade combiner related discontinuities for those introduced by the image acquisition and processing pipeline. Intrinsic parameters of the camera, including the lens properties, sensor characteristics, and camera settings (e.g., exposure time, ISO, and white balance), introduce noise, geometric distortion, motion blur, defocus blur, and color cast. Virtual objects rendered free of those distortions stand out as relatively crude but synthetically perfect elements of the scene. Methods to emulate camera distortions or stylize the entire scene can reduce this effect, but may not be suitable for all applications (Fischer, 2015).

2.7.4 Tracking Methods

Combining real and virtual scenes in a spatially coherent fashion is the essence of AR (Azuma & Bishop, 1994). See Figure 2.15. To maintain accurate "registration" (alignment) of the virtual and real world scenes in three dimensions, AR devices must determine their position and orientation in the world, or "pose" (You & Neumann, 2015). This process, known as tracking, typically uses methods from computer vision to estimate the pose of a camera based on features identified in its video stream.[14] In general, this process involves three steps: recognition, tracking, and pose estimation (Yang & Cheng, 2015). Once the camera's real world pose is aligned with the virtual coordinate system virtual objects can be rendered in the scene with appropriate scale, orientation, and placement.

Recognition identifies features in the 2D imagery and matches them to corresponding points in a database of 3D features. Typically, the database consists of image, model, or area feature types, which are described in greater detail below. Recognition and tracking are interrelated problems, where the former is used to initialize the latter, or reinitialize it when tracking performance degrades. Tracking updates the position of recognized features over time to reduce the computational costs associated with recognition (You & Neumann, 2015).

Camera pose estimation calculates the camera's transformation matrix based on the tracked features. It is achieved by solving the perspective-n-point (PnP) problem for 2D-3D pairs based on intrinsic camera parameters (e.g., focal length, aspect ratio, lens distortion). PnP is a fundamental computer vision problem with many modern applications. The details of PnP are beyond the scope of this work but the essence of the problem

[14] 3D registration for navigational purposes is commonly achieved using a combination of GPS related technologies, but the results are not sufficiently accurate for AR applications.

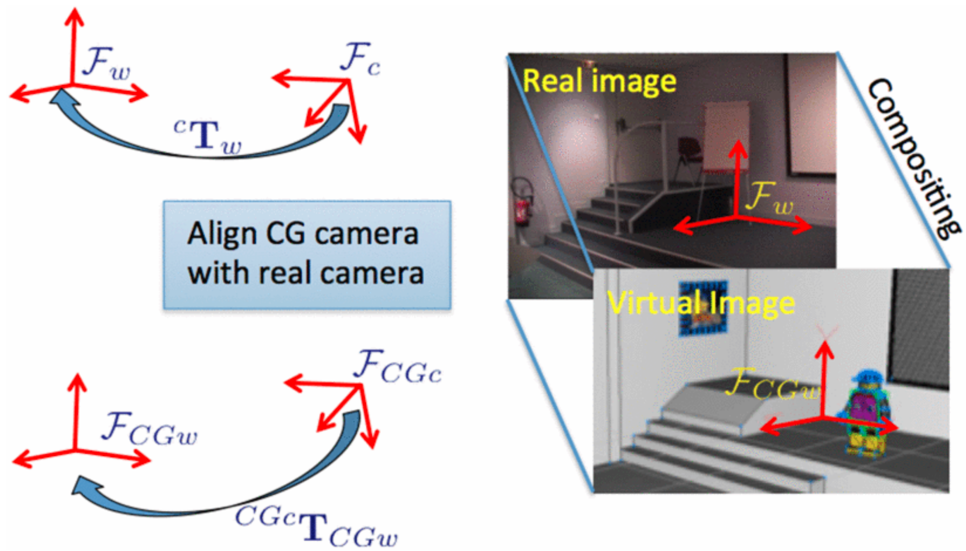


Figure 2.15: AR Tracking: spatial coherence is achieved by aligning real and virtual cameras with shared intrinsic properties. Marchand et al. (2016)

is captured in Figure 2.16. For more information, including a survey of implementations, see Marchand et al. (2016).

Tracking methods are typically characterized by the features used in the registration process. This is an active area of research where terminology and implementations vary, but image, model, and area feature types are common. Image based tracking relies on 2D pixel data. Model and area based methods use discrete and continuous objects of 3D geometry, respectively. Hybrid methods are also used.

Image-based methods use either photographic image data, graphic symbols called templates, or barcode-style marker designs. The feature database is created through an offline preprocess which identifies critical reference points in the image data and encodes them as vector representations. During recognition a similar process is used to encode reference points identified in the live imagery, which are then matched to the feature database using nearest neighbor methods. This process is resource intensive for arbitrary image and template data (Yang & Cheng, 2015).

Marker-based AR affords simplifying assumptions for the registration process with standard fiducial designs optimized for all stages of the tracking process. Black-and-white encoding patterns and clearly delineated boundaries aid recognition and tracking. The corners emphasized by square marker designs provide four coplanar, non-collinear points required for PnP pose detection (Yang & Cheng, 2015). DensoWave's Quick Response (QR)

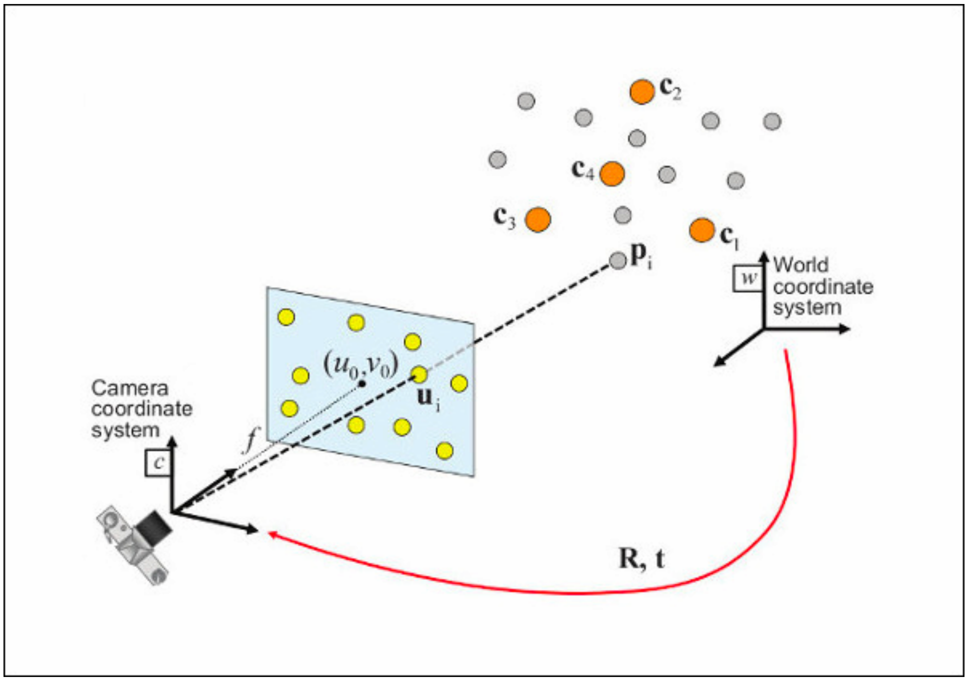


Figure 2.16: The pose estimation problem. "OpenCV Perspective-n-Point (PnP) Pose Computation" (n.d.)

codes store 2953 bytes of easily-decoded binary data and are widely used for AR applications (ISO, 2015). The marker-based process is depicted in Figure 2.17.

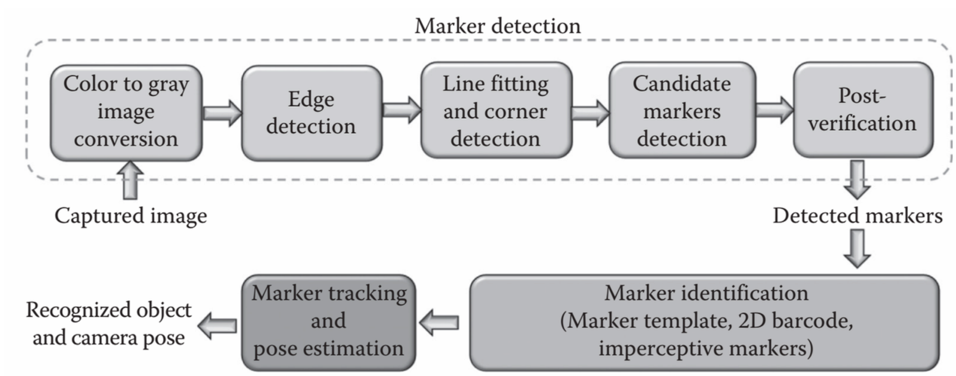


Figure 2.17: Marker-based tracking pipeline. Yang and Cheng (2015)

Marker-based optical tracking systems use two approaches. The "inside-out" approach places markers on the target object and camera pose is estimated from images of the marker in the observer-borne camera. In the "outside-in" case, markers are placed on the observer, who is localized by a set of static cameras surrounding the scene. The outside-in method, more commonly used in motion capture applications, requires an additional pre-calibration process that establishes the pose of the target object. Both cases require prior instrumentation of the scene with markers or cameras, the number and placement of which determines the trackable volume. This deployment process and the visually obtrusive nature of markers may be inappropriate in some applications (Gay-Bellile et al., 2015). Additionally, markers are often impractical in controlled and/or outdoor environments and are sensitive to occlusion (Ventura & Höllerer, 2015; Yang & Cheng, 2015). Together, these shortcomings compel the use of more sophisticated tracking methods.

Instead of image data, model based tracking relies on a 3D model of the target object for feature identification. The process is analogous to what is described above: key points encoded from the 3D model data are compared with features extracted from the live scene data and corresponding pairs are used for pose estimation. Live scene data can consist of imagery or 3D geometry generated from camera data using SFM (structure from motion) (Schonberger & Frahm, 2016) or SLAM (simultaneous localization and mapping) (Durrant-Whyte & Bailey, 2006) related methods (You & Neumann, 2015). Alternatively, active scanning systems using LiDAR (light detection and ranging) or TOF (time of flight) can be used to reconstruct scene geometry in real time (Behzadan et al., 2015).

The accuracy of model based tracking methods suffers when geometric or photometric details are not easily discerned. As a result, it is sensitive to lighting conditions (color, intensity, direction) and visibility (small in FOV, occluded, or outside DOF). Area based tracking uses SFM / SLAM to address those shortcomings by tracking a 3D model of the entire scene rather than discrete elements of it. This greatly increases the likelihood of achieving the confluence of 2D-3D matches required to achieve recognition (Gay-Bellile et al., 2015).

Vision based methods are often supported by incorporating additional sensor data to augment the tracking process. A complementary source of orientation and translation data can be derived from GPS data fused with signals from a trio of inertial measurement units (IMUs): accelerometer, gyroscope, and magnetometer (Ventura & Höllerer, 2015; Yang & Cheng, 2015).

Proper tracking is the essence of AR/MR devices and a critical element of pictorial consistency in both OST and VST devices. But the accurate placement of virtual objects in the real scene does not guarantee spatial coherence. Such objects must also appear naturally occluded.

2.7.5 3D Occlusion

As discussed in reference to Stereopsis and Depth Perception, occlusion occurs when objects nearer the viewer naturally obscure background objects. Real world scene depth is largely informed by our perception of this. Thus, proper occlusion of virtual objects in the real world is essential to the user's understanding and acceptance of a mixed reality scene, as well as their interaction with it.

The graphics pipeline and depth sensors of a modern XR HMD can provide the information required to enable per-pixel masking of virtual objects for accurate depth sorting. The believability of the combined scene is dependent on the resolution, dynamic range, and opacity of the virtual object. So-called "hard-edged occlusion," where virtual objects appear opaque and naturally occluded, with crisp edges, is the ideal. This requires blocking light from the scene at a pixel level, which is achievable on VST devices using traditional digital compositing methods (Kiyokawa, 2015).

OST devices rely on optical compositing techniques, with little control over the scene's natural dynamic range and displays unable to match its brightness. As a result, virtual objects on OST AR/MR devices have a ghostly, semi-transparent look. This may be suitable for overlays and other augmentations, but falls short of enabling a cohesive mix of real and virtual objects. Currently, few optical methods are capable of addressing this problem. Pixel dimming is often suggested as a compromise in OST HMDs. This method, also known as soft-edge occlusion, selectively dims areas of the real world to help virtual objects stand out (Kress, 2020).

Despite significant research and development efforts, occlusion remains an unsolved problem in OST devices. Few implementations of soft-edge occlusion exist in the market, and hard-edge solutions remain entirely absent. The details of the challenges involved are beyond the scope of this work, but are well summarized by Karl Guttag, a recognized expert in graphics processors and display systems. Guttag identifies technical and physical roadblocks for both approaches and declares a general solution to hard-edge occlusion "infinitely complex" for current optical architectures (Guttag, 2021).

2.8 Ensuring Comfort

Well-designed hardware and software can exhibit the fidelity, responsiveness, interactivity, and believability necessary to promote immersion through the user's overall sensory comfort. The previous section outlined many of the key technical considerations in doing so. In time, many of shortcomings identified will likely be overcome.

Meanwhile, comfort related considerations must help guide the necessary tradeoffs. For the effects of immersion to take hold, the experience must limit distractions due to wearable, social, vestibular, or visual discomfort.

2.8.1 Wearable Comfort

Wearable comfort refers to general ergonomic traits, including size, weight, and balance, as well as surface treatments and thermal management features. Overall usability and safety are also factors. For example, the safety and mobility benefits offered by a direct view of the environment and cable-free use motivated the HoloLens' untethered OST design (Kress & Cummings, 2017).

2.8.2 Social Comfort

Social comfort concerns are primarily related to privacy and acceptable public use. The suitability of a design's aesthetic and form factor is one consideration (Cook et al., 2019), as is allowing an unaltered view of the wearer's eyes. The number and packaging of outward-facing sensors, and the nature and use of the data they collect, entails a number of public privacy concerns that influence social comfort (Kress, 2020). Each of these balance the wearer's willingness and right to wear the device with the needs of the public, and are strongly influenced by the context and manner of intended use. Bass et al. (1997) describe the ultimate test of social comfort as "whether or not a wearer is able to gamble in a Las Vegas casino without challenge."

2.8.3 Vestibular Comfort

Due to the interrelated nature of the human visual and vestibular senses, it is difficult to clearly separate the relevant comfort issues. Here, vestibular comfort is primarily concerned with motion sickness. However induced, motion sickness is XR's most common and significant adverse health effect. In VR and VST AR devices the primary contributors are movement and visual effects.

The most widely accepted explanation for sickness caused by real or apparent motion attributes it to a mismatch of sensory inputs. In XR, visual and auditory stimuli are experienced through the HMD while the vestibular and proprioceptive signals are coming from body motion. When discrepancies occur, motion sickness can follow. Sensory mismatch in XR is commonly caused by latency or unnatural motion. When MTP latency is excessive, the perception of body motion and corresponding visual stimuli are not synchronized, leading to visual-vestibular mismatch. Unnatural motion is often implemented with the intention of improving the experience. For example, head bobbing or strafing motions commonly used to add dramatic or interactive effect in screen-based experiences can have unintended effects in XR. The negative health effects of latency partly motivated the push for high frame rates and sensor fusion optimizations common today. Intended unnatural motion is a content design issue easily addressed through best practices (Jerald, 2016).

Visually induced motion sickness (VIMS) is "a subcategory of motion sickness that specifically relates to nausea, oculomotor strain, and disorientation from the perception of motion while remaining still"(UL, 2022). Several characteristics of VR and VST AR HMD design directly contribute to VIMS, including optical design issues, the presence of motion artifacts, and tracking / sensor fusion issues, all of which contribute to scene instability.

Elevated levels of vergence-accommodation conflict (VAC) are known to cause discomfort and nausea in OST AR devices. Our visual reflexes naturally work together to look at (vergence) and focus on (accommodation) objects in the FOV. But most modern AR/MR devices use fixed focal length displays in which all virtual objects appear in focus at the same distance from the eye point, typically 2m. Virtual objects that occur at any other depth in the scene will lead to conflicting signals from the eyes' vergence and accommodation demands. When that occurs, depth and focus cannot be reconciled, leading to eye strain and disorientation (Kiyokawa, 2015). For example, mixed reality experiences that rely on arm's length interactions are focused on an area 30-70 cm from the user. This is well within the headset's fixed 2m focus, and often leads to VAC-induced discomfort (Kress, 2020). Extended VAC exposure can lead to visual adaption, temporarily decoupling vergence and accommodation. The reduction of depth perception that results can create a hazardous situation. As such, UL 8400 recommends that users avoid sensorimotor-demanding activities (e.g., taking the stairs, driving, bike riding) for 30 minutes after each session (UL, 2022).

2.8.4 Visual Comfort

The primary visual comfort related considerations are vision correction, eye box design, and the limits and parasitic effects of screen-based display technology.

HMD designers cannot ignore the fact that a large portion of the population have some form of vision impairment, yet the method and degree of corrective support varies. Depending on the device type and form factor, interchangeable lenses, adjustable focal length, or custom corrective lenses may be integrated. Correction is particularly important in OST HMDs, and many are designed to accommodate the wearer's prescription glasses. This has an impact on the eye box design.

Ideal optical system designs provide a clear, consistent, and unobstructed view of the entire FOV. A key contributor to that outcome is the size of the eye box: the volume of "3D space in which the viewer's pupil can be positioned to see the entire FOV" without a reduction in brightness or distortion near the extents (Kress, 2020). Eye box designs vary with user anthropometry (inter pupillary and temple to eye distances), system design (combiner thickness, optical architecture, and eye relief), and pupil size. Though mechanical adjustments may allow users to optimize a system's eye box for their static anthropometry, scene visibility will still vary with their pupil size. For example, the edges of the display may become blurry when the pupil dilates in bright conditions. The complexity of eye box design and ambiguities of the "easy viewing" requirement make this a challenging problem (Kress et al., 2014). Large eye box designs can improve visual and wearable comfort (fit), but at a cost to perceived brightness (luminance) due to physics based constraints (étendue).

Current hardware continues the trend of exploiting the latest advances in components designed for the screen-based smartphone and tablet markets, sometimes with little effect. In particular, flat panel display technologies used as immersive near to eye (NTE) displays are inherently limited by fixed focus, low brightness / contrast, and optical invariants including étendue. These pixel-based displays are also susceptible to a variety of parasitic effects. The screen-door effect appears when the optical quality, typically expressed in terms of MTF,[15] is high enough to see gaps between the pixels of the display device. Aliasing is the visible side-effect of representing continuous visual phenomena with discrete pixels. Where aliasing is the spatial artifact of sampling, motion blur is its temporal side effect. The Mura effect describes an unevenness of the display caused by imperfect illumination or screen

[15] Modulation transfer function (MTF) is a quantitative measure of the ability of an optical system to reproduce contrast detail. It is known to correlate with our perception of image quality. MTF is the magnitude of the optical transfer function. https://en.wikipedia.org/wiki/Optical_transfer_function

geometry. Each of these effects can be mitigated with hardware and/or software methods (Kress, 2020).

2.9 Design Tradeoffs

Due to directly conflicting requirements common in XR systems, there is no "one size fits all" solution. Amidst the hype surrounding this promising but immature technology, it is important to have an accurate understanding and realistic expectations. Numerous considerations important to the design, selection, and implementation of XR solutions are detailed above. Figure 2.18 assesses the importance of selected optical requirements in HMD devices across common market segments.

Specs	Smart Glasses	VR Headsets	Industrial HMDs	Defense HMDs
Industrial design	++++	+	–	–
Power consumption	++++	–	+++	–
Costs	+++	++	+	–
Weight/size	+++ (Forgettable)	+	++	+ (Helmet mounted)
Eye box	++++ (Minor mech. adjustments)	– (Dial in)	+++ (Minor mech. adjustments)	– (Dial in)
Rx glasses integration	+++ (Combo/ monolithic)	– (Dial in)	+++	– (NA)
Full color operation	+ (Mono- to full color)	++++ (Full color)	++ (Multicolor)	– (Mono-/ multicolor)
FOV	– (≤15°)	++++ (>90°)	++ (>30°)	++++ (>100°)
System contrast	+ (≥100.1)	++ (Occlusion)	++++ (>500.1)	++++ (>500.1)
Environmental stability	++	–	++++	+++
See-through quality	++	(Occlusion display)	+++	+++
Mono-/binocular	Monocular	Binocular 3D	Monocular	Binocular 2D

Figure 2.18: Requirements from various HMD market segments. Number of '+/-' markers indicates magnitude of positive/negative criticality. Kress (2015)

Tradeoffs should be informed by requirements specific to the context, manner, and goals of intended use, prioritizing human factors related to perception. Aligning the desired outcomes with the primary affordance of the chosen device is essential. These choices are aided by an understanding the theoretical underpinnings of those affordances.

2.10 AR/MR Potential for Industrial Training Applications

While XR has yet to provide consumers with a value proposition that is broadly compelling and sustainable,[16] the industrial, healthcare, and military markets have embraced its potential for cost savings and competitive advantage.

Across industry, preliminary studies have shown that AR's essential connection to reality (e.g., guiding a surgeon's hand) can have a variety of benefits. AR has improved learning rates, reduced errors, increased yields, improved quality, and enhanced designs. By enabling collaborative design, remote expert guidance, and enhanced monitoring, it has also improved the end-user experience (Azuma, 2019). In industrial settings HMDs are typically designed to support operators in an unobtrusive fashion, allowing them to focus on a task in the physical world, e.g., inspection, maintenance, repair, and order picking. In doing so they can reduce cognitive and/or physical load thru supplemental hands-free displays (Starner, 2015). Manufacturing, healthcare, and defense are three industries that have invested heavily in the early development of this technology.

2.10.1 Applications in Manufacturing

In manufacturing, supporting operators and repair technicians with digital work instructions has been a common application of AR research and development since the early 1990s (Azuma & Bishop, 1994). AR is a core component of I4.0 that allows intuitive, real-time access to contextually appropriate information. It provides the ideal visual interface for collaborative problem solving as described in Grieves' original vision for digital twins (Grieves, 2015). Due to the many benefits described above, a 2020 study documented applications in operations, maintenance, quality control, safety management, design, visualization, logistics, and marketing (Oztemel & Gursev, 2020). XR shows particular promise as a source of innovative tools and technologies for training workers at a time when finding skilled labor is increasingly difficult due to high retirement rates, global expansion, and increasing specialization (Kress, 2020).

A compelling case study for the use of AR in manufacturing comes from the automotive sector, where its benefits can be leveraged across the entire product life cycle (Gay-Bellile et al.,

[16] This claim may well be tested in 2024, with the recent introduction of Apple's Vision Pro, a state of the art video pass-through device, and Meta's Quest 3, which is positioned primarily as a VR device, but also offers video-pass through.

2015). During design it reduces the need for and cost of physical mock-ups. AR complements the traditional design process, enhancing physical prototypes with virtual elements that can be accurately evaluated in real-world context. Prior to production, AR can reduce the impact and cost of factory planning. Operators can evaluate simulated workspace changes virtually integrated into the real environment without disturbing production or requiring a complete and accurate 3D model of the existing environment. During production AR can benefit tasks related to assembly, picking, and quality control by delivering instructions naturally, in the ideal context (what, when, where, how needed). AR allows the operator to remain focused on the work area while augmenting their perception with relevant data from sensors and/or information systems. Sales efforts benefit from AR's ability to communicate aspects of the vehicle that are not otherwise observable, (e.g., performance characteristics), or demonstrate unaccessible features (e.g., options for models not in inventory). Unlike other methods, the AR based approach retains accurate perception of dimensions, volumes, and other cues that are subtle but important to human perception. During the operation phase, AR can enhance the driver experience in many ways by visualizing system characteristics, highlighting potential dangers, aiding perception in degraded conditions, and augmenting instructional materials and support services.

2.10.2 Applications in Healthcare

In healthcare, XR has therapeutic and educational applications ranging from pain management and the diagnosis of mental disorders to medical decision-making and surgical support (Aqlan & Hui, 2020). AR in various forms has been adopted by the medical industry to improve patient / procedure outcomes and safety while reducing radiation exposure, recovery time, and costs. It is a well-suited complement to the trend towards minimally invasive procedures, where access and vision are limited (Yaniv & Linte, 2015). In those cases, AR eliminates the need for surgeons to map preoperative data to the patient from an adjacent monitor. It allows direct cognition of the operator's movements relative to patient anatomy. This form of image-guided surgery is achieved by tracking the surgical instruments and visualizing them over preoperative data registered to the patient. The challenge, time, and error of these procedures are less than screen-based alternatives that require filtered cognition (Kersten-Oertel et al., 2015). Medical necessity will likely drive the adoption of wearable devices that include AR functionality. Conditions like diabetes and macular degeneration can be monitored and/or improved with such devices. Eye worn sensors are being developed to address both of these medical necessities by improving the user's perception (Barfield, 2015).

2.10.3 Applications in Defense

In the military market, where many of these technologies were first proven, there remains a strong and growing demand for custom XR hardware solutions. In addition to the traditional HMD / Heads-Up-Display (HUD) systems common in fixed and rotary wing aircraft, there are efforts underway to outfit service members with AR devices that support their mission (Kress, 2020). The US Army's IVAS (Integrated Visual Augmentation System) is the most ambitious current example. Originally awarded in March of 2021, IVAS is a $22b partnership with Microsoft to improve "Soldier sensing, decision making, target acquisition, and target engagement" ("PEO Soldier PM IVAS", n.d.). While these custom hardware solutions provide further evidence of XR's adoption and future, educational applications for XR in the military are more relevant to this work. The wide range of applications include training and briefing support for pilots (Alexander et al., 2019), maintainers, leaders (Clayton & Straub, 2020), and officers (Millican, 2017).

2.10.4 Key Benefits for Industrial Training

The reviewed literature underscores the significant potential of AR/MR technologies in industrial training applications, particularly in manufacturing, healthcare, and defense sectors. Despite the gradual pace of consumer adoption, these industries recognize the potential advantages of leveraging AR/MR to enhance operations, reduce costs, and improve workforce development.

The key benefits of AR/MR in industrial training contexts stem from its ability to provide contextually relevant, spatially registered information and instructions integrated with the real-world environment. This promises to enhance learning, reduce errors, improve task performance, and ultimately contribute to a more skilled and efficient workforce.

2.11 Theoretical Basis

Before we proceed, it is essential to differentiate AR and MR from VR and establish the theoretical basis of their advantages for learning and retention.

2.11.1 Differentiating AR from VR

As we saw in Section 2.5, the XR devices lie along a continuum of user experience. For the purposes of this discussion, we will consider MR a subset of AR, with the added ability to manipulate virtual objects within the real-world scene.

AR and VR devices are often confused with one another and/or mistaken as new means for the consumption of traditional content. Both are head-mounted devices that display believable sensory stimuli to augment or reproduce real-world interactions. Both do so in a manner that is contextually cohesive and responsive to a wide range of body-centered inputs. Despite their commonalities, AR and VR are fundamentally different from one another and other modern media. Where VR is designed to immerse the user in a synthetic world, AR is intended to strengthen the user's connections with reality. Failing to recognize and leverage their unique affordances severely limits the utility of these devices (Leonard & Fitzgerald, 2018).

VR provides a form of interactive sensorimotor simulation that, when immersive enough to enable presence, the brain interprets as a lived experience. This enables situated learning experiences which, if designed to be appropriately challenging and/or visceral, can be enhanced by flow and may elicit an emotional response (Kappes & Morewedge, 2016; Kwon, 2018; Millican, 2017). The learning effect of a VR experience is thus largely grounded in the theoretical requirements and benefits of immersion & presence, experiential learning, and flow theory.

As an active learning method[17], VR is best suited for the development of higher-order cognitive skills. The potential for emotional impact also makes VR a useful tool for affective learning. Because the experiences are simulated, VR enables training that is otherwise impractical or impossible. Finally, the digital nature of VR experiences makes them easy to repeat, instrument, scale, and distribute. These practical benefits are accurately summarized as offering "experience on demand" in Jeremy Bailenson's[18] popular book of the same name (Bailenson, 2018).

AR allows augmentation of the real world with virtual objects that are informative and/or interactive, thus enhancing our understanding of and connection with the world. The es-

[17] Active learning and other educational theories mentioned in this section will be detailed in the *THEORY SECTION*

[18] Jeremy Bailenson is a prominent figure in the field of VR and its applications, particularly in education and behavioral change. As the founding director of Stanford University's Virtual Human Interaction Lab, his work focuses on how VR can affect users' cognition, behavior, and social interactions.

sential affordance of AR is direct interaction with virtual objects in which visual and spatial queries take the form of natural object manipulation in everyday surroundings. Applying embodied cognition and animate vision theories in the context of learning suggests that, by retaining proprioception and sensorimotor function, AR experiences are more aligned with human cognitive architecture than metaphorical digital interfaces. AR interfaces provide a combination of procedural and configurational spatial knowledge via haptic and pictorial sources. Visual, spatial, and sensorimotor feedback provides multiple reference frames that enhance perception and cognition. By reducing the overall cognitive load or better distributing it across multiple sensory pathways, AR improves the uptake of sensorial-based knowledge (Shelton & Hedley, 2003).

AR is also an active learning method best suited for higher-order cognitive development. Its affordances are well-suited for task-related learning because of the inherent connections between visual perceptual activity and physical movement. These effects are enhanced by untethered, hands-free OST HMDs which improve mobility and enable unencumbered use. AR facilitates local collaboration and remote assistance. Where VR excels at delivering discrete packages of simulated experience, AR is best applied to the continuous enhancement of action in the real world (Leonard & Fitzgerald, 2018).

Neither AR nor VR have proved more effective than traditional classroom methods for the recall-oriented learning outcomes found low in Bloom's cognitive domain, including remembering, understanding, or applying. However, both demonstrated other benefits in line with theory. VR users perform better on high-order questions related to analyzing, evaluating, and creating. It is also known to improve student attitudes, including engagement and self-efficacy (Cook et al., 2019; Kwon, 2018). AR users demonstrate improved perception, performance, and understanding of spatial concepts, with student outcomes correlated to physical engagement with the content. The psychological benefits of AR include reduced test anxiety and increased self-efficacy (Chen et al., 2019; Shelton & Hedley, 2003). These benefits have broad industrial and military applications.

2.11.2 Theories of Learning and Cognition

The perceptual, cognitive, and learning benefits of XR devices are generally attributed to theories rooted in experiential and constructivist learning, as well as related cognitive theories, all integral to the concept of active learning. These theories collectively emphasize the importance of direct experience, active engagement, and integrating all human faculties in the learning process. By applying these principles, XR devices are posited to optimize learning outcomes, assuming other factors are conducive. The following section will delve deeper into these theoretical frameworks, explaining their relevance and application in the context of XR-enhanced learning.

Active Learning Theories

Active learning theories (ALT), particularly constructivism and experiential learning theory (ELT), describe the relationship between situated experiences and educational outcomes, where the self-directed construction of new knowledge occurs through activity in a supportive environment (Clayton, 2017). Fundamentally, these ideas have epistemological origins in empiricism, rationalism, and pragmatism, which consider the role of experience, reason, and action in knowledge.

The idea of learning by doing is ancient, but the origins of modern ELT are usually attributed to John Dewey and his 1938 work, *Experience and Education* (Dewey, 1938). Jean Piaget's theory of cognitive development later introduced the idea of constructivist learning theory (CONLT), wherein learners build new understanding through the interaction of prior knowledge and experience (Piaget, 1928). Russian psychologist Lev Vygotskii's "Zone of Proximal Development" (ZPD) emphasized the learner's need for knowledgeable support, along with the social aspects of constructivist learning (Vygotskiĭ & Kozulin, 1986). These ideas were expanded on by Jerome Bruner's theory of "instructional scaffolding." Bruner claimed that understanding is developed through carefully guided and supported learner experiences that build on their current knowledge (Bruner, 1960).

In 1984 David A. Kolb, a protégé of Bruner's, published his cycle of experiential learning, which identified four stages: concrete experience, reflective evaluation, abstract conceptualization, and adaptive experimentation (Kolb, 1984). Kolb's conceptual model incorporated elements from previous theories and is widely used to operationalize ELT concepts today. Later, Lave and Wegner's Situated Learning Theory emphasized the contextual aspects of ELT. They claimed that an environment relevant to the subject matter helped situate the learner's mind, strengthening the experience and thus the learning effect (Lave & Wenger, 1991).

Active learning theories are grounded in andragogy and its methods, as espoused by Malcolm Knowles (Knowles, 1970). Where andragogy emphasizes the self-directed methods described above, pedagogy is primarily concerned with the delivery of knowledge and skills by an instructor. Modern educational systems are commonly designed to maximize the uptake of content knowledge using the later approach (Leonard & Fitzgerald, 2018). Pedagogy is well suited to the developmental and intellectual needs of young learners focused on the cognitive domain of Bloom's *Taxonomy of Educational Objectives* (Bloom, 1956). The objectives in this domain, as revised in 2001, are: remember, understand, apply, analyze, evaluate, and create. The extended taxonomy also describes the domain of affective (emotional) development (Simpson, 1966). Where pedagogy excels at delivering content knowledge, andragogical methods better support "higher order" cognitive and affective learning. For example, andragogy is commonly employed in the development of 21st Century Skills, including critical thinking, innovation, collaboration, and problem solving (Millican, 2017).

Flow Theory

Focused activity can lead to a state of psychological absorption. This intuitive phenomena is known as 'flow,' a term coined by Mihály Csikszentmihályi[19] who described it as the "optimal experience" (Csikszentmihalyi, 1990). Flow is a cognitive and affective state in which individual attention and motivation feel in harmony with the situation. This leads to a period of absorbed productivity wherein the normal concern for our immediate needs abates. Most of us recognize this highly gratifying experience, which is colloquially known as being in the zone or groove. Many previous studies have established flow's positive influence on learning effects (Kwon, 2018).

Csikszentmihályi's work claims that activities leading to flow must have structure and direction, provide clear and immediate feedback, and balance perceived challenges and skills. These interrelated requirements enhance the sense of competence and self-efficacy, in a way that is highly engaging without creating anxiety (Csikszentmihalyi et al., 2014). The so-called "flow channel," in which challenge and skill are appropriately balanced for the individual, is similar in concept to Vygotskii's ZPD, as previously described.

Cognitive Load Theory

Cognitive Load Theory (COGLT) is a framework for instructional design that aims to optimize learning by managing the cognitive load placed on learners. It is based on the assump-

[19] Mihály Csikszentmihályi was a renowned Hungarian-American psychologist and researcher whose work has been influential in various fields, including psychology, education, and business. His last name is pronounced *me-high chick-sent-me-high*.

tion of a limited working memory and an unlimited long-term memory (Sweller et al., 1998). COGLT suggests that effective instructional material should direct cognitive resources towards relevant learning activities (Chandler & Sweller, 1991). It identifies three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load is determined by the nature of the material, while extraneous cognitive load is caused by poorly designed instructional materials (Sweller, 1994). Germane cognitive load, on the other hand, is the cognitive load that contributes to learning by promoting the construction and automation of schemas[20].

Like Flow Theory and the ZPD concept, COGLT can inform both active learning theories and pedagogical practices to optimize learning experiences. The former both deal with aligning the challenge level of learning activities with the learner's abilities to promote engagement and learning. Meanwhile, COGLT deals more directly with how the presentation of information affects memory and learning processes. Together, these theories provide a comprehensive framework for designing effective and engaging learning experiences.

Embodied Cognition

Theories related to embodied cognition (EC) are concerned with the role of mind-body relationship in cognitive processes, and how those processes are influenced by interaction with the environment. EC makes diverse claims, some of which are controversial. Fundamentally, it asserts that cognition and sensorimotor processing are deeply intertwined. "On-line" cognition, which occurs in the context of the real world, involves perception. In that case, the purpose of the mind is to guide responses in real-time, and interactive experimentation with the environment is often used to aid cognition. But much of human cognitive activity occurs "off-line," separate from the environment (e.g., planning, analysis). In those times, cognitive processes are often informed by simulations of sensorimotor activity, including mental imagery, spatially-oriented mental models, and procedural memory. Thus, EC ultimately claims that perceptual and motor systems are not merely peripheral input and output services; they are essential components of an integrated mind-body process which is highly reliant on real or simulated interaction with the world (Wilson, 2002).

Mental practice is an instructive example of off-line cognition, defined as mentally rehearsing or "visualizing" a motor task in the absence of physical movement. These sensorimotor

[20] In learning theory, a schema is an organized pattern of thought or behavior that helps in processing, interpreting, and storing information in long-term memory. Schemas allow learners to categorize and assimilate new information efficiently by integrating it with existing knowledge.

simulations typically entail detailed mental representations of a specific real or hypothetical event. Compared to the corresponding physical experiences, they are shown to engage similar neural and conceptual systems and have corresponding effects on perception, cognition, motivation, and action. This form of mental simulation is known to be effective in a range of cognitive and physical skill-based tasks, including golf putting, rock climbing, piano playing, and surgery. The effects of mental practice appear to come from improved connections between action planning, movement, and proprioception, demonstrating that the brain responds similarly to imagined and real experiences (Kappes & Morewedge, 2016).

Spatial Cognition Theory

EC is related to spatial cognition theory (SCT), which describes the forms and sources of spatial concepts. Spatial knowledge, it claims, comes in three forms: procedural, declarative, and configurational. Procedural knowledge relates to navigating spaces or things. Simple facts about a space and the entities therein are the basis of declarative knowledge. Configurational knowledge concerns the relative positions and orientations between spatial entities, as well as their relationships. Likewise, three sources of spatial knowledge have been identified: haptic, pictorial, and transperceptual. Haptic knowledge is formed by touch or body movement. Visual information is the source of pictorial knowledge. Transperceptual knowledge is synthesized over time from multiple sources (Shelton & Hedley, 2003).

Animate Vision Theory

Though our language of human vision shares terms and ideas with cameras and photographs, the relationship is only analogous. A photo may resemble the mental image of what we perceive, but it is a shallow, incomplete representation of the experience (Greenwold, 2003). The operation of human vision is less like a camera than it is a computational imaging system with multiple sensory inputs and a brain-based CPU.

Animate vision theory (AVT) proposes that "vision is not the transformation of light signals into a representation of the enveloping 3D world, but … a tool used for sensory exploration of the environment," in which humans "sample a scene from the world in ways suited to their immediate needs" (Shelton & Hedley, 2003). Human vision involves physical and visually-related behaviors that iteratively construct a cognitive map of the environment. With each cycle, those mental representations guide movements and actions that redirect perception. New information acquired in each iteration is used to refine the cognitive map.

In this visuo-motor model, motor movement is essential to vision as it provides valuable information about the relative location of objects in the environment and the movement of the perceiver in relation to them (Clark, 1997).

2.11.3 Implicitions to Instructional Design for Augmented Training

This review emphasizes the importance of differentiating AR from VR when considering their application in learning and training contexts, particularly in manufacturing settings. While VR excels at delivering self-contained, emotionally engaging simulations, its fully immersive nature disconnects users from the real world, making it less suitable for supporting manufacturing operators who need to interact with physical tools, machines, and workpieces.

In contrast, AR's ability to enhance the user's connection with the real world aligns well with the demands of manufacturing tasks. These claims are supported by well-established theories of experiential and constructivist learning, including embodied, spatial, and visual cognition. By preserving the user's connection to the real world and leveraging natural perception-action couplings, AR is believed to align more closely with human cognitive architecture in ways that may enhance the acquisition of spatial and procedural knowledge. These affordances make AR particularly well-suited for enhancing real-world task performance and skill acquisition in manufacturing contexts, where operators need to navigate complex spatial arrangements, manipulate physical objects, and execute precise procedures.

Cognitive load theory and flow theory offer additional insights about balancing cognitive load and the level of challenge to enhance engagement and motivation. Ultimately, the practical and theoretical implications of these theories must be carefully considered during the instructional design of AR/MR-based training in order to meet the specific learning objectives and demands of the manufacturing industry.

Together, these theories inform a cohesive approach to instructional design for augmented training methods. As depicted in Figure 2.19, instructional design should be based on Active Learning Theories and informed by Cognitive Load Theory, while applying Educational Best Practices. Active learning theories are comprised of experiential and constructivist components, along with related theories of cognition, embodiment, and flow.

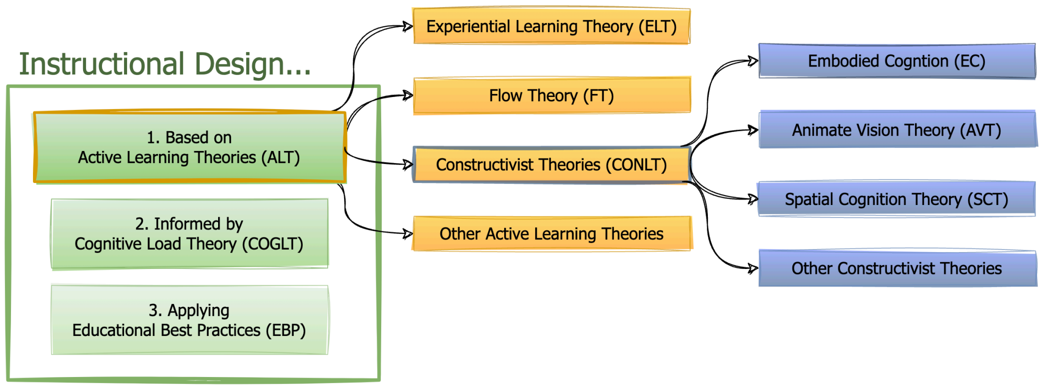


Figure 2.19: Instructional Design for Augmented Training Methods

2.12 Barriers to Adoption in Manufacturing

XR, particularly AR/MR, is still relatively immature. Despite promising results from pilot studies, widespread industry adoption of AR/MR for training requires clear justification in terms of return on investment (ROI) and measurable improvements in training outcomes. A number of other important technical, market, and social / legal obstacles must also be overcome (Azuma, 2019).

Doolani et al. (2020) conducted a comprehensive review of the current state-of-the-art in the use of XR technologies for manufacturing training. The review included 52 peer-reviewed articles published between 2001 and 2020, covering applications of VR, AR, and MR in various manufacturing training domains, such as maintenance, assembly, and human-robot collaboration. The authors found that XR technologies are effective in improving performance, reducing errors, and increasing engagement compared to traditional training methods. They also identified key benefits of using XR in manufacturing training, including enhanced safety, cost-efficiency, and scalability. However, the review highlights current barriers to XR adoption, such as hardware limitations and the need for further research on the application of AR in later phases of the manufacturing process. The authors conclude that XR technologies are powerful tools for manufacturing training, with each technology having unique capabilities and applications. They emphasize the need for future research to focus on developing interactive training interfaces and addressing the limitations of current XR systems to facilitate wider adoption in the manufacturing industry.

2.12.1 Measurable Improvement of Outcomes

In this section, we specifically focus on quantitative studies that evaluate the effectiveness of AR in enhancing instructional techniques. Our inclusion criteria are centered on case studies that require participants to learn and apply new cognitive and/or physical skills in practical, hands-on tasks within a manufacturing context. Such studies must also involve AR technologies that enable hands-free interaction. From an initial pool of 44 generally relevant studies, only 10 were found to align with these stringent criteria.

Upon closer review, two cases were later found less relevant than originally understood. Gonzalez-Franco et al. (2017) primarily assessed knowledge retention through fact-based quizzes, and not the acquisition of practical assembly skills. Wang et al. (2021) was designed to compare different instructional designs using the same AR device. Those studies were retained in the literature review but excluded from further consideration in the interpretations and conclusions that followed.

Tang et al. (2003) explores the comparative effectiveness of AR versus traditional and other computer-assisted instructional media in an assembly task utilizing LEGO Duplo blocks. In a carefully designed between-groups experiment involving 75 undergraduate students with no previous AR experience, participants performed an assembly task under one of four instructional conditions: traditional printed manual, computer-assisted instruction (CAI) on an LCD monitor and see-through HMD, and spatially registered AR instructions through an HMD. The assembly task, involving 56 procedural steps, was chosen for its generalizability to a wide range of assembly tasks across sectors. Key performance metrics included task completion time, error rate, and perceived mental workload, measured by the TLX. The authors discovered that spatially registered AR instruction significantly reduced assembly errors and decreased participants' mental effort compared to other media, highlighting AR's potential to offload cognitive processing. However, while AR outperformed the printed manual in completion time, it did not significantly outpace the other CAI conditions. The study underscores the risk of attention tunneling in AR, where users might become overly-reliant on its cues and become less aware of their physical surroundings. The authors suggest that AR systems should be carefully designed to balance those inputs.

Gonzalez-Franco et al. (2017) examines the effectiveness of MR against traditional training in manufacturing. As seen in Figure 2.20, the study uniquely employed an OST HMD setup to facilitate a face-to-face training where participants and instructors collaborated using a virtual model of an aircraft maintenance door. Twenty-four employees of the institution, without prior manufacturing knowledge, were recruited for this between-groups study. Knowledge retention tests and practical application assessments were used to determine the effectiveness and knowledge transfer. Analysis unexpectedly revealed that no significant differences were found in knowledge retention and interpretation scores between the MR and traditional methods. Task times did increase for MR training, attributed to the complexity of and user inexperience with HMDMR. The research highlights a unique capability of MR as equivalent training tool that can support, not replace some forms of face-to-face training in the future.

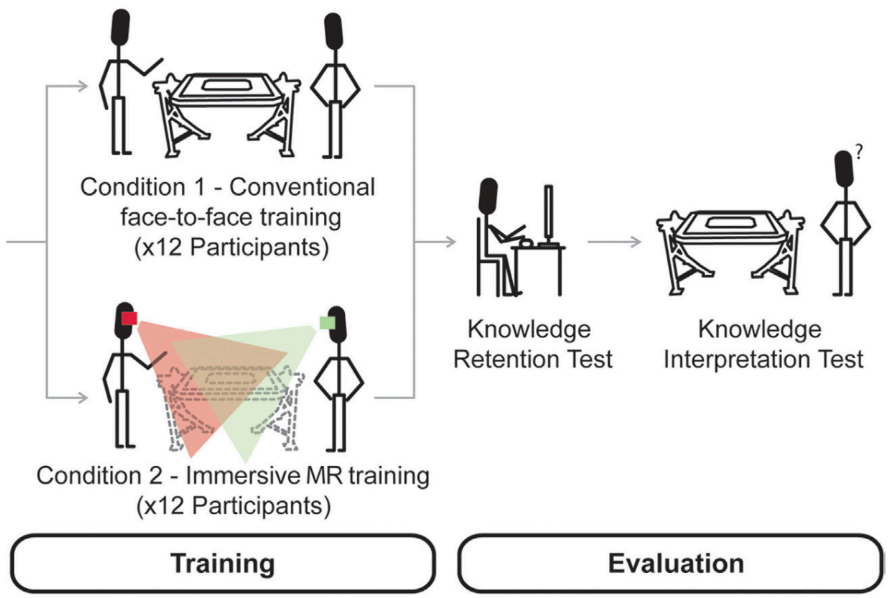


Figure 2.20: Experimental Setup from Gonzalez-Franco et al. (2017)

Chu et al. (2020) investigates the comparative effectiveness of instructional methods for assembling models of traditional Chinese architecture. The between-groups study recruited 48 engineering students to compare traditional paper instructions with a 3D viewer and an AR-assisted system. Each treatment was designed to include a progression of instructional affordances, as seen in Figure 2.21, based on validated paper-based instructions. Despite this, paper methods were associated with the most part-fetching errors, suggesting they lacked the necessary clarity. The AR system showed a trend towards reducing assembly errors and improved the accuracy of component placement, albeit at the expense of longer assembly times. Participants indicated a preference for the interactive features of AR, but a comparison of TLX responses showed no significant difference in perceived workload. The authors conclude that while AR has the potential to support complex manual assembly, the longer assembly times suggest areas for improvement in AR-assisted systems, such as reducing part confirmation time and addressing user fatigue. They also emphasize the importance of well-designed instructional content and user interaction methods in AR-assisted assembly systems, as these factors can significantly impact assembly performance and user experience.

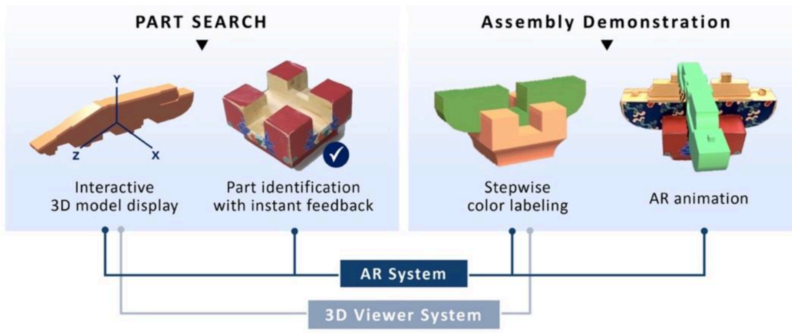


Figure 4. Two augmented reality (AR)-assisted systems.

Table 1. Assisted functions provided by three media.

Function	Paper-Based	3D Viewer	AR
Interactive 3D model display	no	yes	yes
Part identification with feedback	no	no	yes
Stepwise color labeling	yes	yes	yes
AR animation	no	no	yes

Figure 2.21: Progression of Instructional Affordances from Chu et al. (2020)

Büttner et al. (2020) investigates the efficacy of projection-based AR systems compared to personalized training and paper manuals for industrial assembly work training. The between-groups study simulated assembly tasks using a Fischertechnik construction kit. Training cycles, training time, error rates after 24 hours and 1 week, and quiz scores were tracked across 24 participants without prior AR experience. Personalized training outpaced both projection-based AR and traditional paper manuals in immediate learning efficiency. While AR systems somewhat improved training efficiency by preventing systematic mislearning through immediate feedback, they did not significantly outperform other methods in terms of training speed or long-term recall precision. The approach emphasized impact on the learning process—–training efficiency (rate of skill acquisition) and sustainability (recall and retention)—–over immediate task performance metrics like error rates and task completion time. The authors conclude that while projection-based AR can prevent mislearning, it does not offer significant benefits over paper manuals. They suggest exploring ways to incorporate aspects of personalized, adaptive training into AR systems to potentially improve training efficiency.

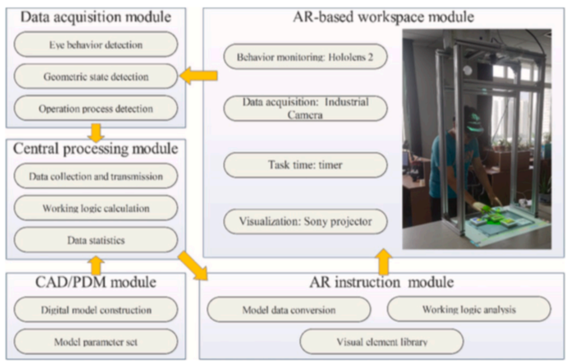
Hoover et al. (2020) examines the efficacy of using a first generation Microsoft HoloLens (HL1) for delivering AR guided assembly instructions against traditional and tablet-based digital instructions. Data for desktop and tablet model-based instructions, along with tablet AR conditions were drawn from prior studies. Participants in this between-groups study completed a mock aircraft wing assembly task in 46 steps. This task, created in partnership with the Boeing Company, was designed to reflect the complexity and variety of operations required in aircraft construction. The study found that HL1 AR instructions significantly improved task completion efficiency and accuracy, though floor effects make those accuracy findings less definitive. While outperforming non-AR instructions, HL1 AR led to significantly fewer errors than desktop MBI and tablet MBI but not tablet AR. User satisfaction measured by Net Promoter Score was lower for HL1 AR than tablet AR, attributed to comfort issues like the device being heavy and 3D tracking problems identified in qualitative feedback. The authors recommend using HL1 AR for complex assemblies with minor changes like toggling instructions on/off, and employing SUS for more rigorous user experience evaluation.

Vanneste et al. (2020) examines the comparative efficiency of projected AR, oral, and paper instructions in enhancing assembly operations, particularly for workers with cognitive or motor disabilities. In this within-groups study, various outcomes were measured, including productivity, quality, and help-seeking behavior. Stress was professionally observed and a modified version of the TLX was administered post-hoc. The findings reveal that AR instructions, specifically projection-based ones, significantly improved task quality by reducing error rates and aided operators in achieving better task comprehension and independence, as evident from reduced help-seeking behavior compared to oral instructions. However, AR did not outperform other media in terms of productivity or physical effort. The authors conclude that while AR has the potential to provide cognitive support by reducing perceived complexity and stress for novice learners, these advantages seem to diminish with repeated attempts as operators gain experience.

Havard et al. (2021) assesses the impact of AR against traditional PDF instructions on performing complex maintenance tasks within industrial settings, focusing on task complexity and operator competency. The authors claim novelty in their approach of separating out and measuring consultation duration as distinct from physical execution duration. In this between-groups study involving a 27-step drilling module maintenance task, measures like maintenance duration, consultation times, error rates, and satisfaction (TLX, SUS, feedback) were evaluated. The study found no significant differences in total maintenance duration between AR and PDF tablets for either competency group, regardless of if AR search time was included. Like other studies, it found that AR users were less prone to skip steps due to the direct feedback provided. However, AR was found to provide particular benefits over PDF for operation steps involving parts that are small, hidden or hard to locate, or easily confused, and steps requiring coordinated gestures. The study shows that AR acquisition and tracking delays account for a 34% increase in consultation times compared to PDF instructions, but the mean number of consultations was lower. The same delays impacted performance, especially for less experienced operators who faced greater usability issues and gave lower SUS ratings for AR, despite an overall "good" score. It did not find significant differences in mental workload between AR and PDF for either competency group. The authors conclude that, if tracking delays are overcome, AR exhibits promise for facilitating complex industrial tasks. In particular, they find it is well suited for frequently repeated or complex operations (due to accumulated consultation savings) and situations involving high operator turnover. This is especially true when the benefits previously enumerated can be leveraged, and operator competency is considering during deployment.

Kolla et al. (2021) explore the efficiency of AR against paper-based instructions. Participants in this study constructed a planetary gearbox using a variety of operations representative of a real manufacturing scenario. Both AR methods—HoloLens and a mobile device—notably reduced errors and improved system usability over traditional paper instructions, albeit without significantly affecting task completion times or workload. The authors underscore the critical role of thoughtful application design in AR's efficacy, highlighting how leveraging benefits like spatial mapping and speech recognition, while addressing limitations like occlusion and collision, contribute to smoother user interfaces and more positive task outcomes. The study's within-groups design with counterbalancing helps control for individual differences and learning effects. Participant responses to TLX and SUS surveys further confirmed the superior user experience offered by AR instructions. However, the authors suggest that further research with a larger sample size is needed to investigate task completion time and workload more conclusively. They recommend future work to validate AR's effectiveness in real assembly or training tasks within enterprises.

Wang et al. (2021) investigated the effectiveness of user-centered AR instruction in improving assembly performance and reducing cognitive workload compared to traditional 2D paper-based instruction. The study recruited 30 participants with an engineering background but no prior AR experience. Each were given the task of locating the centroid of a triangle, which they completed for both treatments. The crossover design of this study counterbalanced the order of conditions to help control for learning effect. As seen in Figure 2.22, AR instructions were delivered through a projected display system, while a HL2 was used to collect eye-tracking data. Assembly time, error rates, and NASA-TLX scores were also measured. Results showed significantly faster completion times, fewer errors, and lower cognitive workload for the AR condition. The authors conclude that augmented instruction, when designed to meet users' cognitive needs, enhances spatial understanding and task performance for novices.



(a) Projected AR System Configuration

(b) User's View of AR Operation

Figure 2.22: Experimental Setup from Wang et al. (2021)

Alves et al. (2022) investigate the efficacy of three AR methods—Mobile, Indirect, and Optical See-Through HMD—in supporting assembly tasks. Specifically, this study aims to address the lack of research using equivalent task designs to compare multiple AR methods and their relative advantages. The crossbalanced, within-groups study recruited 30 participants from the university community, each with varying exposure to AR assembly support. Participants were asked to prioritize accuracy and speed while constructed an 18-step LEGO Duplo assembly. Uniquely, they were given the choice to to either superimpose the virtual assembly or view it adjacent to the workpiece. Mobile AR was associated with significantly higher task completion times than both Indirect AR and HMD AR, while no significant difference was found between the latter two. Indirect AR, often overlooked, led to significantly fewer location errors compared to the other methods, and along with Mobile AR, was more prone to shape errors than HMD AR. Notably, the analysis focused heavily on workload evaluation, with Indirect AR demonstrating significantly lower mental and physical demand as measured by "raw" (unweighted) TLX scores. The study also found a significant difference in the error types most common to each treatment and a tendency of participants not to leverage beneficial affordances. The authors conclude that while all three methods were adequate, factors like price, comfort, usability, and control would determine the best fit for the application, highlighting the need to understand their relative advantages for the task and outcomes of interest. Specifically, they identify monitor-based

Indirect AR implementations as a very promising yet relatively unexplored option. Finally, despite stipulating that "Spatial AR" has been found to provide the best overall results, a lack of capable equipment prevented its inclusion in this study.

Summary of Study Results

The results of these studies are summarized in Table 2.1, including columns for Sample Size (SS) and AR/MR treatment type (AR), along with the primary results: Time, Errors (Err), Workload (Work), and Usability (Use). Where a study included more than one AR/MR treatment type, the one that best leverages the available affordances is listed. Cells for each of the four primary results denote the nature and significance of measured differences between the identified intervention and control (paper or digital work instructions). This approach maximizes the theoretical benefits, providing a "best case" interpretation of the results. For studies that involved two sessions (Büttner, Hoover), the outcome represents an approximate average of the findings.

The letter *P* is used to indicate a positive effect, where negative effects are indicated with an *N*. Asterisks indicate varying levels of significant effect, where one, two, and three stars correspond to increasing levels of statistical significance ($p < 0.05, p < 0.01$, and $p < 0.001$). Indicators without asterisks denote situations where a difference was reported without a test for significance. Dashes indicate no effect and empty cells were not measured.

Table 2.1: Summary of Results, AR/MR Case Studies

Paper	SS	AR	Time	Err	Work	Use
Tang et al. (2003)	75	HMD	P*	P*	P	
Gonzalez-Franco et al. (2017)	24	HMD	N*			
Chu et al. (2020)	48	Mobile	N**	P*	—	
Büttner et al. (2020)	24	Proj	—	—		
Hoover et al. (2020)	30	HMD	P**	P***		N
Vanneste et al. (2020)	40	Proj		P**		
Havard et al. (2021)	42	Mobile	—	P		P*
Kolla et al. (2021)	30	HMD	—	P*	—	P*
Wang et al. (2021)	30	Proj	P**		P*	
Alves et al. (2022)	30	HMD	P***	N	P*	

Most studies found that AR significantly reduced error rates compared to traditional instructional methods. However, Chu et al. (2020) noted that only part-fetching errors were significantly reduced in AR, while Büttner et al. (2020) noted that AR prevented mislearning, but found no significant improvement in short or medium-term recall. Alves et al. (2022) reported mixed results.

The impact of AR on task completion time was less consistent across studies. Some studies reported significant improvements, while others found increased times or no significant differences. Notably, Havard et al. (2021) found longer consultation times due to tracking delays but fewer overall consultations with AR, resulting in similar overall task times.

Several studies assessed cognitive workload using the NASA-TLX or modified versions, with many finding that AR significantly reduced workload compared to traditional methods. Tang et al. (2003) did not support that finding with pair-wise analysis, while neither Chu et al. (2020) nor Kolla et al. (2021) found significant differences in perceived workload.

Only three studies evaluated usability using standardized instruments, with mostly positive results. Hoover et al. (2020) found lower user satisfaction with AR compared to tablet-based instructions due to comfort and tracking issues, but the study did not report significance. Havard et al. (2021) and Kolla et al. (2021) reported improved usability with AR.

Havard et al. (2021) suggests that the benefits of AR may be more pronounced for complex tasks or in situations involving high operator turnover. However, Vanneste et al. (2020) found that the advantages of AR may diminish as operators gain experience with repeated task performance. Alves et al. (2022) noted that "indirect AR," as pictured in Figure 2.23, is a particularly promising and generally overlooked option.

This literature review demonstrates broad support for the preliminary findings previously discussed. These eight case studies, drawn from various domains and with a range of task types and complexity, provide empirical evidence that aligns with the promised improvements to learning transfer, accuracy, and performance compared to traditional instructional methods. However, that effectiveness is shown to depend on various factors such as task complexity, user experience, and application design. We will explore this claim further in the following section.

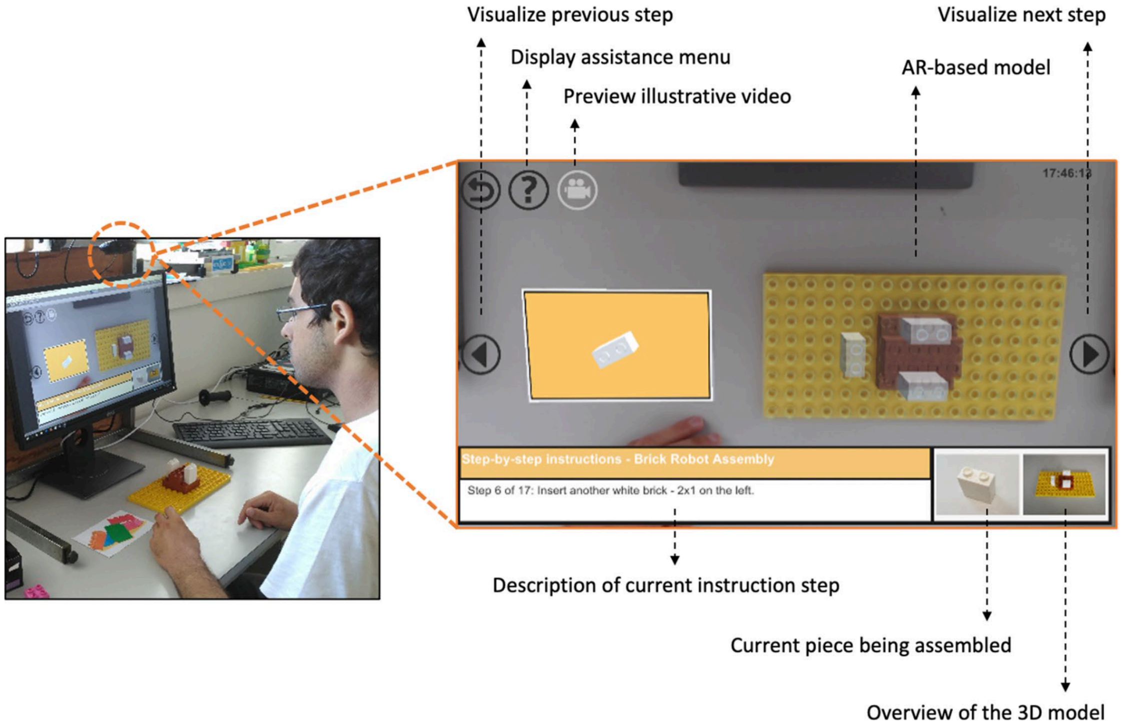


Figure 2.23: Indirect AR from Alves et al. (2022)

Summary of Study Designs

While the outcomes of these studies provide valuable insights, a comprehensive understanding of their collective significance requires a closer examination of their design and features, as summarized below. Table 2.2 includes columns for Relevance (Rel), and AR/MR treatment type (AR), as well as the instruments used for assessing Workload (Work) and Usability (Use).

Relevance is an overall measure of how closely the study's task resembles real-world assembly tasks, designed to facilitate the assessment of each study's ecological validity.[21] It was assigned based on the nature and complexity of the task design, using a standardized 5-point scale. Purely abstract tasks were given scores in the 1-3 range, LEGO assemblies 2-4, and realistic tasks 3-5. The final determination was based on the assigned range and relative complexity. Two studies were assigned an overall relevance of zero as they did not meet the criteria for inclusion, as described above.

[21] In the context of this review, ecological validity pertains to how well the study's task design mirrors authentic manufacturing assembly tasks in terms of complexity, tools, and environment. Studies with higher ecological validity would, therefore, be considered more relevant and informative for understanding the effectiveness of AR technologies in real-world industrial settings.

Table 2.2: Summary of Design, AR/MR Case Studies

Paper	Task	Rel	AR	Work	Use
Tang et al. (2003)	Abstract LEGO Assembly	3	HMD	TLX	
Gonzalez-Franco et al. (2017)	Aircraft Door Assembly	0	HMD		
Chu et al. (2020)	Architectural Model Assembly	3	Mobile	TLX	
Büttner et al. (2020)	Industrial Model Assembly	4	PAR		
Hoover et al. (2020)	Realistic Aircraft Wing Assembly	5	HMD		NPS
Vanneste et al. (2020)	Assembly & Quality Control Tasks	3	PAR	MTLX	
Havard et al. (2021)	Drill Maintenance Operation	5	Mobile	TLX	SUS
Kolla et al. (2021)	Realistic Gearbox Model Assembly	4	HMD	TLX	SUS
Wang et al. (2021)	Abstract Spatial Procedure	0	PAR	TLX	
Alves et al. (2022)	Simple LEGO Assemblies	2	HMD	RTLX	

The reviewed studies employed a wide range of task types, relevance, study designs, and AR/MR technologies. All studies assessed immediate learning effects, while only Büttner et al. (2020) assessed recall or retention. Workload was commonly measured using the TLX or variations thereof. In all but one case, usability was evaluated using the SUS. Hoover et al. (2020), after using the Net Promoter Score (NPS), noted plans to switch to SUS for future studies for improved rigor.

Most studies used paper instructions as the control condition, though two utilized digital equivalents. Some studies, such as Alves et al. (2022), compared multiple AR methods using equivalent task designs to assess their relative advantages. All but one (Büttner et al., 2020) measured task completion times. All studies measured error count, but only Chu et al. (2020) and Tang et al. (2003) measured error types. Chu et al. (2020) and Havard et al. (2021) broke down time by task step.

Several studies incorporated unique design features or methodological approaches. Büttner et al. (2020) focused on training efficiency and sustainability, using quizzes and training cycles as additional measures of knowledge capture. Havard et al. (2021) and Vanneste et al. (2020) were the only studies to measure consultation time, providing insights into help-seeking behavior and AR tracking delays. The latter's work included participants with cognitive or motor disabilities.

Kolla et al. (2021) and Chu et al. (2020) designed treatments with affordances in mind, emphasizing the importance of leveraging AR's unique capabilities. The latter employed deliberate instructional design with progressive affordances across treatments. Though it was otherwise excluded from this summary, Wang et al. (2021) demonstrated the benefits of instructional design for AR-assisted learning outcomes.

All studies employed either between-groups or within-groups designs. In order to help control for learning effect, all within-groups studies were counterbalanced via task ordering. All but Vanneste et al. (2020), Havard et al. (2021), and Kolla et al. (2021) employed a toolless task design to control for previous experience.

Finally, it is important to note that neither Chu et al. (2020) nor Hoover et al. (2020) were entirely hands-free designs. The former required some manipulation of the device and the latter utilized a wrist-mounted wireless button for input. Hoover et al. (2020) chose this over voice or gesture control of the HL2, which "are not always feasible in a factory environment."

2.12.2 Other Factors

Technical Considerations

Among technical limitations, general concerns about usability and immaturity are commonly noted (Leonard & Fitzgerald, 2018). Usability is primarily concerned with qualities of the application software, including user interface design, which are outside the scope of this work but obviously critical to the user experience and thus adoption. Technical immaturity relates to display fidelity (e.g., resolution, FOV, brightness, and contrast) and pictorial consistency. Of the latter, robust tracking is the most fundamental. AR devices must provide accurate, stable tracking in a variety of environmental conditions (Azuma, 2016;

Gay-Bellile et al., 2015). Related to tracking, and of particular concern to OST AR devices, is occlusion. Accurate compositing and occlusion require an understanding of the structure and illumination of the real world scene. These so-called scene semantics also allow for advanced interactions that build meaningful connections with the world (Azuma, 2017; Fischer, 2015). Finally, VAC mitigation techniques are necessary to eliminate it as a source of user discomfort in fixed focal length displays (Kress, 2020).

Over time, compounded incremental improvements promise to address many of the issues related to display fidelity and world tracking. Fast, accurate, universal eye tracking is premiering in the latest generation of XR devices, enabling other critical technologies. VAC mitigation methods that utilize gaze direction to perform discrete or continuous focus tuning should soon follow, along with foveated displays (Kress, 2020). Scene semantics is an active area of research in the deep learning community, and promising methods are emerging (Roberts & Paczan, 2021). Hard-edged occlusion in OST AR and multifocal displays still seem intractable with modern optical designs. Future advancements will likely rely on innovative methods, including light field and digital holographic displays that allow for layered or even per-pixel scene depth (Kress, 2020). Until then, tradeoffs guided by human factors and a deep understanding of customer needs will be required to deliver solutions with optimal product-market fit.

Market Considerations

Meanwhile, market considerations will limit adoption, even for XR systems that are "good enough" for today. Key among those are interoperability, standards, validation, metrics, organizational readiness, and access to content. Interoperability promotes open and/or standardized interfaces between systems. Commercial XR solutions are frequently built on stacks of interconnected technology that rely on other systems for data, etc. As such, interoperability is essential to the development of reliable, cost effective systems (Gay-Bellile et al., 2015). Interoperability depends heavily on the emergence of standards created to promote and enable it.

Here, standards is a broadly interpreted term. It includes publications from "standards bodies" like UL, ISO, and ANSI; similar publications from professional organizations; written frameworks that guide organizational processes and decision-making; and software frameworks, including APIs, libraries, or stacks that facilitate development. Together, these standards provide informational scaffolding, development support, tools, and even legal cover that many organizations need to reduce uncertainty and ease adoption. Relevant examples include the UL 8400 safety standard (UL, 2022), IEEE 1589 AR Learning Experience Model (IEEE, 2020), ETSI's Augmented Reality Framework ("ETSI Augmented Reality Framework", n.d.), and Microsoft's Mixed Reality Toolkit ("Microsoft Mixed Reality Toolkit", n.d.).

Validation and metrics both relate to demonstrating the claimed benefit of these systems. For industrial applications, adoption depends on quantifying the system value in terms of ROI and/or other metrics. Domain-specific modeling methods and evaluation metrics are needed to facilitate direct assessment and comparison of these systems (Kersten-Oertel et al., 2015). Organizational readiness is an overall assessment of a company's ability to adopt an XR solution. It includes considerations that are both cultural (e.g., leadership, attitude, risk tolerance) and practical (e.g., budget, goals, capacity) in nature (Cook et al., 2019). In part it is a measure of how well equipped the organization is to recognize and leverage the innovative benefits of XR, along with their willingness and ability to adapt to them (Leonard & Fitzgerald, 2018). The final market consideration is access to content. At this stage of adoption, most industrial XR systems will be custom applications, with few consumer-off-the-shelf (COTS) solutions. That said, software frameworks are available that enable low / no-code alternatives for common application types. Also, there is a growing network of specialized development studios and value-added resellers available for XR development.

Social and Legal Considerations

Important other social and legal barriers to XR adoption exist, most related to social comfort. Issues related to privacy and wearer's rights, censorship / disinformation / propaganda, driving with headsets, medical regulations, and related liabilities is a partial list of potential impediments (Barfield, 2015). While important to consider in the context of XR adoption, this category is outside the scope of this work.

In order to address many of the barriers identified above, a number of frameworks have been proposed to guide the specification, design, and implementation of XR solutions.

2.12.3 Gaps and Opportunities

The adoption of AR/MR for manufacturing support and training faces a number of important obstacles. Technical limitations, such as usability issues, display fidelity, tracking robustness, occlusion handling, and vergence-accommodation conflict mitigation, pose significant challenges. While ongoing research and incremental improvements are expected to address many of these concerns over time, tradeoffs guided by human factors and a deep understanding of customer needs will be necessary to deliver optimal solutions in the near term.

Market considerations, including interoperability, standards, validation, metrics, organizational readiness, and access to content, also play a crucial role in the adoption of AR/MR technologies. The development of open and standardized interfaces, along with the emergence of industry standards and frameworks, will be essential to promote cost-effective and reliable systems. Organizations must also be equipped to quantify the value of AR/MR solutions in terms of cost-benefit and other relevant metrics. Organizational and user readiness, encompassing both cultural and practical aspects, will determine a company's ability to recognize and leverage the innovative benefits of AR/MR technologies. Finally, important social and legal barriers must be addressed.

Fundamental to any industry adoption process is fact-based decision-making. To that end, the case studies reviewed showed AR/MR assisted instruction can help address the needs of manufacturing assembly training, but is not a one-size-fits-all technology. Its effectiveness varies with task complexity, user experience, the specific technology used, and other factors. The exact nature of those relationships is still not well understood.

Meanwhile, researchers should consider whether AR/MR needs to be "better" than traditional methods. This may seem counterintuitive in our age of high-tech wonders, but merely equivalent performance, when combined with other benefits, such as scaleability, cost-efficiency, repeatability, and safety, could be enough to drive adoption in the short term (Kaplan et al., 2021).

When examining the specific technologies used in these studies, HMDs stand out as particularly relevant for manufacturing assembly tasks due to their hands-free interaction methods, spatial registration, and unrestricted field of view. It is still essential to recognize the potential benefits of other AR technologies, such as mobile, projected, and indirect AR, as each has its own unique advantages and limitations. This is especially true as full-featured AR/MR headsets still suffer from technological limitations.

An important and related insight from these studies is the importance of well-designed instructions minimize AR's limitations while leveraging its affordances. Effective instructional design must consider user needs, abilities, and the context of the task. As discussed in Section 2.11.2, when done correctly, this leads to lower cognitive load, improved performance, and higher user satisfaction.

Despite the promising findings, there are notable gaps and limitations in the existing research. Most studies focus on immediate learning effects, with minimal coverage of long-term retention. The lack of industry recruitment in these studies may limit the ecological validity of their findings, as the tasks and settings may not fully represent real-world manufacturing contexts. Additionally, the highly abstract nature of some tasks (e.g., Tang et al. (2003)), may have hindered some participants' ability to form the mental models required for learning.

Measuring user satisfaction and usability through instruments like the NASA-TLX and SUS is crucial for assessing the quality of AR/MR implementations and guiding iterative improvements in tool development. By considering user feedback and needs, researchers and developers can create effective, engaging training solutions that address human problems and fit seamlessly into users' workflows. Adopting a human-centered design approach that incorporates user perspectives throughout the development process is essential for success.

Upon closer review, the heterogeneity of study designs emerges as a key concern. The wide variety of tasks, technologies, methodologies, and measures employed across these studies, while representative of the broader field, may hinder our ability to draw generalizable conclusions about the effectiveness of AR/MR in manufacturing assembly training. This issue is not unique to the domain and is identified in related studies.

Kaplan et al. (2021) conducted a meta-analysis comparing XR training's efficacy with traditional methods. Specific inclusion criteria were employed to ensure the validity and relevance of the included studies. Twenty-five studies were identified that quantified performance among adults after XR training for cognitive, physical, or mixed tasks. The analysis focused on learning transfer as a critical measure of the direct effect of training on real-world performance, and used a random-effects model to allow for direct comparison of results across diverse study designs. The authors concluded that the heterogeneity of study designs complicates the search for standardized efficacy metrics in XR training. They identified a need for more empirical studies and called for a unified methodological approach in those future explorations.

Further evidence of this gap is found in Moro et al.'s (2021) meta-analysis of VR/AR for anatomy and physiology knowledge acquisition, which found substantial unexplained heterogeneity ($I^2 = 72\%$) across eight studies. This suggests that the studies were not measuring the same effect, and makes it difficult to interpret the overall results. The source of this heterogeneity could not be identified by removing outliers or conducting a post hoc sensitivity analysis, and authors ultimately noted it as worthy of further exploration. Here again, it is most likely due to the small number of studies and their diverse designs.

Together, these studies support our interpretation that more uniform and rigorous study designs are required to identify key factors influencing the success of AR/MR interventions in real-world industrial contexts.

As a final note, these case studies span over two decades (2003-2022), during which time AR technology, instructional design, and manufacturing needs have all evolved significantly. This evolution may contribute to the improved results observed in more recent studies and highlights the importance of ongoing research to identify the key factors that influence the success of AR/MR interventions in real-world manufacturing contexts.

2.13 Tools for Development and Assessment

As shown in the previous section, the successful adoption and implementation of AR/MR technologies for manufacturing training requires careful consideration of various factors, including technical feasibility, user acceptance, organizational fit, and economic viability. Researchers and practitioners have developed a variety of methods, frameworks, and instruments to guide this process. These tools ensure that AR/MR solutions are aligned with the specific needs and requirements of the manufacturing domain, and are designed and implemented for users in a way that maximizes their effectiveness.

2.13.1 Development Methods

This section will discuss literature related to methods for specifying, designing, and implementing AR/MR systems for industrial training applications.

Specification

Palmarini et al. (2017) proposes a questionnaire based strategic decision making tool to guide the selection of AR systems for maintenance applications. The authors noted that these selections are challenging, with many considerations and a fragmented market of hardware and software solutions. The authors developed 30 questions, grouped into four questionnaires, based on the analysis of AR system characteristics described in related papers. Each questionnaire is designed to address the main hardware, software, and content choices involved, along with the overall suitability of an AR based solution. The author notes that the approach is not validated, does not address economic or ergonomic considerations, and does not generalize to other applications. Additionally, the resulting recommendations are general in nature and exclude VR and MR options.

Design

Borsci et al. (2015) describes the importance of alignment between training objectives, contents, method, and expected outcomes, along with the criteria used to evaluate those outcomes, in program design. This alignment is considered essential in the field of training assessment but usually overlooked in VR/AR studies. The authors found that experimental methods ignored important factors, did not employ standardized instruments, and failed to consider organizational or environmental needs. As a result, most studies are not reliable or generalizable. They concluded that a common framework is needed to address these issues in the design and assessment of XR training systems.

Taylor (2021) proposes a framework for adapting live training events to distance learning via immersive environments. Flow Driven Learning Experience Design (FLXD) integrates flow and transactional distance theories into Kolb's experiential learning model. FLXD describes how the designer can combine traditional and immersive learning methods in a way that best meets the unique needs in each of ELT's four stages. Taylor's work was designed to meet the needs of the large, diverse population of learners typical in military training programs.

Implementation

Longo et al. (2017) details SOPHOS-MS, a methodological framework and reference implementation for augmented operators in I4.0 based on Lee's 5C architecture. Their framework adopts a human-centered approach wherein the operator is essential to the optimal integration of real and virtual assets. By providing real time feedback, support, and access to the IT knowledge-base, SOPHOS-MS extends operator capabilities. This is accomplished via a verbal natural language interface using a variety of XR hardware. Their approach is suitable for a both on-line and off-line purposes, including training, collaboration, and support. Tests of this versatile implementation showed that operators trained with it outperform traditionally-trained counterparts throughout a two week period of use.

Geng et al. (2020) notes that industrial AR adoption is hindered, in part, by its reliance on custom software that is rarely reusable or flexible. The authors propose an adaptive no-code authoring system that allows end users to quickly customize and deploy ARWI (AR work instructions). The structure of their system enhances its adaptability to user needs, training environments, work processes, and system configurations. Its data driven design and form-based authoring tool are flexible, modular, and easily extensible. A collaborative implementation approach ensures that process requirements are accurately portrayed. Authoring tasks alternate between engineers and operators as each ARWI moves through four stages of development. Together, these features, and many more described therein, provide an agile alternative to rigid systems bottlenecked by their reliance on experienced developers.

Laviola et al. (2021) identified a lack of standards for the design of AR work instructions, without which choices are based on personal preference. This can lead to unnecessarily complex visuals that negatively impact cost and performance without improving the user experience. The authors proposed a standard process for AR work instruction design that conveys only the information required to accomplish a task, considering real objects involved, end-user needs, and task complexity. Experiments confirmed this “minimal AR” approach did not degrade any measured variable of user performance for various levels of task complexity.

2.13.2 Assessment Methods

Here, we review literature related to the assessment of XR systems, including well-known frameworks and popular instruments.

Frameworks

Kersten-Oertel et al. (2015) described their DVV Taxonomy for describing AR image-guided surgical systems, and proposed a framework for their assessment. DVV is an acronym of the three components identified in the taxonomy: data, visualization processing, and view. Those components, their classes and subclasses, and the relationships between them are considered at each step of the surgical scenario. The framework assesses image-guided surgical systems based on technical parameters, reliability, surgical performance, patient outcomes, economics, and social / legal / ethical aspects of use. Each component is evaluated in terms of the primary components of the operating room environment - surgeon, patient, and AR system - and the relationships between them.

Jetter et al. (2018) identified key performance indicators (KPIs) that influence user acceptance of AR for industrial applications. From a list of 16 candidate KPIs identified in a structured literature review and semi-structured expert interviews, the authors identified reduction of time and errors, spatial representation of contextual information, cognitive workload, and ease of use as the most predominant and suitable factors to study. Hypothesizing that the perceived usefulness of AR is influenced by those factors, a theoretical framework based on the Technology Acceptance Model (TAM) was developed to evaluate their effects on user attitudes and intentions. Their qualitative study found that all four KPIs had a positive role in users' perceived usefulness of AR, and thus their attitude towards and intent to use it. Despite that positive outcome, they also found that users are not yet convinced of AR's benefits, suggesting the importance of clear and convincing use cases.

Masood and Egger (2019) identify factors that influence the success of industrial AR using a research model based on the Technology, Organization, and Environment (TOE) for the adoption and implementation of innovation. Where implementation success (IS) is often measured in the literature by measures of worker performance improvement, here it refers to the benefits received by the company and their willingness to make further investments. Quantitative analysis found technological considerations, including system configuration along with technology hardware readiness and compatibility, and organizational fit had the most impact on IS. Their study also included a qualitative survey, which identified important challenges to IS. Together, these results provide a valuable, cohesive, and holistic depiction of success factors.

The following year Masood and Egger (2020) extended their prior research with 22 experiments conducted in an industrial setting and designed to identify challenges and success factors for IAR adoption. Using a combination of quantitative and qualitative analysis, the authors found that user acceptance, system stability, and organizational fit were the primary factors for success. Likewise, user rejection, system incompatibilities, technical maturity, and content creation issues were the main challenges. These findings can help guide strategic planning and requirements development for new IAR initiatives. In addition, the study gathered diverse industry feedback related to each context of the TOE model. A key implication of this study is that the relative importance of technological and organizational considerations vary, where the latter are more relevant in industry.

Danielsson et al. (2020) developed and applied a framework to assess the state of AR for industrial assembly applications. From a manufacturing engineering perspective, the authors considered authoring, infrastructure, and validation. Technical maturity concerns focused on the Technical Readiness Levels (TRLs) of available devices. Key requirements and enabling technologies were described. From both perspectives, AR is rapidly improving but still only suitable for limited usage. The authors identified a need for strategic decision-making guidelines for the integration of these systems. Such guidelines should need to be validated and account for economic considerations.

Instruments

Witmer and Singer's (1998) Presence Questionnaire consists of 32 items and measures the degree of presence experienced in a virtual environment. The same publication describes the Immersive Tendencies Questionnaire which measures the tendency of an individual towards immersion with 29 items. Both instruments use a seven-point scale where the endpoints are anchored by opposing descriptors (e.g., not compelling / very compelling).

The Flow State Scale by Jackson and Marsh (1996) is a 36 item instrument used to measure nine dimensions of the flow state described by Csikszentmihályi. It uses a 5-point Likert-type scale anchored with strongly disagree / strongly agree descriptors.

Kennedy et al. (1993) derived the Simulator Sickness Questionnaire from a prior instrument intended to measure real-world motion sickness. Differences in the origin, type, and severity of simulator sickness symptoms demanded it. Users self-report the presence of 16 symptoms ranging from general discomfort to nausea. Each symptom is measured on a scale of none, slight, moderate, severe. Three principal factors of this instrument are interpreted as clusters of oculomotor, disorientation, and nausea symptoms.

Hart's NASA Task Load Index (TLX, 2006) has been used to estimate workload for almost 40 years. It assesses overall task workload based on the magnitude of mental, physical, and temporal demands imposed by the task, the operator's emotional response to those demands (effort, frustration), and their perceived ability to meet them (performance). These six factors are weighted based on the factors each subject feels best describe the workload associated with the task under study.

The System Usability Scale (SUS) was designed by John Brooke (1996) to provide a "quick and dirty" assessment of usability for industrial systems, where detailed analysis is often expensive and impractical. The design of SUS was partly informed by his work on ISO 9241-11 (International Organization for Standardization, 2018), a standard for the definition and measurement of usability. It describes usability in terms of effectiveness, efficiency, and satisfaction in the context of use. Because the first two are difficult to compare across systems, SUS focuses on user satisfaction (Brooke, 2013). The resulting score is only indicative in nature. The SUS is not diagnostic and can not pinpoint specific usability issues. Despite its limitations, multiple studies have shown the SUS is a valid and reliable high-level measure that is applicable to a wide range of technologies (Bangor et al., 2008; Sauro, 2011).

2.13.3 Needs and Recommendations

The reviewed methods, frameworks, and instruments share common themes and goals. They aim to guide the specification, design, and implementation of AR/MR solutions to align with the specific needs and requirements of the manufacturing domain, assess the effectiveness and impact of these systems in terms of user acceptance, performance, and organizational fit, and provide structured approaches to support informed decision-making and optimization of AR/MR adoption in manufacturing training.

Several connections can be drawn between the methods discussed. Palmarini et al.'s (2017) questionnaire-based tool and Danielsson et al.'s (2020) framework both focus on guiding strategic decision-making for AR/MR adoption in industrial contexts. The emphasis on alignment between training objectives, contents, methods, and outcomes in Borsci et al. (2015) is echoed in the design considerations of Taylor's (2021) FLXD framework and Laviola et al.'s (2021) "minimal AR" approach. Additionally, the human-centered approach of Longo et al.'s (2017) SOPHOS-MS framework aligns with the user-centric focus of both Jetter et al.'s (2018) TAM-based framework and Masood et al.'s (2019, 2020) TOE-based model. Crucially, three studies [Palmarini et al. (2017); Borsci et al. (2015); danielsson2020augme] explicitly state the lack of validated tools.

This review highlights the importance of considering user needs, systematic fit, and technical feasibility when designing and implementing AR/MR systems for manufacturing training. It demonstrates the potential of structured approaches to guide the development and assessment of AR/MR solutions, ensuring their alignment with domain-specific requirements to maximize their effectiveness.

Ideally, these frameworks should serve to align all aspects of the system's design with the business objectives (Borsci et al., 2015), consider user needs related to usability and benefits to deliver a compelling value proposition (Jetter et al., 2018), and address technology and organizational issues that threaten short and long term success (Masood & Egger, 2019). Priority should be given to user acceptance, technical integration, organizational fit, and content creation considerations (Masood & Egger, 2020). However, the need to validate and refine existing tools and frameworks through empirical research in real-world manufacturing contexts is evident (Danielsson et al., 2020).

2.14 Next Steps

This section provides a high-level recap of findings before describing a novel affordance-based approach to study design for AR/MR assisted learning assessment. Finally, it will enumerate the key gaps and limitations identified in the literature, which will provide the basis for the problem statement and study design.

2.14.1 Summary

Turnover in the manufacturing workforce and the lack of skilled labor necessitates scalable, efficient training methods. Furthermore, the shift from mass production to mass customization forces operators to contend with wide variance in the assembly steps required at each workstation. Together, these trends demand innovative methods for operator training and support.

Preliminary studies suggest that emerging AR/MR technologies may provide a solution to address these challenges. These systems offer real-time, contextually relevant instruction, the educational benefits of which are grounded in well-established learning and cognitive theories. However, despite their proclaimed advantages, the manufacturing industry has been slow to embrace augmented training systems. That adoption has been hindered by various factors, including technical limitations, market considerations, and business requirements.

Researchers and practitioners have developed various tools, frameworks, and instruments to help overcome those obstacles. These tools aim to guide the specification, design, implementation, and assessment of AR/MR systems, ensuring their alignment with the unique requirements of the manufacturing industry.

Various case studies have also been conducted within the context of manufacturing. Unfortunately, their results do not yet provide a clear picture of the value proposition of AR/MR in manufacturing training. The research landscape is characterized by a limited number of empirical studies, heterogeneity in study designs, and insufficient validation in real-world manufacturing contexts. These factors make it challenging to draw definitive conclusions about the effectiveness and generalizability of AR/MR interventions in the domain.

2.14.2 An Affordance-Based Approach

To address some of the identified limitations and provide a more comprehensive understanding of the factors influencing the effectiveness of AR/MR in manufacturing training, this research proposes an affordance-based framework. The framework conceptualizes AR/MR technologies as bundles of affordances that, when appropriately leveraged and implemented using best instructional design practices, can lead to improved learning outcomes and performance. The development of this framework is grounded in the theoretical bases and informed by the insights gained from the literature review.

Parsons and MacCallum (2021) emphasizes the benefits of this approach over a feature-based perspective. They claim affordances are more generalizable than specific implementations and enable comparison across contexts, while still being highly contextualized to the domain of interest. Their systematic review of 21 empirical studies found that "studies that did not address any of the key affordances identified as relevant ... showed relatively poor learning outcomes" (2021, pp. 89-90). This suggests paying close attention to relevant affordances when designing AR systems may lead to better results.

Through their review, the authors synthesized five key affordances of AR/MR that can enhance learning in medical education: (1) reducing negative impacts like risk and cost, (2) visualizing the invisible, (3) developing practical skills in a spatial context, (4) enabling device portability across locations, and (5) facilitating situated learning grounded in the professional context. By highlighting the rationale for an affordance-based approach and the specific affordances identified as relevant for training in this hands-on domain, the authors provide a strong framework for adopting a similar approach.

While Parsons and MacCallum (2021) affordances captured high-level organizational goals like risk reduction and operational flexibility, our approach focuses on identifying specific benefits that can directly optimize learning processes and outcomes in AR/MR manufacturing training environments. Rather than focusing on broad potential benefits, our affordances directly apply established learning theories and instructional principles that promote hands-on practice, reducing cognitive load, improving spatial awareness, and creating an intuitive user experience within the manufacturing training context. The ten affordances are summarized in Table 2.3.

Table 2.3: Proposed Affordances for AR/MR Manufacturing Training

#	Affordance	Description
1	Task Instructions	A description of how to complete the task.
2	Hands-On Engagement	The learning method involves physical interaction with the subject matter.
3	Direct View of Work	The work area is viewed directly, without requiring a shift of focus from the workspace to a separate display.
4	Freedom of Movement	The device does not hinder the user's movement with a bulky or tethered design.
5	Step-Wise Guidance	Instructions are presented sequentially, adapting to user needs and pace.
6	Feedback Mechanisms	The system provides real-time feedback on user actions.
7	Workspace Integration	Instructional materials are integrated with the workspace.
8	Sensor-Based Interaction	The system is controlled with sensor-based input devices, eliminating the need for physical controllers.
9	User-Centric Display	Instructions are displayed in the user's view, rendered from their perspective.
10	Freeform Interaction	The system allows for natural manipulation of the workpiece.

These affordances were identified based on their direct applicability to the learning tasks within an AR/MR environment, their alignment with a carefully chosen set of instructional treatments, and their foundation in educational theories known to influence learning outcomes positively. Each affordance serves to operationalize these theories within the context of the experimental design, with the expectation that their integration into the instructional treatments will lead to measurable improvements in learning and performance.

As discussed in Section 2.11.2, the theoretical benefits underpinning these affordances are rooted in educational theories that are particularly relevant to AR/MR learning environments. *Active learning theories*, including *experiential learning theory* support the idea that learning is enhanced through direct experience and reflection, which is fundamental to several of the identified affordances, including "Hands-On Engagement," "Step-Wise Guidance," and "Feedback Mechanisms."

Flow theory emphasizes the importance of a state of heightened focus and immersion for optimal learning, which is fostered by affordances that engage users in a compelling and intuitive way, such as "Egocentric Display" that ensures the instructional content is seamlessly integrated into the user's field of view.

The theory of *embodied cognition* posits that cognitive processes are deeply intertwined with the physical actions of the body. In an AR/MR setting, affordances that align with this theory, such as "Freeform Interaction," allow for a more natural and intuitive learning process by leveraging the body's movement and spatial orientation. Other *constructivist* theories, including *animate vision* and *spatial cognition theory* are similarly represented by "User-Centric Display," and "Workspace Integration."

Cognitive load theory provides a framework for understanding how information is processed and suggests that well-designed instructional materials can reduce unnecessary cognitive load, making learning more efficient. This directly relates to affordances like "Sensor-Based Interaction" which simplifies the user interface, and "Workspace Integration," which eliminates context switching associated with referencing instructions away from the work surface.

Lastly, the affordances have been selected with *educational best practices* in mind, ensuring that they not only align with theoretical perspectives but also adhere to the principles of effective instruction design, such as clarity, engagement, and scaffolding.

This design links the chosen affordances with the framework for instructional design with AR/MR augmentation for industrial training applications that was proposed in Section 2.11.3 and illustrated by Figure 2.19.

2.14.3 Advancing the Research

The findings of this literature review underscore the need for further research to address the gaps and limitations in the current understanding of AR/MR technologies in manufacturing training. To advance the field, future studies should prioritize the following nine considerations, listed in no particular order:

1. **Address Ecological Validity**: Conducting research in real-world industrial settings, with suitable tasks and participants to help ensure that the findings are directly applicable and relevant to the unique challenges and requirements of manufacturing training.

2. **Incorporate Instructional Design Best Practices**: Firmly grounding study designs in learning and cognitive theories will optimize the effectiveness of AR/MR training solutions. By leveraging these principles, researchers can provide a model for future implementations and contribute to the development of evidence-based guidelines for designing AR/MR training programs.

3. **Employ Rigorous Methodologies**: Using well-controlled experimental designs, reliable and valid measurement instruments, and appropriate statistical analyses to establish the reliability and generalizability of the findings.

4. **Compare Multiple AR/MR Technologies**: Comparing Mobile, HMD, Projected, and Indirect methods to provide insights into their relative effectiveness and suitability for different manufacturing training scenarios.

5. **Study Learning Outcomes Hollistically**: Providing a more comprehensive understanding of the impact of these technologies on skill acquisition and maintenance over time by assessing training outcomes not just in terms of immediate learning effects but also longer term recall and retention

6. **Collect User Feedback**: Including data and analysis on user satisfaction, usability, and workload to inform iterative improvement and user-centered design, and help ensure that the resulting systems are effective, engaging, and intuitive for the target audience.

7. **Use an Affordance-Based Approach**: Designing treatments and interpreting their effects not in terms of transient hardware capabilities, but as a bundle of affordances each with corresponding theoretical benefits.

8. **Apply a Standard Methodology**: Reducing the heterogeneity of study designs will facilitate the direct comparison of results and synthesis of findings, improve their collective generalizability, and provide a common language for researchers and practitioners alike.

9. **Provide Practical Recommendations**: Framing research and findings in a way that supports the successful design and implementation of these systems, and translating those into fact-based decision and planning frameworks will accelerate industry adoption.

This proposed affordances framework serves as a foundation for the current study, which, in part, aims to empirically validate its application in a real-world manufacturing training context. This work will apply the affordance framework to the design of the treatments, allowing us to interpret effects based on the underlying benefits, which are ephemeral, rather than any transient technologies. We trust this will provide valuable new insights into the most influential factors in the value of augmented instruction for learning, recall, and retention, thereby contributing to the development of best practices for their implementation in real-world industrial settings.

2.14.4 Closing

This literature review has provided a comprehensive examination of the current state of research in the domain, critically analyzing empirical studies that assessed the efficacy of AR/MR interventions while also identifying persistent gaps, limitations, and adoption challenges. Moreover, the review introduces a novel affordance-based framework as a theoretically-grounded approach to guide the design and evaluation of AR/MR training solutions. The following chapter will articulate the specific problem statement, research questions, and hypotheses that guide this endeavor.

3 Problem Statement

This chapter identifies the research gap and frames our study's objectives. It begins by summarizing the problem context and need detailed in the preceding chapters. It then briefly outlines the problem, bounded by gaps and limitations of the current literature. The primary research question and supporting questions are identified. Finally, the work's primary contributions are identified.

3.1 Problem Identification

Despite a demonstrated need for operator assistance, the emergence of promising technology, a theoretical basis for performance enhancement, a "solution" that connects it all, and frameworks to support its development and assessment, XR adoption remains slow in manufacturing. Until a compelling business case can be made for the substantial investment required, it is unlikely that industry will move beyond isolated proof of concept studies. With continued delays to adoption comes the risk that hardware providers lose their appetite for the massive ongoing R&D budgets these systems demand and reduce or end their commitment in the sector. Therefore, there is a pressing need to assess the business benefits of augmented instruction, which may come in the form of improved operator performance, learning outcomes, long-term retention, user comfort, quality, or other measurable returns on the investment.

As enumerated in Section 2.14.3, our review of the literature has highlighted a number of gaps. This list includes few direct comparisons of AR/MR technologies, a focus on immediate performance measures, a lack of ecological validity, and heterogeneous designs. More specifically, too many studies focus on comparing a single AR/MR intervention to traditional training methods, leaving unanswered questions about how various AR/MR technologies stack up against each other, and which are most suitable for specific training contexts and desired learning outcomes. The prevailing emphasis on immediate performance measures, such as task completion time and error rates, provides a limited view of training success and knowledge transfer. Furthermore, a significant portion of existing research on AR/MR in manufacturing training relies on simplistic tasks, irrelevant environments, or novice participants. While these controlled studies provide valuable insights into the potential of AR/MR technologies, they may not accurately reflect the complexity and challenges of real-world industrial settings. Finally, the literature includes a wide variation in research designs, technologies, training content, and outcome measures employed across studies. While this diversity reflects the breadth and complexity of the field, it also poses significant challenges for comparing results, synthesizing findings, and drawing conclusive insights about the effectiveness of AR/MR interventions.

Collectively, these limitations hinder our ability to draw definitive conclusions about the effectiveness of AR/MR technologies in manufacturing training, leaving organizations without clear guidance on how to leverage these tools for optimal learning outcomes and return on investment. In short, additional data and more comprehensive results are needed to provide an accurate assessment of AR/MR efficacy across the diverse problem and solution space.

To address these shortcomings, this research proposes an affordance-based framework that conceptualizes AR/MR not as transient technologies, but as bundles of theoretically-grounded affordances designed to optimize learning processes. By systematically operationalizing these affordances through carefully designed instructional treatments, we aim to provide empirical insights into the factors that most influence the efficacy of AR/MR-augmented training. Grounded in the literature and tailored to real-world manufacturing contexts, the proposed study will yield actionable recommendations to accelerate the adoption of these innovative technologies where their benefits can be maximized.

3.2 Research Questions

This gives rise to the central research question of this work:

> *How do different AR/MR instructional methods, designed to leverage specific affordances, impact operator learning, recall, and retention in a real-world manufacturing assembly training context?*

Several supporting questions are also identified:

A. What are the relative effects of various AR/MR technologies on immediate learning outcomes, such as task completion time and error rates, compared to traditional paper-based instructions?

B. How do these AR/MR technologies influence long-term recall and retention of assembly skills, as measured by performance on the same task after a designated period without further training?

C. To what extent do the specific affordances of each AR/MR technology, such as hands-free interaction, spatial registration, and user-centric displays, contribute to the observed learning, recall, and retention outcomes?

D. How do operator characteristics, such as related experience or demographics, influence the effectiveness of each instructional method?

E. What are the perceived workload, usability, and user satisfaction associated with each AR/MR technology, and how do these factors relate to learning, recall, and retention outcomes?

3.3 Key Contributions

By addressing key gaps in the existing literature, this research will contribute to the development of evidence-based guidelines and best practices, ultimately advancing the field and informing future efforts to optimize AR/MR-assisted training programs for improved learning outcomes and return on investment. Three key contributions of this work are:

1. An innovative, affordance-based study design for comparing the impact of multiple AR/MR technologies on learning, recall, and retention in a real-world manufacturing training context. A pilot study is incorporated to assess the ability of this design to address limitations identified in the literature while also incorporating its best practices.

2. A comprehensive analysis of the immediate learning effects of different AR/MR technologies, considering multiple performance metrics, cognitive load, usability, qualitative feedback, and the role of specific affordances.

3. An examination of the factors influencing the long-term effectiveness of AR/MR training, including the relationships between initial learning outcomes, recall and retention, user experience, and individual differences.

To address these research questions and deliver the outlined contributions, a comprehensive methodology is outlined in the following chapter. It details the the selection of AR/MR technologies, development of instructional treatments, and the methods for data collection and analysis.

4 Methods

This study addresses the identified gaps by comparing three AR/MR technologies against each other and a paper-based control in a real-world manufacturing assembly training context. The approach is unique in its use of an affordance-based framework and comprehensive assessment strategy that examines immediate learning outcomes, recall, and retention, along with workload, usability, and qualitative feedback. Authentic assembly processes and validated training methods help ensure ecological validity.

The experimental design, data collection procedures, and analysis strategies are detailed herein. This chapter begins with an overview of the key components of the research methodology, followed by a comprehensive description of the measures and variables used to assess learning, recall, and retention outcomes. The experimental procedures are then outlined, including participant recruitment, random assignment to treatment conditions, and the conduct of the training sessions. Considerations for compliance with ethical guidelines and the steps taken to ensure participant safety and confidentiality throughout the conduct of trials are also discussed. Finally, data extraction and analysis procedures are described in detail, highlighting the use of both quantitative and qualitative methods to gain a holistic understanding of the impact of different IMTs on operator performance and learning outcomes.

The insights gained from this research have the potential to advance our theoretical understanding of how AR/MR technologies support learning and skill acquisition, while also informing the practical application of these tools in manufacturing training contexts.

4.1 Experimental Design Overview

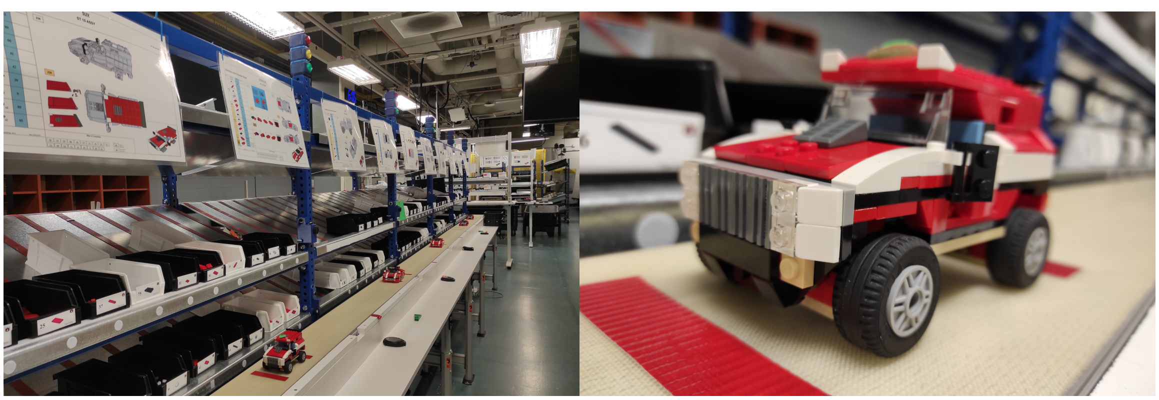
To assess the effect of augmented instruction on operator performance, human subjects were asked to learn and repeat a simulated manufacturing assembly task. A convenience sample of adult participants without relevant experience were recruited from the Auburn University community and randomly assigned to one of four instructional treatments, each with a different level of augmentation. This between-groups approach was adopted to allow for direct comparison between different levels of the treatment and to minimize the learning effect that would otherwise occur over repeated trainings.

The experiment included two phases, where learning and recall were tested for the assigned treatment level while performance measures were recorded. After each phase, validated instruments were used to assess the participant's perceived workload and their impression of the treatment's usability. Several weeks after the initial intervention, participants were invited back to an event where retention was also tested.

Data related to the timing, errors, and ultimate outcome for each task was recorded, allowing for a detailed performance assessment at the task and participant level. Subsequent analysis of all performance data quantified the treatment effect. The addition of retention data allows the study to also compare the long-term effectiveness of each treatment. Demographics, perceived workload, and usability data were then used to identify other contributing factors and better understand the results.

These experiments were conducted in the Tiger Motors Lean Education Center[1] (aka the Lean Lab). Designed to simulate modern automotive manufacturing and teach best practices in a real-world setting using LEGO® vehicle assembly, this award-winning facility provided an ideal setting for the study. Pictured in Figure 4.1a, Tiger Motors is an integral research and education component of the Industrial and Systems Engineering Department in the Samuel Ginn College of Engineering at Auburn University. The facility and simulation design were primarily the work of graduate students (Moyo, 2013).

Participants acted as operators learning part of the Model T (SUV) assembly process. A completed SUV model is pictured in Figure 4.1b. This process has been repeated thousands of times in lean education courses without significant incident.



(a) Conveyor and final assembly workstations (b) Completed SUV model

Figure 4.1: Tiger Motors Lean Education Center and SUV Model

[1] Located in the basement of the Shelby Center for Engineering Technology, room 0317. Address: 345 W Magnolia Ave, Auburn, AL 36849.

Each participant completes the same task sequence with one of four different Instructional Media Treatments (IMTs). Each IMT offers a different level of augmented instruction, ranging from traditional paper instructions with no augmentation to a mixed reality head-mounted display where interactive instructions are superimposed on their field of view.

Instructional design for all augmented IMTs was based on the paper work instructions, adapted only as needed to leverage the specific affordances of each technology. This helped ensure that any differences observed in the study were due to the augmentation level of each IMT, not variations in instructional content.

The remainder of this chapter will describe in detail all aspects of the study's methodology, including participant recruitment, human subjects considerations, experimental design, data collection, and analysis. The chapter will conclude by considering the limitations of this study.

4.2 Study Design

In this section, we will provide a detailed description of the study design. We will begin by discussing the task and the authentic manufacturing environment in which the experiments are situated. Next, each of the four treatment levels are described. Finally, the affordance-based nature of this design will be detailed, wherein the underlying affordances are identified for each treatment and inherent tradeoffs are described.

4.2.1 Task and Context

The Lean Lab assembly line consists of two manufacturing cells followed by a conveyor with five additional stations. Each manufacturing cell consists of five workstations arranged in a U-shape. Each participant is tasked with the operation of one of 15 workstations involved in the SUV assembly.

The experiment is run at ST-8, which is located in the middle of the second manufacturing cell, flanked on either side by workstations six through ten. This arrangement is pictured in Figure 4.2.

All workstations in both cells are similarly equipped with work surface and trays for parts bins. Work instructions for both car models normally produced on the line are displayed above the bins. The bins are removable to facilitate part resupply, but their arrangement at

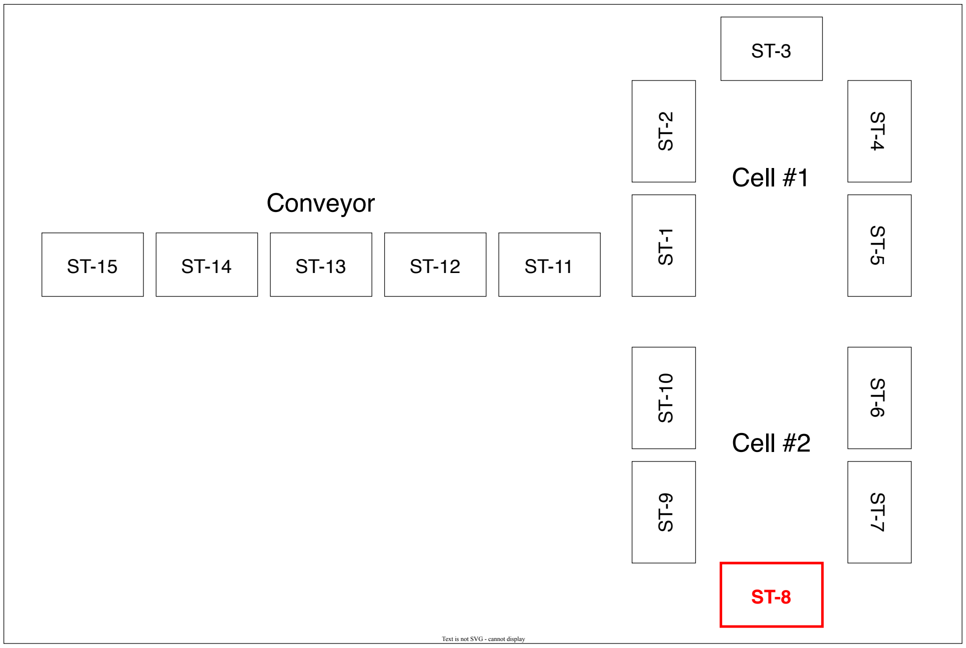
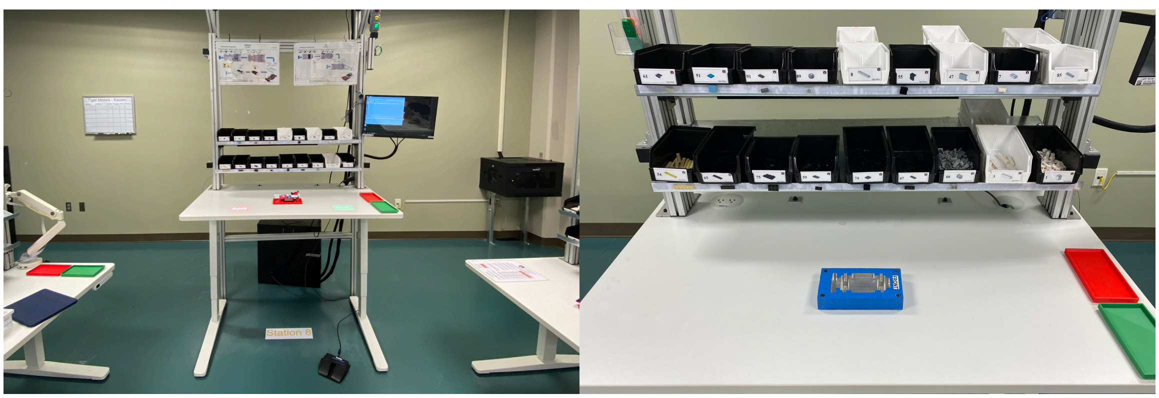


Figure 4.2: Tiger Motors Assembly Line

each station is standardized and specified on the work instructions. Figure 4.3 shows ST-8 and its bin layout.



(a) ST-8 (b) Bin Layout

Figure 4.3: Workstation Eight

The green and red trays pictured on the right of the work surface are for finished goods and rework, respectively. The central fixture provides for standard placement and secure retention of the workpiece. All of these are removable to allow the lab to operate with varying levels of Lean practice in effect.

4.2.2 Treatment Levels

As described in Section 4.6.4, each participant in this between-groups design is randomly assigned to a single level of the Instructional Media Type (IMT) treatment. There are four levels of this treatment, each with increasingly augmented work instructions: (1) traditional paper work instructions, (2) projected augmented reality, (3) head-mounted optical see-through augmented reality, and (4) head-mounted optical see-through mixed reality. All are detailed in the sections that follow.

Paper Work Instructions

Paper work instructions (PWI) are printed instructions traditionally used in manufacturing assembly processes. Unlike all other IMTs in this study, PWIs are inherently static and do not adapt to the operator in any fashion.

The process at workstation eight (ST-8) is a three step sequence in which 16 pieces, consisting of eight different part types, are affixed to the workpiece. The PWIs for this process are pictured in Figure 4.4.

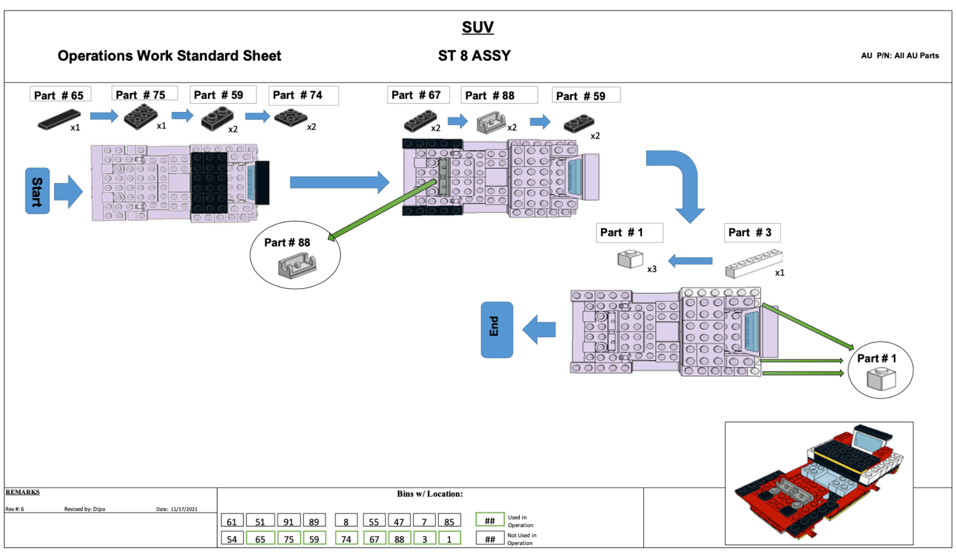


Figure 4.4: SUV assembly instructions for ST-8.

The instructions consist of one top-down view of the workpiece for each step, plus an isometric view of the ST-8 assembly. In each step, the car's prior state is shown in grey and new parts have the correct color. A sequence of parts is shown for each step, with the unique part number and quantity required. It is up to the operator to discern the correct placement of those parts from the diagram. Additional detail views are incorporated to each PWI where more clarity or detail are required. At the bottom center of each PWI a map of the workstation's parts bins depicts which are used in the task.

Manufacturing simulations in the Lean Lab are expected to run at a takt time[2] of one minute. Therefore, the instructions for each of the 15 workstations have been carefully designed to include one minute of work content.

While it may seem trivial to complete the work at ST-8 correctly in one minute or less, experience shows that is not the case. The Lean Manufacturing Systems class at Auburn University (INSY 5800/6800), has validated these instructions through countless lab sessions in the administration of that course. They served as the basis for the design of all other IMTs, and as the control for this study.

Projected AR

Projected AR (PAR) systems integrate work and instruction by projecting the latter onto the work surface. Work steps are displayed sequentially, either under operator control or automatically triggered.

ST-8 is equipped with a PAR system by LightGuide[3] (LG), a Michigan-based company focused on innovative, AR-based manufacturing solutions. As shown in Figure 4.5, Their system uses a Windows PC, industrial-grade projector, and 2D or 3D vision system, along with other optional input devices and tools. All components are commercially available and integrated by the LG software, where digital work instructions are both authored and played back.

Note that the Lean Lab's LightGuide system predates this study. Its specification, installation, setup, and configuration, along with the design and implementation of the instructional materials it uses at ST-8, were conducted by prior teams.

The system at ST-8 uses a video projector and depth-sensing camera mounted above the paper work instructions. The output of a PC running Windows 10 and LightGuide software

[2] Takt time is a critical measure of the overall efficiency and performance of manufacturing systems. It sets the pace of production to align with customer demand, thus dictating the maximum time allowed at each workstation. By balancing the production line in this way, waste due to overproduction or delays is reduced (Ali & Deif, 2014). Takt time is integral to the Toyota Production System (TPS) developed by Taiichi Ohno (1988).

[3] Lightguide: https://www.lightguidesys.com/

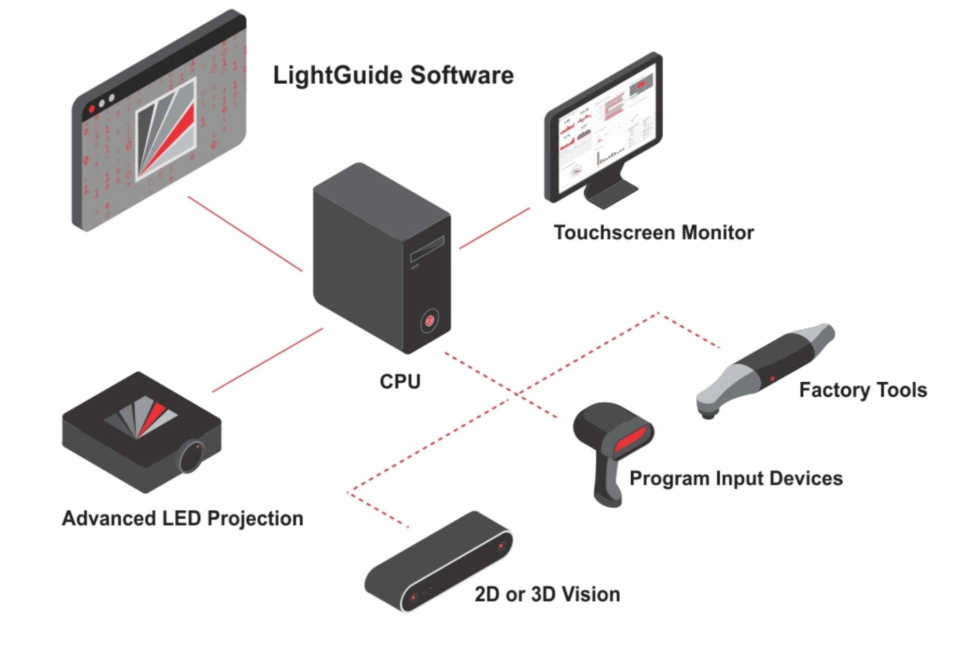


Figure 4.5: LG System Architecture. Source: LightGuide website

is displayed on a monitor conveniently situated to the right of the bin trays. All equipment, except the PC itself, is mounted on a robust structure of extruded aluminum components from 80/20[4] to ensure operator safety, consistent alignment, and stability. A foot-operated switch provides an alternative to gesture-based control of the system, but it was not used in our study. The complete installation is shown in Figure 4.6, with each component labeled.

To create a consistent user interface across all XR treatments, this study leveraged the LG's vision-based system for operator control. In this mode, an infrared time-of-flight camera provides a real-time depth map of the work area. This 2D video signal encodes the distance of objects from the camera as a color value at each pixel. That output, commonly known as a depth map, is processed using computer vision techniques to detect motion and identify any obstructions in the workspace. Figure 4.7 demonstrates the output of this system in a simulated operator interaction.

With this information the LG can, for example, recognize the operator's hand placement and trigger system actions, warnings, or data logging events. In the ST-8 implementation, some applications of this method include triggering a green or red overlay when the operator reached into the right or wrong part bin, advancing to the next assembly step when

[4] 80/20: https://8020.net/

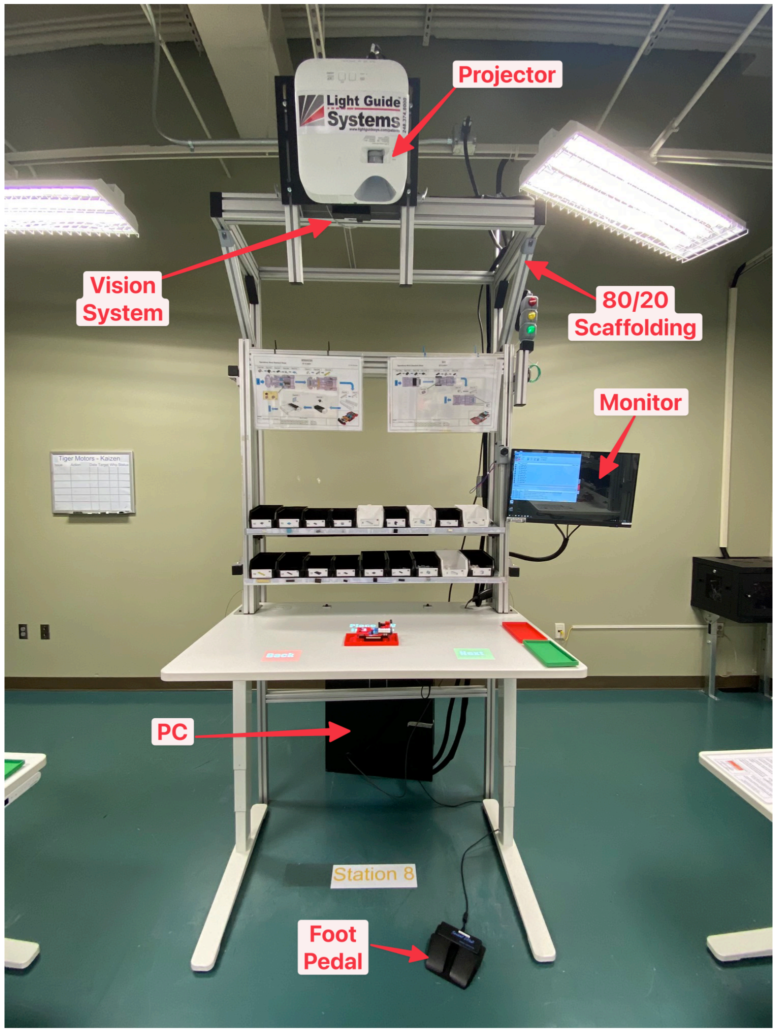
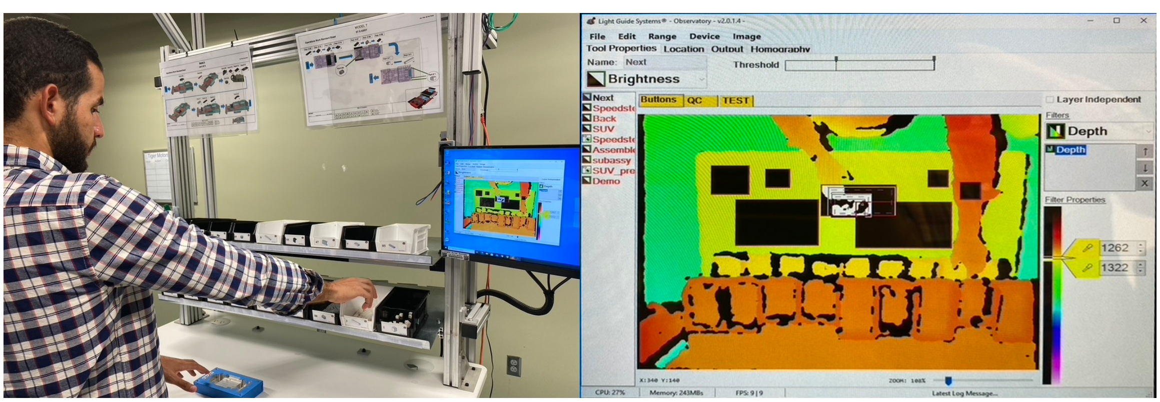


Figure 4.6: ST-8 LG Installation



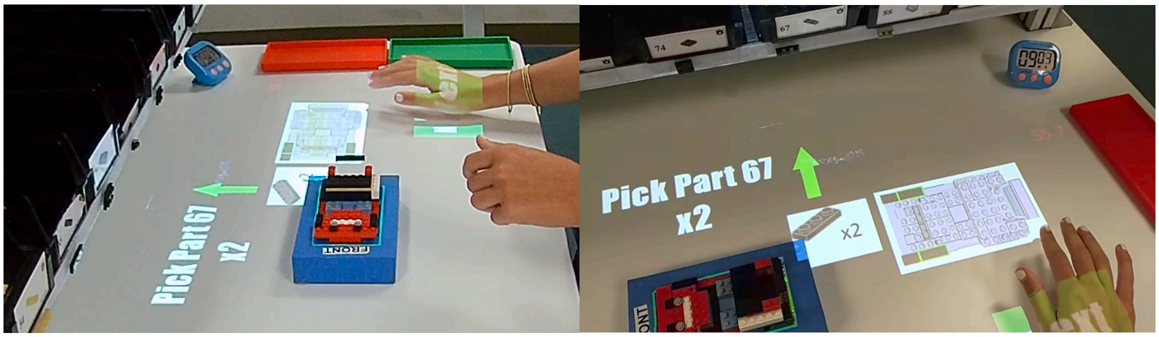
(a) Demonstrating LG Vision

(b) LG Vision Input

Figure 4.7: LightGuide Vision System

the operator swipes over the virtual `NEXT` button, and automatically advancing when it appeared that the operator had installed a piece.

The images below were grabbed from PAR trial recordings. Figure 4.8a shows a side view of the workspace and Figure 4.8b shows the same moment from the operator's perspective. Both images are cropped to focus on the work surface, but the operator view maintained the original aspect ratio with minimal cropping.



(a) PAR Side View (b) PAR Operator View

Figure 4.8: PAR in Operation

The PAR instructions were designed and implemented by students and graduate assistants in the Lean Manufacturing class. They were adapted from the same PWIs used as the control in this experiment. Like the PWIs, the PAR system has been validated through extensive use in the lab.

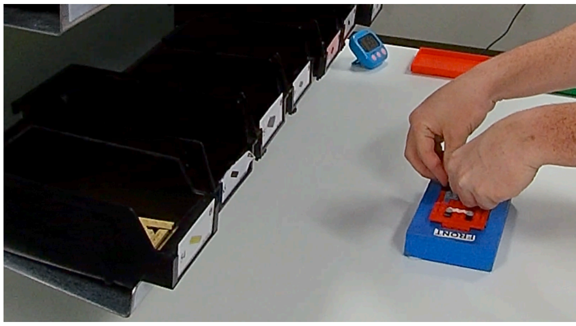
Head-Mounted AR

Like PAR, head-mounted AR (HMDAR) systems integrate dynamic instructions into the work area. Whereas PAR relies on traditional projectors and classical computer vision techniques, HMDAR employs sophisticated displays and fused sensor data to superimpose instructions and virtual controls directly into the operator's field of view, properly aligned with the work area.

As detailed in the subsequent section on System Development, Microsoft's HoloLens2[5] (HL2) was used for this treatment. The HL2's area-based tracking capabilities were used to align the virtual and physical coordinate systems, enabling the proper in-view placement of UI objects, independent of the position and orientation of the operator's head. As with the PAR treatment, user motion was the only input modality, implemented via the HL2's

[5] HoloLens2: https://www.microsoft.com/en-us/hololens

more sophisticated hand tracking and gesture recognition systems. Figure 4.9 shows the side and operator view of this treatment at the same moment.



(a) HMDAR Side View

(b) HMDAR Operator View

Figure 4.9: HMDAR in Operation

The instructional content for this treatment was intentionally designed to mirror that of the PAR system. By controlling for variables such as instructional content and task complexity, the study design allows for a direct comparison based on the distinct affordances of each technology—namely, the nature of their display and interface. This approach ensures that differences in operator experience and performance can be attributed most directly to the technological medium.

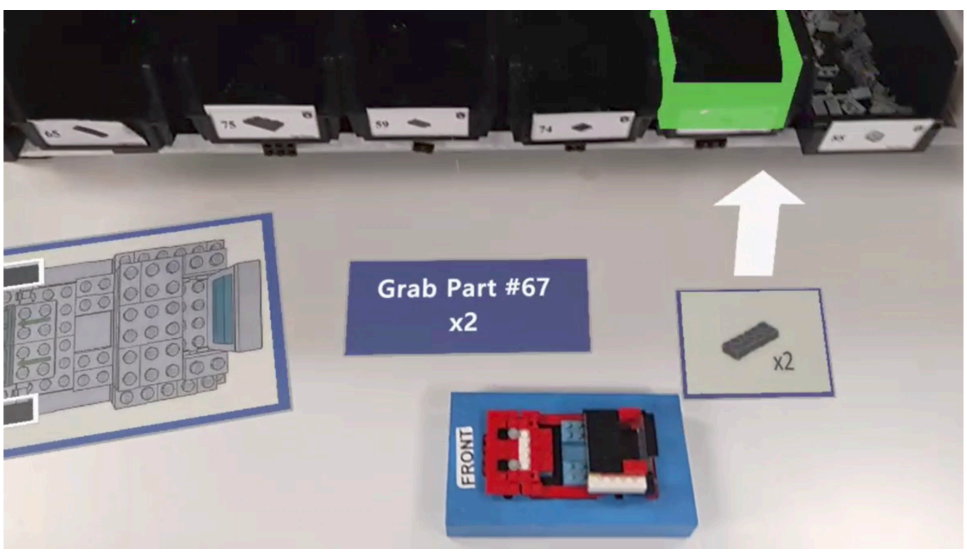


Figure 4.10: HMDAR at Pick Part 67

For comparison purposes, Figure 4.10 shows the HMDAR operator's view of the same instructions pictured in Figure 4.8 for the PAR treatment. The two differ only in that the digital work instruction is placed on the operator's left in the HMDAR version to prevent overlap.

Head Mounted MR

Head-mounted MR (HMDMR) extends the capabilities of the HMDAR treatment with a more sophisticated tracking method that allows for more natural interactions with the workpiece. For all other treatments, the workpiece remains in a fixture, forcing the operator to align themselves, both physically and mentally, to its placement.

HMDMR eliminates the need for a fixture and allows the operator to rotate and/or lift the work off the surface if it seemed natural to them to do so. It achieves this by incorporating model-based tracking to align instructions with the workpiece itself. The added flexibility was expected to make the process more intuitive and ergonomic, enhancing operator performance.

Using model-based tracking, part placement indicators remain properly aligned with the workpiece, regardless of its position and orientation in space. This is demonstrated in Figure 4.11, where participant #1040 has rotated the workpiece approximately 45 degrees from its normal orientation to facilitate the installation of parts at the rear of the model. Note that the green part proxies and white placement arrows, all virtual, remain properly oriented.

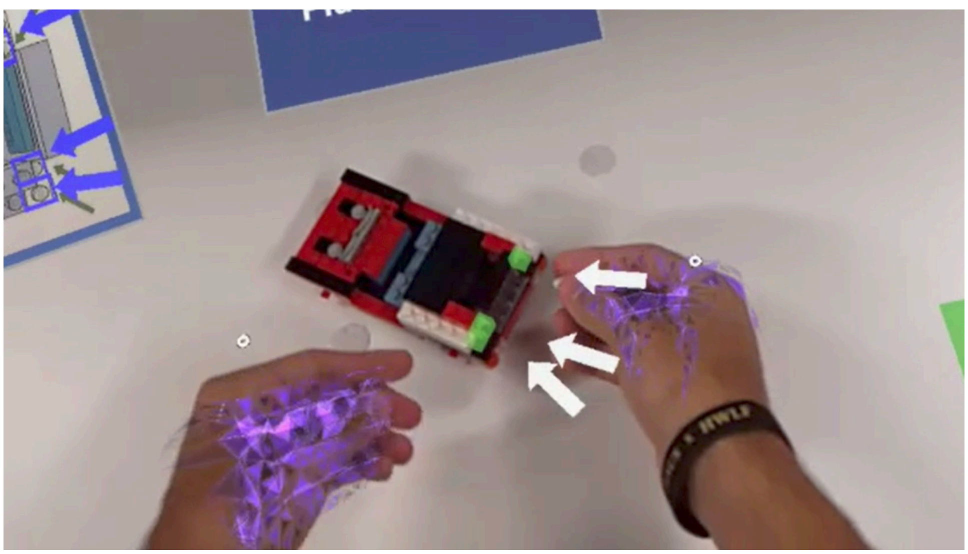
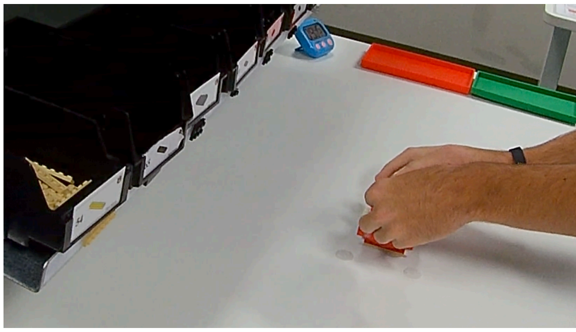


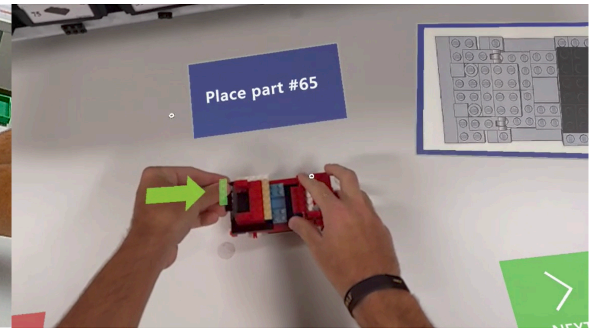
Figure 4.11: HMDMR Tracks Model Orientation

Figure 4.12 provides a side and operator view of this treatment, and illustrates the lack of a fixture. Otherwise, the HMDMR treatment was unchanged from the HMDAR, allowing for direct comparisons of all treatments based only on the affordances of interest.

4.2.3 Affordances and Theoretical Benefits



(a) HMDMR Side View



(b) HMDMR Operator View

Figure 4.12: HMDMR in Operation

Table 4.1: Summary of Affordances by Treatment

Affordance	**PWI**	**PAR**	**HMDAR**	**HMDMR**
Task Instructions	Y	Y	Y	Y
Hands-On Engagement	Y	Y	Y	Y
Direct View of Work	Y	Y	Y	Y
Freedom of Movement	Y	Y	Y	Y
Step-Wise Guidance		Y	Y	Y
Feedback Mechanisms		Y	Y	Y
Workspace Integration		Y	Y	Y
Sensor-Based Interaction		Y	Y	Y
User-Centric Display			Y	Y
Freeform Interaction				Y

Table 4.1 maps each treatment to its set of affordances. It demonstrates a progression of complexity and sophistication. "Task Instructions," "Hands-On Engagement," and "Direct View of Work" provide the baseline capabilities for all treatments. Additional affordances add layers of instructional complexity and interactivity.

While it may be tempting to make a-priori claims about the most influential affordances or theories in this context, the relationships are not so clear-cut. Each identified affordance can claim some benefit from most identified theories; all are interrelated and contribute to the overall learning experience.

Instead, this work hypothesizes that treatments leveraging more affordances will result in better learning outcomes. But the success of these treatments also depends on the quality of their implementation and the balance between enhanced functionality and added complexity. The latter may have a negative impact on the overall user experience, increasing cognitive load, breaking flow, and otherwise offsetting the intended gains. Finally, the interplay between different affordances — how they complement or interfere with one another — is also a significant consideration.

4.2.4 Tradeoff

Both HMD treatments were designed to resemble the PAR, and, by extension, PWI experiences as closely as possible. This choice was made to deliberately control for the instructional design and focus treatment differences on the benefit of underlying affordances. While any of the treatments could have been “improved” with additional system-specific functionality, that would have been detrimental to the experimental design. For example, obvious shortcomings of the PWI could have been corrected, or participants could have been given the option to use the PAR’s footswitch. Either would likely have improved the results for those treatments but led to less relevant comparisons. This methodology acknowledges the potential limits to each treatment’s efficacy, but upholds the integrity and clarity of the study’s comparative analysis.

4.3 Measures and Variables

In this study, we evaluate the efficacy of different instructional methods in manufacturing assembly training, focusing on key variables that impact learning, recall, and retention. This section outlines these crucial variables, alongside a comprehensive set of hypotheses for both primary and secondary outcomes, ensuring a robust and systematic assessment of each instructional treatment’s effectiveness.

4.3.1 Dependent Variables

The phenomena of interest in this study are learning, recall, and retention. This section will define each before discussing how they are operationalized through observed and calculated measures.

Constructs

Learning, recall, and retention are the outcomes of interest in this study. They were chosen for two reasons: (1) they are commonly used in related studies, and (2) they are supported by research in cognitive and educational psychology. But learning, recall, and retention are all psychological *constructs*; abstract concepts or ideas used to explain a phenomenon or behavior (Braun et al., 2001). They are fundamental to the research questions, but not directly observable, and therefore must be inferred from behaviors, actions, or outcomes.

Learning is the process through which knowledge, skills, behaviors, or values are acquired or modified (Bloom, 1956). In manufacturing assembly, learning encompasses not just understanding the theoretical aspects of an assembly process but also acquiring the practical skills to execute tasks efficiently and accurately.

Recall refers to the ability to access information from memory without being re-exposed to it after initial training. Recall is especially relevant in manufacturing settings where operators often need to perform tasks without step-by-step guidance, relying on their memory of the training.

Retention is the ability to maintain information, skills, or knowledge over time. In manufacturing, where precision and consistency are vital, an operator's retention can significantly impact production quality and efficiency.

Each of these constructs are a function of how effectively the instructional process encodes and stores knowledge and skills in long-term memory. These claims are supported by established theories in cognitive psychology, including the Information Processing Theory, which seeks to explain how humans process, encode, and retrieve information (Atkinson & Shiffrin, 1968). It is also supported by Ebbinghaus (Ebbinghaus, 2013) and works that followed. Originally published in 1885, Ebbinghaus first demonstrated that memories decay over time without reinforcement or repetition. Additionally, research in educational psychology, particularly studies on effective instructional strategies and their impact on long-term skill acquisition and knowledge retention, provide empirical backing to these concepts.

The focus on learning is justified as it provides insights into how different instructional methods (traditional vs. augmented) influence the speed and depth with which participants master new tasks. This aligns with cognitive load theory, which posits that reducing extraneous cognitive load and optimizing intrinsic and germane loads can enhance learning efficiency. The emphasis on recall is supported by active learning theories, which suggest that engaging, hands-on experiences (as facilitated by augmented reality technologies) can lead to more durable learning. Retention ties back to the experiential and constructivist learning theories, which argue that knowledge constructed actively by the learner is more likely to be retained over time. Furthermore, retention data can inform training protocols, indicating when refreshers or additional training may be needed to maintain proficiency.

Observed Measures

As constructs, learning, recall, and retention cannot be directly measured. For this study, they are operationalized through a variety of commonly-used measures. Learning is measured by how quickly a participant progresses towards the level of proficiency expected of a qualified operator. Specifically, it is a function of the dependent variables quality (uncorrected error count and type) and performance (task completion time). It can also be measured by the participant's reliance on PWI consultation (PWI count and duration) during the second experiment. Recall is assessed by how well participants can perform the task after the learning phase, without additional instructional support. This can be observed through the same quality and performance measures as learning. Finally, retention is measured by re-testing recall some time after the initial training, without further exposure to the task or instructional materials.

Calculated Measures

Each of the identified constructs can also be assessed using a variety of calculated measures, including Learning Rate (LR), Transfer Effectiveness Ratio (TER), and Overall Equipment Effectiveness (OEE). LR is a measure of the rate of change of learning and can be calculated in a number of ways. TER quantifies the value of time spent training (Kaplan et al., 2021; Roscoe, 1971), based on the amount of time required to reach certain training outcomes with (Y_c) and without (Y_x) augmentation, as seen in Equation 4.1.

$$TER = \frac{Y_c - Y_x}{Y_c} \times 100 \tag{4.1}$$

OEE has emerged as a fundamental and widely accepted KPI in manufacturing (Ng Corrales et al., 2020). Introduced by Seiichi Nakajima (1988) as part of Total Productive Maintenance (TPM), OEE evaluates overall manufacturing performance as the product of system availability, productivity, and quality, as shown in Equation 4.2. It is instrumental in pinpointing areas for improvement in equipment utilization and production processes.

$$OEE = Availability \times Productivity \times Quality \tag{4.2}$$

These three values are the percentage of measured vs expected speed, yield, and up-time, respectively. As seen in Equation 4.3, productivity is simply the number of units produced multiplied by the takt time, divided by the operating time. Completed units includes those with errors, but not those retired by the operator.

$$Productivity = \frac{UnitsProduced \times TaktTime}{OperatingTime} \tag{4.3}$$

Quality is the number of units produced less the number with errors, divided by number produced. This intuitive calculation is shown in Equation 4.4. Once again, retired units are not included in these counts.

$$Quality = \frac{UnitsProduced - DefectiveUnits}{UnitsProdcued} \tag{4.4}$$

OEE's third and final component, availability, is commonly calculated by dividing the amount of time that the system was operational (aka measured up-time) by the scheduled time. This is shown in Equation 4.5.

$$Availability = \frac{UpTime}{ScheduledTime} \tag{4.5}$$

In the context of this study, availability is out of the participant's control, but, as discussed elsewhere, can manifest in some system related issues that were encountered. Where appropriate, we will calculate availability by deducting time lost to system issues from the available time.

Integration

Together, these constructs and measures support the comprehensive analysis of the effects of augmented instruction in manufacturing assembly training. Together with workload and usability data, and qualitative feedback collected, we aim to provide a more complete picture of the roles that human factors, instructional design, and system performance have in achieving optimal results for skill acquisition, knowledge retention, and long-term performance.

4.3.2 Independent Variables

The primary independent variable in this study is the treatment level assigned to each participant. Of the four treatments, three serve as interventions: PAR, HDMAR, and HMDMR. The fourth treatment, PWI, is the standard for manufacturing assembly training and therefore a natural choice for the control.

Post-hoc analysis is conducted to determine if any of the participant demographics should be considered secondary independent variables. Those of particular interest include age and prior experience with LEGO or manufacturing. The results of this analysis may have implications for the generalizability of the study.

4.3.3 Controlled Factors

To ensure the validity of the outcomes, the study design carefully controlled for various factors, isolating the impact of the treatments.

1. Participant Sampling: The recruitment strategy aimed for a diverse and representative sample within the constraints of the study's target population.
2. Random Treatment Assignment: This method was utilized to evenly distribute potential confounding variables across the different treatment groups, thereby minimizing biases.
3. Uniform Device Usage: All participants wore the HL2 during the learning and recall experiments to standardize any potential impact of using the device.
4. Screening for Prior Experience: Prospective participants were screened and excluded if they had prior experience with similar AR/MR devices or Lean Lab assembly tasks.
5. Task Consistency: The content, complexity, and duration of the task were uniform for all participants, ensuring that any learning differences were attributable to the treatments rather than task variability.

6. Standardized Session Conduct: The environment and methodology of conducting each session were kept consistent, further ensuring that differences in outcomes were treatment-related.

These control measures were integral to maintaining the integrity of the study and ensuring that the results accurately reflect the effects of the instructional treatments.

4.3.4 Primary Outcomes

The primary outcomes of this study are the results of the hypotheses tests for each phase, as outlined below.

Learning Phase Hypotheses

The first group of hypotheses are designed to test the effect of each treatment on training outcomes.

H1: Learning

How does each IMT affect performance during the learning phase?

$$H_{1a}\text{: Average time per car varies with treatment}$$

$$H_{1b}\text{: Learning rates vary with treatment}$$

$$H_{1c}\text{: Average error count per car varies with treatment}$$

To better understand those results, additional analysis considers the treatment effect on error types, task completion rate, and first-task performance. Finally, treatment groups are investigated to determine which had the highest percentage of "qualified operators" at the end of the 10-min session. This is assessed relative to expected performance metrics, including OEE and takt time.

This approach provides a robust assessment of the instructional treatments' effectiveness during the learning phase. It examines key performance metrics—efficiency, accuracy, and learning progression—to capture a comprehensive understanding of participant performance. The evaluation against real-world manufacturing standards further ensures the study's relevance to practical training contexts. This design allows for a nuanced interpretation of how each treatment influences learning outcomes in manufacturing assembly training.

Recall Phase Hypotheses

The second group of hypotheses aim to evaluate the residual impact of each treatment on recall — specifically, participants' ability to correctly replicate the task without further training.

H2: Recall

How does each IMT affect performance during the recall phase?

$$H_{2a}\text{: OEE varies with treatment}$$

$$H_{2b}\text{: PWI reliance varies with treatment}$$

In contrast with the RQ1 analyses, which focused on the learning *progress*, these hypotheses are designed to assess the training *effectiveness*. OEE was chosen as the primary measure due to its practical relevance in answering "Did the training work?" It concisely quantifies the participants' ability to utilize acquired skills under conditions that mimic real-world expectations, where both speed and accuracy are crucial. For the purpose of this analysis, reliance will be measured by the number of times a participant refers to the printed instructions, and the duration of each.

Subsequent analyses will investigate the primary drivers of OEE (efficiency vs. quality) and PWI reference duration (frequency and length of references). Additional exploration into error types, task completion rate, first-task performance, and ongoing learning rate may also provide additional insight into the nuances of recall performance across treatment types.

Retention Phase Hypotheses

Phase three of the study is designed to assess the residual impact of each treatment on retention. This describes the durability of the learning and is measured by testing recall several weeks post-intervention. No additional training is provided in the meantime.

H3: Retention

How does each IMT affect performance during the retention phase?

$$H_{3a}\text{: Change in OEE varies with treatment}$$

TODO: update this hypothesis

Here we look at the change in operator performance since the recall experiment. We expect performance to degrade for all treatments, but are interested to find if there are significant differences in the magnitude of change based on the instructional method. Both absolute and percentage change will be tested, and the primary driver of OEE identified as before.

Given the limited data available from the retention phase (only one replication per volunteer participant), and the variable delay between experiments, these results will be treated as exploratory. While they may not be conclusive, these findings can help illuminate underlying patterns in treatment effects.

4.3.5 Secondary Outcomes

The study design supports several other areas for statistical, exploratory, and qualitative analysis. A variety of secondary outcomes, each described in the sections that follow, are incorporated to provide better context for and understanding of the primary outcomes.

Statistical

A variety of additional statistical tests were performed, involving workload, usability, demographics, and performance variability. Specifically, we plan to investigate the following relationships:

- TLX composite score (workload) and performance across treatments.
- SUS composite score (usability) and performance across treatments.
- Demographics (e.g., age, prior experience) and performance across treatments.
- Within-group performance variance

Exploratory

In addition to commonly used descriptive statistics and visualization methods, this study employed a variety of other statistical methods to explore second level effects.

- The TLX components (e.g., mental workload, frustration) with the greatest influence on performance.
- The relationship between workload, usability, treatment, and their interactions on performance.
- Factors influencing within-group performance variance during the recall phase.

Qualitative

Qualitative feedback from participants is an essential complement to the other findings. Through thematic analysis, we aim to systematically identify and interpret patterns in comments gathered during exit interviews and other interactions throughout the study. These insights are crucial for integrating and enriching our findings beyond what quantitative data alone can reveal.

4.4 Experiments

This study is organized into two sessions, encompassing three distinct experiments, each aimed at evaluating one of the three measures of training effectiveness described above. This section details the methods, variables, and rationale behind the design of each experiment. For a step-by-step description of the conduct of each session, see Section 4.8 and Section 4.9.

4.4.1 Surveys and Instruments

Two surveys and three instruments were completed during the course of this study, all during the first session. Each is described below, and copies are included in the IRB approval forms.

Participant Intake Form

The *Participant Intake Form* (PIF) is a survey designed to gather information, essential to understanding the participant demographic makeup, assessing the generalizability of the study, identifying potential confounding factors, and facilitating post-hoc analyses involving matching or grouping:

1. General demographics, including gender, age, height, race, ethnicity, country of origin, primary language, education level, and student status.
2. The presence of color blindness.
3. The need for corrective lenses, and whether they will be used during the experiment.
4. Any other condition that might affect their performance during the study.
5. Self-rated experience with LEGO building and background in manufacturing.
6. The method by which they learned about the study.

General Feedback Sheet

This simple form is used to record each participant's qualitative feedback on their overall experience. It is administered by the PI in a very open-ended manner. A list of standard questions is available for participants that aren't forthcoming or otherwise benefit from prompting. It is only used at the end of the first session.

NASA Task Load Index

As discussed in Section 2.13.2, the *NASA Task Load Index* (TLX, Hart, 2006) was designed to assess the perceived workload of a task. The primary outcome of the TLX is a weighted average of six factors that contribute to overall workload. Three of those factors are related to the mental, physical, and temporal demands placed on the participant by a task. The remaining three measure the participant's perceived effort, frustration, and performance during it.

The TLX is comprised of weighting and scoring processes that are repeated by participants upon task completion. First, to account for the subjective nature of workload, each factor is weighted by the participant. This *Sources of Workload Evaluation* is accomplished by having them indicate which element of each pair made the greatest contribution to their perceived workload. This is illustrated in Figure 4.13. For example, given the pair "Mental

Demand or Effort," a participant would likely indicate *effort* for a task like lawn work, but *mental demand* for a philosophical debate. This is repeated for all fifteen possible combinations of the six factors.

Effort or **Performance**	**Temporal Demand** or **Frustration**	**Physical Demand** or **Performance**	**Temporal Demand** or **Mental Demand**	**Mental Demand** or **Physical Demand**
Temporal Demand or **Effort**	**Physical Demand** or **Frustration**	**Frustration** or **Effort**	**Performance** or **Mental Demand**	**Effort** or **Physical Demand**
Performance or **Frustration**	**Physical Demand** or **Temporal Demand**	**Performance** or **Temporal Demand**	**Mental Demand** or **Effort**	**Frustration** or **Mental Demand**

Figure 4.13: Sources of Workload Evaluation

Next, participants complete the *Workload Rating Scales* form to assess the magnitude of each factor for the given task. This uses a Likert-like (Likert, 1932) scale with 20 equal intervals and bipolar descriptors. No numeric values are given and participants are instructed to mark it freely. The mental demand rating scale shown in Figure 4.14 is representative.

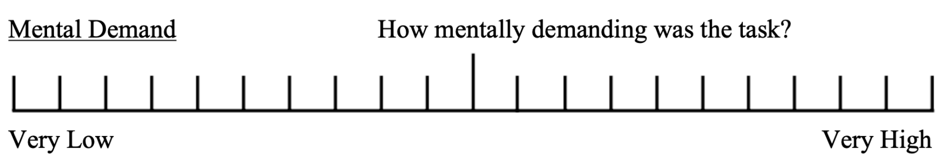


Figure 4.14: Workload Rating Scale, Mental Demand

The results of the TLX are expected to provide insight into the overall user experience for each treatment, along with how workload varies across treatments and what underlying factors contribute to both. This will aid in contextualizing the positive or negative influence that workload might have on learning outcomes.

System Usability Scale (SUS)

As discussed in Section 2.13.2, the *System Usability Scale* is a widely used instrument designed to quickly and reliably assess the overall usability of a product or service. The SUS is composed of 10 questions that participants respond to according to their experience with the system. Each is ranked on a five point scale anchored by "Strongly Agree" and "Strongly

Disagree" bipolar descriptors. Note that alternating items are reverse-scored to improve validity. A portion of the form used is shown in Figure 4.15.

		Strongly Agree				Strongly Disagree
1	I think that I would like to use this system frequently.	1	2	3	4	5
2	I found the system unnecessarily complex.	1	2	3	4	5
3	I thought the system was easy to use.	1	2	3	4	5

Figure 4.15: System Usability Scale, Questions 1-3 of 10

The SUS score is calculated by summing the scores of each question. For odd-numbered questions, the formula is the response value minus one; for even-numbered questions, it is five minus the response value. The total is then multiplied by 2.5:

$$S = 2.5 \times \left(\sum\nolimits_{i,\,\mathrm{odd}} (R_i - 1) + \sum\nolimits_{i,\,\mathrm{even}} (5 - R_i) \right) \tag{4.6}$$

where R_i represents the response to each of the ten items. The SUS score, S, calculated by Equation 4.6 will range from zero to 100. Values above 68 are generally understood to represent above-average usability.

Behavioral Control Survey

The final instrument completed during the intake process was the Adult ADHD Self-Report Scale (ASRSv1.1, Green et al., 2019). Referred to simply as a *Behavioral Control Survey* (BCS) to avoid the possibility of biasing participant responses, the data collected are the focus of a separate study (Ballard et al., 2024), and not further discussed in this work.

4.4.2 First Session - Learning and Recall

The first session of the study tested its primary research questions in two phases, as described below.

Learning Phase

Phase one, Learning, compared the effects of the assigned treatment on the speed and accuracy with which participants performed each repetition of the task. Speed was measured as task completion time, while the number and type of uncorrected errors were used to represent accuracy. These measures, tracked for each assembly during the 10-minute session, were later used to assess both the learning rate and depth of skill acquisition. This approach offers insights into how efficiently and effectively each treatment imparts necessary skills and knowledge for the task.

Participants were instructed to focus on three priorities during the first phase: (1) learn the steps of the assembly process, (2) complete each assembly correctly, and (3) finish as many assemblies as possible in the time allowed. This approach of prioritizing correct and efficient work is in line with measures observed and the principles of OEE, which is used in the second and third phases of this study.

The fixed duration of 10-minutes was deliberately chosen to align with various aspects of the learning process and the operational context of the task. In addition to the OEE considerations outlined above, using a fixed duration ensured that all participants have the same opportunity to learn the procedure. A fixed car count approach (e.g., "make 6 cars") was rejected due to concerns that the absence of a time constraint could lead to uneven learning opportunities and extend session lengths beyond practical limits.

A 10-minute timeline was chosen based on prior experience with learning curves for the task. The Lean Lab is designed around a 60-second takt time, which constrains the work content for experienced operators at each station. Before training, the time to complete these tasks varies widely. Our expectation was that participants would typically complete between three and six cars during their 10-minute session, but any individual participant might complete only one or as many as ten cars.

Recall Phase

The second phase was designed to assess the residual effects of the instructional treatment on each participant's ability to perform the task correctly and efficiently. The assembly task was repeated four times in the control condition, and the same measures were recorded.

Participants were given three priorities: (1) deliver error-free results, (2) reference the work instructions only if necessary, and (3) work quickly. This emphasized working from memory with expectations appropriate for an operator in training. Although participants were encouraged to work quickly, no time limit was set to ensure each produced four complete assemblies.

First Session Data Collection

Manual data collection was limited during the learning and recall experiments. Each assembly was reviewed for correctness and the number and types of uncorrected errors were recorded on the appropriate data sheet. For incomplete (time expired) or retired (breakage requiring rework) assemblies, the final part count was also recorded.

The majority of the data was collected from photographs and video recordings. For each participant, both experiments were recorded on a pair of cameras. One, integrated into the HL2, provided a clear view of the process from the participant's perspective. The other camera was positioned and oriented to record the entire work area from the operator's left, as seen in Figure 4.16.

(a) Side Camera Positioning (b) Side Camera Orientation and FOV

Figure 4.16: Side Camera Setup

Photos were also taken to provide a detailed, high-quality record of the results for each experiment, complementing the video data. This is exemplified by Figure 4.17, which shows that participant #1053 completed three cars in the learning phase. The 4th car pictured here is rotated to indicate that it was incomplete or retired when time expired. Laminated treatment slates are included in these photos to easily embed essential metadata.

Figure 4.17: Sample Learning Result Photo

Subsequent analyses, as detailed in Section 4.12, confirmed the original results and extracted additional data related to timing, error type, PWI usage during recall, and more. While video review was time consuming, this approach allowed us to focus on administering the experiment correctly and carefully observing the participant without the distraction of data collection. This ultimately improved the accuracy and traceability of the results.

The TLX and SUS instruments were both administered twice during this session, once after each experiment. This gave us workload and usability information for all treatment groups during the learning task and for all participants during recall. The latter could be used as baseline measurements for the ST-8 work content.

The PIF, BCS, and a trial run of the TLX were also administered during the intake process of this session. Finally, during the exit interview, general feedback was collected. All of this is detailed in Section 4.8.

4.4.3 Second Session - Retention

The second session took place in the Lean Lab several weeks after the learning and recall experiments, as part of an end of study event. All prior participants were invited to attend. To address confidentiality concerns that might arise due to the public nature of this event, attendance was entirely voluntary.

Based on the number of trials in the first session, the maximum anticipated turnout during this 4-hour event was 40. That suggested a traffic rate of 10 to 15 participants per hour, implying a conservative maximum duration of 3 minutes per trial. This necessarily limited the scope and complexity of the experiment.

This experiment was designed to assess longer-term benefits of the original intervention. Each participant was asked to build a single car from memory, without additional instruction. They were asked to prioritize (1) completing the assembly correctly and (2) finish within the station takt time of 60 seconds. Task completion time was collected, but a generous 3-minute time limit was imposed in the interest of expediency. Compensation was awarded as described in Section 4.6.2.

The interval between the original intervention and this session varied for each participant. This variable time gap, which could range from one to seven weeks, will be a consideration during analysis.

Second Session Data Collection

As in the first session, the emphasis was on minimizing the amount of manual data collection required. Each trial was recorded using only a compact forehead-mounted video camera. Task completion time and error count, along with any essential observations, were documented. The completed car was also photographed, ensuring that the timer and notes were visible in the frame.

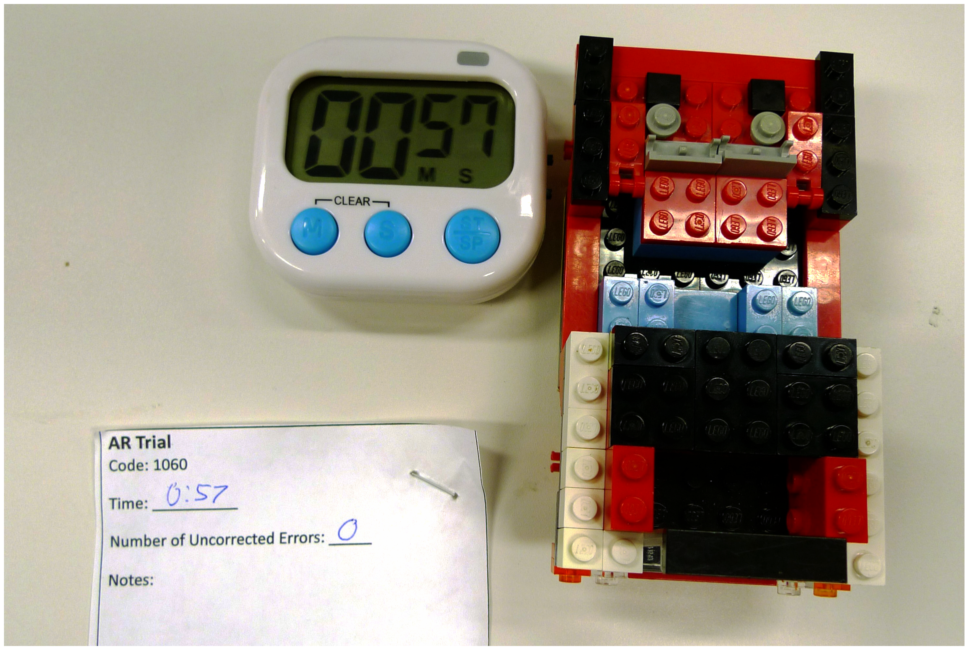


Figure 4.18: Sample Retention Result Photo

This single photo approach streamlined data collection during the event without compromising the integrity of the results. While the photos alone provide sufficient data for our primary analysis, the video recordings offer additional detail if desired.

4.5 HMD System Development

In this section, we will explore the key aspects of the system development process. We will discuss the hardware and software considerations and provide an overview of the development timeline, strategies, and tactics employed. Finally, we will highlight the various challenges encountered during the development process, along with the lessons learned and strategies used to overcome these obstacles.

4.5.1 Hardware

Of the HMDAR systems commercially available in 2021, when the precursor to this study began, only the HL2 was well-suited for enterprise applications, including manufacturing. Originally released for that market in 2019, the HL2's distinguishing features are tabulated in Table 4.2, below.

Table 4.2: HoloLens2's Distinguishing Features[6]

Feature	Description
Natural Field of View	The HL2 is an optical-see-through system (OST) where its display is overlayed on the user's normal view of the world. Its optical design provides a wide and minimally obstructed field of view (FOV), giving a natural and safe user experience. The HL2 also can be used with glasses and features a unique, flip-up design that eliminates the device from view.
Hands-free use	The HL2 relies entirely on natural inputs - gestures and voice controls - rather than physical input devices like tablets or game-style controllers. This leaves operators free to use their hands for their required tasks.
Untethered	The HL2 is a stand alone design with integrated power and compute. This eliminates the need for power or data connections that can encumber users, limit their motion, and introduce tripping hazards.
General Purpose	HL2 is a versatile XR device that supports open development across various industries. Its hardware supports image, model, and area tracking methods to allow for a wide range of augmentation. Unlike some devices, the HL2 is not limited to specific applications or development by authorized partners.

Only the Magic Leap 2[7] design is similar. However, despite being released nearly three years later, it has a more limited FOV, requires prescription inserts for vision correction, offers less extensive developer support, and has achieved less market adoption than the HL2. Both devices are pictured in Figure 4.19, below.

Because the HL2 represented the state of the art in OST HMD enterprise XR devices when this study commenced, it was adopted for the HMD AR/MR treatments. At the time of this writing, nearly four years after its introduction, the HL2 hardware design and the feature set it enables remain largely unchallenged.

Unlike the LG, the HL2 is entirely self-contained, processing data from an array of sen-

[6] HoloLens2 Hardware Details: https://learn.microsoft.com/en-us/hololens/hololens2-hardware
[7] MagicLeap: https://www.magicleap.com

(a) Microsoft HoloLens2. Source: Microsoft (b) Magic Leap 2. Source: Magic Leap

Figure 4.19: HoloLens 2 and Magic Leap 2

sors to enable six degree-of-freedom visual tracking, spatial mapping, gesture recognition, voice commands, hologram rendering, and optical compositing in real-time. Figure 4.20 provides an exploded view of the HL2 components.

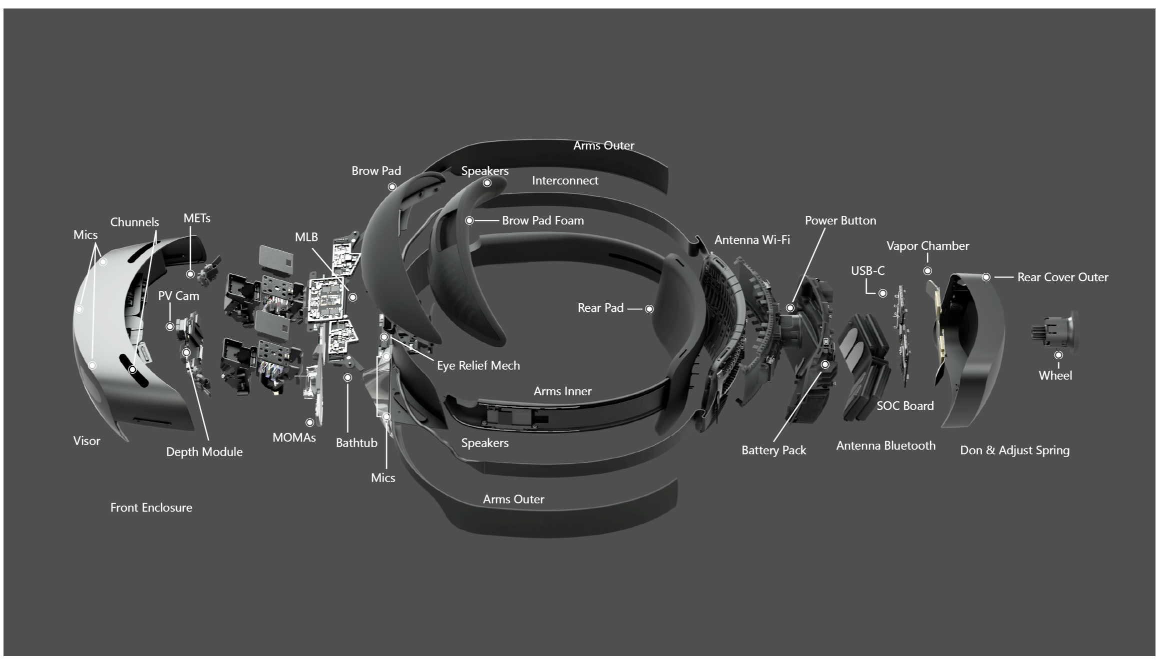


Figure 4.20: HoloLens 2 Exploded View. Source: Microsoft

4.5.2 Software

Development options for XR systems are limited. For creating custom apps on HL2, Microsoft supports and endorses Unity[8] with the Mixed Reality Toolkit[9]. Ultimately, the need for model and area based tracking capabilities required further tooling. Each element is described below.

Unity

Best known as a "game engine," Unity is a very capable tool, commonly used for industrial, commercial, and defense applications. It offers a comprehensive feature set for creating and animating objects, rendering high-quality visuals, programming systems and interactions, simulating physics-based dynamics, processing music and audio, designing user interfaces, and much more.

Despite its complexity, Unity is relatively easy to learn. High quality training, documentation, and support are widely available, both through official channels and from a large and enthusiastic development community. Like most tools of its type, Unity is extremely extensible via plugins, many of which are distributed through the official asset store.

Unity is free for non-commercial and academic applications[10]. For research applications like this one, which will not be distributed or otherwise commercialized, its Terms of Service[11] did not pose major concerns. Crucially, Unity allows users to retain rights to content they create, and makes no ownership claims over it.

Other options, including Epic's Unreal and Vuforia Studio, were carefully considered but ultimately rejected due to a lack of support, functionality, flexibility, or some combination thereof.

Mixed Reality Toolkit (MRTK)

The MRTK is a software development kit (SDK) designed to simplify and accelerate the development for XR hardware, including the HL2. It provides developers access to essential

[8] Unity: https://unity.com/
[9] Mixed Reality Toolkit (MRTK): https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity
[10] Unity Pricing: https://unity.com/pricing#plans-student-and-hobbyist
[11] Unity's Terms of Service and other legal info: https://unity.com/legal

HL2 capabilities, including spatial mapping, hand and eye tracking, natural input modalities, and more, all from within Unity.

MRTK is a free, open-source project that was initiated by Microsoft and first released under the very permissive MIT License (Saltzer, 2020) in 2017[12].

Vuforia Engine

MRTK's built-in tracking support is limited to image-based solutions and spatial anchoring techniques that were insufficient for this project. After exploring available options, it was determined that the Vuforia Engine[13] (VE) by PTC could best address that limitation. This SDK works with Unity, MRTK, and the HL2 to add robust marker, model, and area based tracking capabilities that were necessary for this project.

PTC's Vuforia product line is a commercial product designed for enterprise customers, but the Basic version of the Engine SDK is available at no cost. With that plan a limited number of model and area targets can be generated, so long as the resulting app is not published[14]. Within the constraints of this work, PTC's Terms of Use and Developer Licensing Agreement[15] posed no significant concerns.

4.5.3 Timeline

The software used for both HL2 treatments was based on work originally done during the Summer of 2022 by a team of three undergraduate computer science and software engineering students. Led and directed by the author, that team created an augmented in-situ training prototype for manufacturing operators. The resulting system utilized a HL2 to align in-context instruction with the workpiece using image based methods — QR codes attached to the fixture.

In the following semester the same team adapted and extended the underlying codebase to support this study. This effort primarily consisted of assessing enhanced tracking tools,

[12] MRKT Github Repository and Licensing File: https://github.com/microsoft/MixedRealityToolkit-Unity
[13] Vuforia Engine SDK: https://www.ptc.com/en/products/vuforia/vuforia-engine/ar-app-development
[14] Vuforia Pricing and Licensing: https://developer.vuforia.com/library/faqs/pricing-and-licensing-options
[15] Vuforia Terms of Use and Developer Agreement: https://developer.vuforia.com/legal/tos
https://developer.vuforia.com/legal/vuforia-developer-agreement

integrating the selected SDK, configuring area and model based tracking, and adapting instructional content from the PAR treatment. The project timeline is approximately illustrated by Figure 4.21.

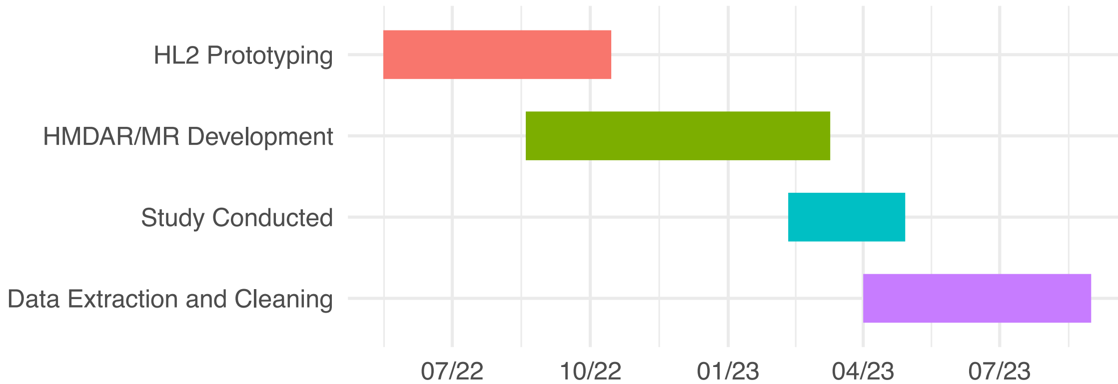


Figure 4.21: Approximate Project Timeline

4.5.4 Strategy

Development was iterative with roughly two-week sprints. Except during the summer, the developers were full-time students and were managed accordingly. Expectations had to reasonably balance their availability and inexperience with the project goals and timeline. Flexibility was critical to everyone's success. Atlassian's Trello[16], a lightweight, web-based project collaboration tool with kanban style task tracking, was used to manage the project.

4.5.5 Tactics

The HMDAR and HMDMR apps were developed in Unity version 2022.x with MRTK v2.7x and VE v10.7x. Development was done entirely on MSI GE76/66 Raider (11UH-053/227) laptops running the 64-bit version of Windows 10 with the latest updates. Both laptop models were equipped with an Intel Core i9-11980HK CPU, NVIDIA GeForce RTX3080 GPU, 32 or 64GB, and 2GB of M.2 NVMe storage.

Systems were configured with the Visual Studio 2019 integrated development environment (VS) as described in the MRTK setup instructions[17]. Unity Version Control (VCS, formerly

[16] Trello: https://trello.com/home
[17] MRTK Setup Instructions:
https://learn.microsoft.com/en-us/windows/mixed-reality/develop/install-the-tools

Plastic) was used to manage all assets and source code, allowing the developers to track changes and collaborate effectively.

All programming was done in C#, as required by Unity. C# is an object-oriented language with strong typing and simplified memory management. It has a familiar syntax that is similar to Java and C++, both of which influenced its design. This enabled the developers, all new to C#, to adapt with relative ease.

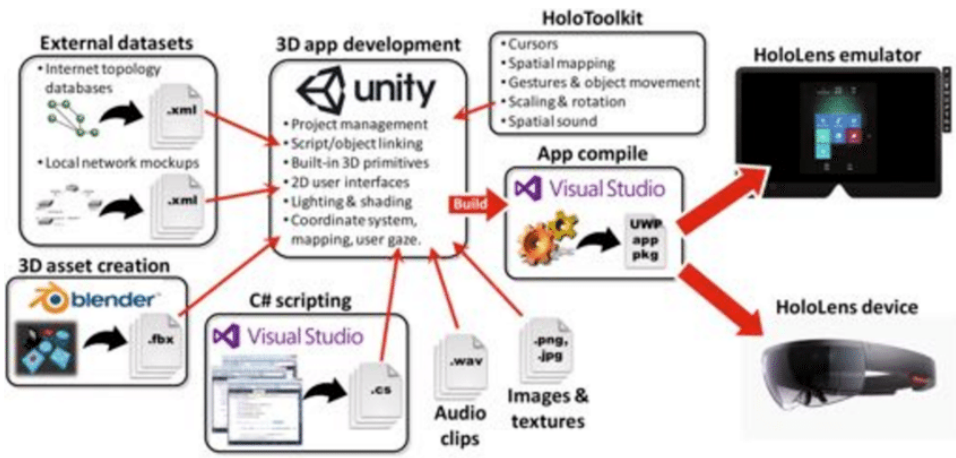


Figure 4.22: Development Workflow

Unity development for HoloLens2 generally proceeds as illustrated in Figure 4.22, and described below.

1. Configure a scene in the Unity editor, including the placement, orientation, and properties of rendered items and UI components, along with cameras, lights, and other "helper objects" used to manage the experience.

2. Write C# scripts to control scene interactions. These scripts are attached to game objects and get events, trigger responses, and pass messages to other objects. Unity's component-driven architecture offers a variety of predefined methods and event functions, which is extended by MRTK.

3. Do initial testing, using the Unity editor to simulate HL2 interactions directly on the laptop screens. This approach allows for rapid iteration and real-time feedback to changes made in the editor, facilitating early debugging.

4. Building for the HL2 device is a two step process. First, Unity generates a bundle of processed data and scripts. The result is then used by VS to compile and package a UWP (Universal Windows Platform) app for the HL2. In the process all C# scripts

are converted into C++, and then into a native binary for the HL2's ARM-based architecture.

5. Finally, the UWP app is transferred to the HL2 via USB.

The inter-dependencies between Unity, MRTK, and VS, each with their own packaging systems, dependencies, and versions, made this an intricate process. Our initial understanding was greatly accelerated by Microsoft's online resources[18] and tutorials[19], along with Ferrone's annually updated *Learning C# by Developing Games in Unity* (2021).

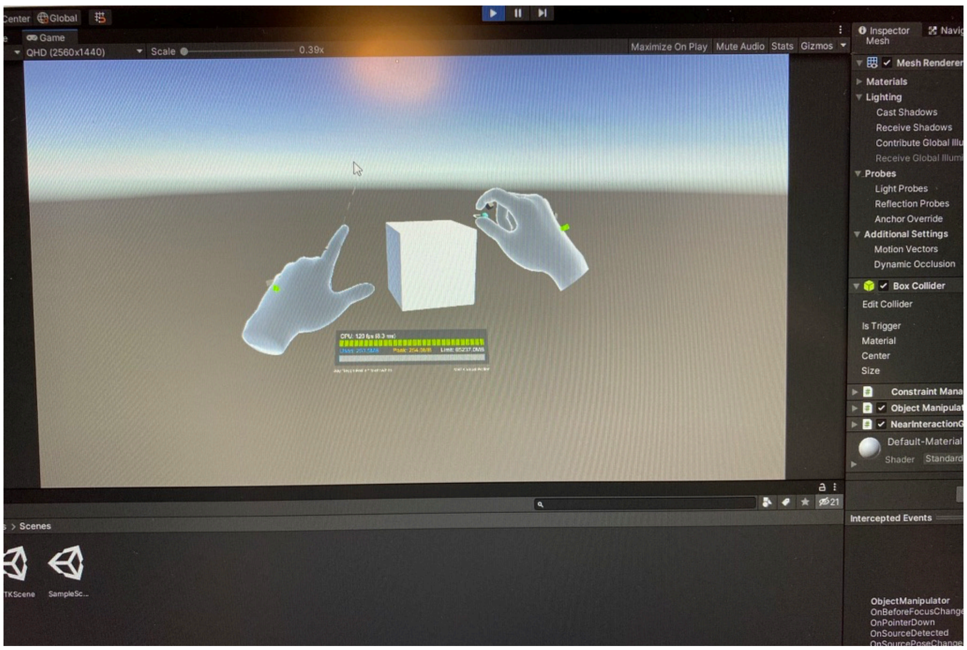


Figure 4.23: Unity HoloLens2 Simulator

4.5.6 Design

HMDAR Treatment

The HMDAR experience was analogous to that of the PAR treatment. Interactions in both were controlled by hand placement / motion in the scene, mimicking button inputs. Both used a fixture to keep the model in the reference position, facilitating the alignment of virtual objects. Essentially, both were projected AR experiences, differentiated primarily by the manner of projection and input detection.

[18] MRTK Unity Documentation: https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2

[19] MRTK Tutorials: https://learn.microsoft.com/en-us/training/modules/learn-mrtk-tutorials/

The PAR system used traditional optics to project virtual instructions onto the work surface and a depth-sensing camera to detect inputs. "Tracking" in this case was static and mechanically set, where the projector's orientation, throw angle, and focal length were fixed based on its location relative to the workpiece.

The HMDAR system used the HL2 to project virtual instructions into the operator's visual field via a sophisticated sensing, display, imaging, and optical systems. This user-centric display is the affordance that differentiates the HMDAR treatment. Tracking was dynamic, based on the operator's position and head angle relative to the work surface, and intrinsic properties of the HL2 system. Input detection utilized the system's hand tracking capabilities, which rely on fused sensor data and machine learning techniques.

HMDMR Treatment

A consistent approach was taken for the HMDMR treatment design, extending the capabilities of the HMDAR treatment. HMDMR used more sophisticated tracking methods to enable freeform interaction with the workpiece, as described in Section 4.2.3.

4.5.7 Implementation

The HMDAR treatment was developed first. Using the HMDAR version as a baseline, the HMDMR version extended it to incorporate model based tracking methods.

HMDAR Implementation

This effort primarily involved: (1) recreating the PAR's instructional design approach, (2) setting up the interaction methods, and (3) implementing the tracking system.

The first was relatively straightforward. Assets were modeled in Unity to resemble components from the PAR instructions. They were arranged in the scene relative to the workpiece location, based on an established scale. Finally, scripts were written to control the scenario logic, changing the scene based on user behavior.

Our interaction implementation relied on HL2 input systems provided by MRTK's modular, component-based architecture. In this system, input actions like `select` or `activate` and the events they trigger act as the bridge between the user's physical actions and the software's response. Physical actions are captured via HL2 sensors and interpreted by MRTK's various modalities, including gesture and hand tracking. Different interaction styles are associated with available pointer types, e.g., ray pointers for distant interactions and poke pointers for nearby. With this approach a diverse range of interactions can be design through careful composition of components.

For HMDAR tracking, we relied on VE's *Area Targets*[20] feature, an implementation of the area-based tracking methods described in Section 2.7.4. Area-based tracking compares a pre-generated 3D model of the workspace with live 3D data of the user's surroundings. This "spatial map" is a polygonal mesh generated in real-time by the HL2, using data from the depth sensor, visible light cameras, and inertial measurement unit. A sample is shown in Figure 4.24. From this comparison, the system can determine the current position and angle of the user's head.

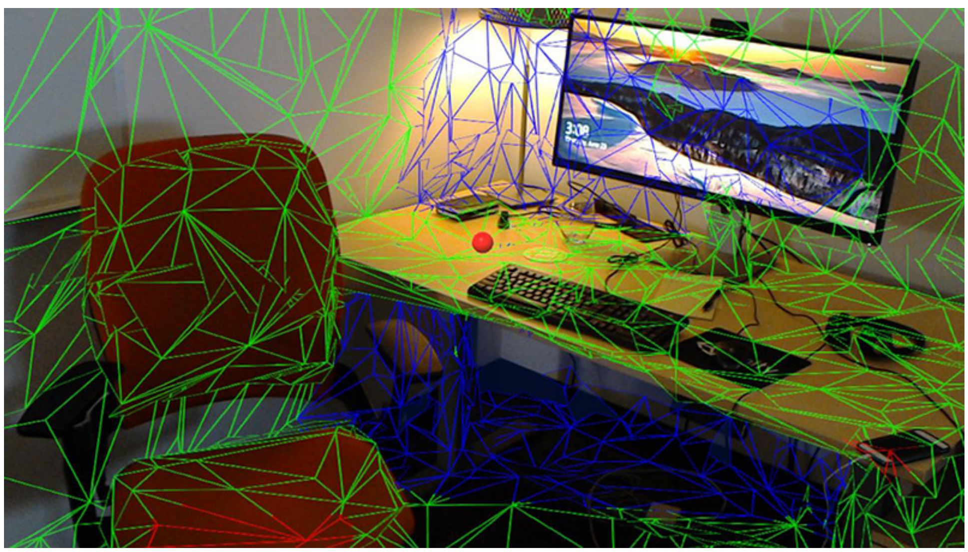


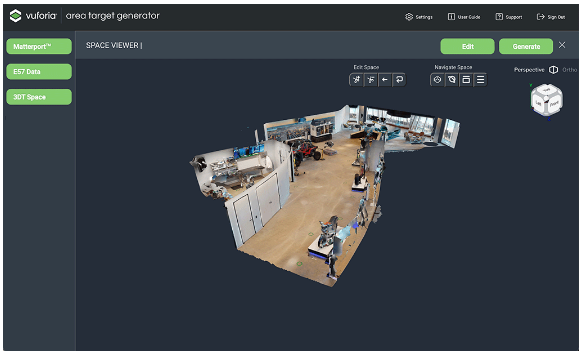
Figure 4.24: HoloLens Spatial Mapping

Area-based tracking was selected over image or marker-based methods to minimize tracking loss. Marker-based tracking will fail if there are no markers within view of the HMD's sensors. Area-based tracking provides a continuous map of registration points for the entire scene, greatly reducing drop-outs. It also tends to (re)acquire tracking more quickly than marker-based methods, again due to the number and distribution of features available.

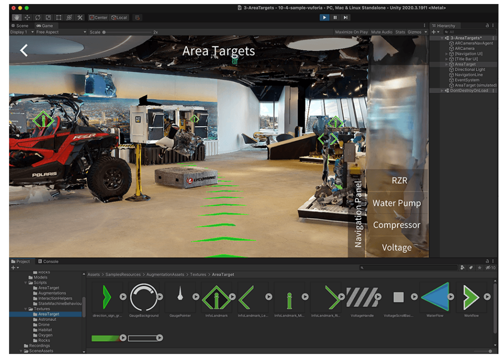
The 3D model is created offline, first by scanning the area with the Vuforia Creator App[21]. Pictured in Figure 4.25a, this tool uses a LiDAR equipped iPhone / iPad Pro to capture an accurate model of the area in E57 format, per ASTM E2807 (ASTM, 2019). The E57 data is

[20] Vuforia Area Targets: https://developer.vuforia.com/library/environments/area-targets
[21] Vuforia Creator App: https://developer.vuforia.com/library/tools/creator-app

then processed to generate an Area Target asset package for Unity, which includes all the required geometry, textures, and metadata.



(a) Vuforia Creator App



(b) Area Target in Unity

Figure 4.25: Vuforia Area Target Processing

Once imported into Unity and properly configured[22], the 3D model is used to align the placement of virtual objects in the real world scene. This allows the developer to design the experience in the context of the real world model, as seen in Figure 4.25b.

Finally, at runtime, key points encoded from the Area Target mesh are compared with the real-time spatial map to estimate the operator's head position and angle relative to the workpiece, thus establishing a coordinate system for the spatially coherent placement of virtual objects.

HMDMR Implementation

To support freeform interaction, the system needed a way to properly place instructional cues on the workpiece, regardless of its position and orientation. This was accomplished with model-based tracking, using the Model Targets[23] feature from VE. Note that this treatment continues to use Area Targets for workspace pose estimation, but adds Model Targets for those involving the workpiece. Otherwise, the HMDMR implementation is unaltered.

The distinct technical requirements of model and area based tracking necessitate separate implementations. Where area-based tracking focuses on spatial orientation within a static environment, the prime challenge for model-based tracking is dynamic object recognition and pose estimation. While there might be some overlap in the fundamental computer

[22] Vuforia Area Targets in Unity: https://developer.vuforia.com/library/develop-area-targets/area-targets-unity
[23] Vuforia Model Targets: https://developer.vuforia.com/library/objects/model-targets

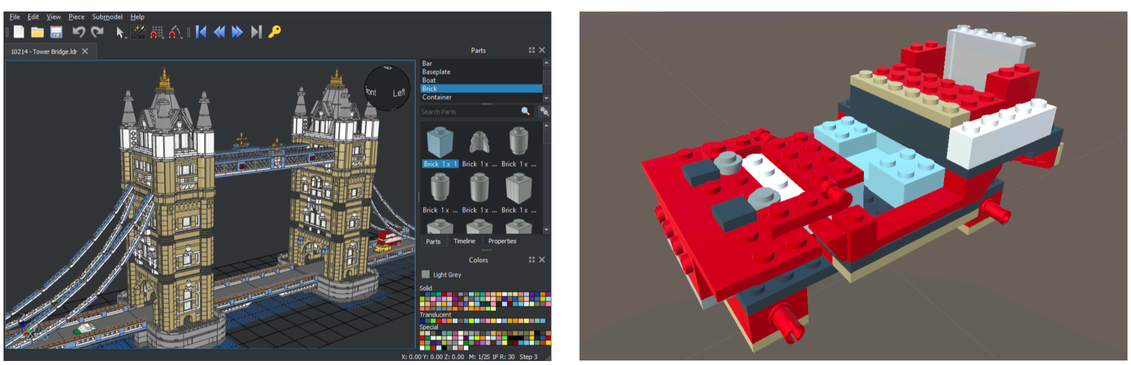
vision and machine learning techniques used, the specific algorithms and their optimization differ significantly, catering to the unique challenges of each tracking type.

Overall, the Model Target implementation process was similar to that for Area Targets. First, a 3D model of the object was constructed and converted into a Model Target using VE tools. These steps were completed offline. At runtime, the system again compares live sensor data with Model Target data to recognize and then track object(s) in the scene.

Model source data was constructed in LeoCAD[24], a tool for designing models using LEGO bricks. LeoCAD uses the comprehensive LDraw[25] database of LEGO parts, most of which are modeled from actual pieces. This standard uses a proprietary unit of measure, the LDraw Unit (LDU), which is based on the smallest stud-to-stud spacing on a standard LEGO brick: 1 LDU = 0.4mm.

LeoCAD's design adheres to LEGO design principles, ensuring that the way parts attach in the software reflects the real-world equivalent. The combination of LDraw's precision and LeoCAD's tooling ensures that the resulting models are faithful representations of their physical counterparts.

Figure 4.26a is a screenshot of the LeoCAD interface, highlighting some of its capabilities. Both LDraw and LeoCAD are unofficial, open source, community run, multi-platform tools that are free to use.



(a) LeoCAD Interface Screenshot (b) Model Target Constructed in LeoCAD

Figure 4.26: LeoCAD UI and Model Target

Our source model was constructed in LeoCAD based on the PWI, using LDraw models of the specified LEGO parts. The result, seen in Figure 4.26b, was exported as a 3DS[26] file and im-

[24] LeoCAD: https://www.leocad.org/
[25] LDraw: https://ldraw.org/
[26] 3DS File Type: https://en.wikipedia.org/wiki/.3ds

ported into Unity as a part hierarchy. There, it was scaled from LDUs to conventional units and exported in the FBX[27] interchange format for VE's Model Target Generator (MTG).

The MTG supports two Model Target types. *Advanced Model Targets* can be automatically recognized and tracked from any angle, without the manual alignment and recognition process required by *Standard Model Targets*. The requirement for the HMDMR treatment to provide freeform interactions in an otherwise equivalent experience mandated the use of Advanced Model Targets.

The MTG process begins by checking for model suitability. Model targets must be free of errors and unnecessary internal geometry, with rigid geometry and real-world scale. Ideal candidates for Advanced tracking also feature optically stable surface features, minimal symmetry, and accurate surface colors. Highly reflective, transparent, or featureless surfaces provide insufficient visual cues and highly symmetric objects make it difficult to determine orientation.

Next, the model geometry is analyzed and a set of Guide Views are automatically generated for a 360-degree recognition range. This step leverages deep learning methods trained to generate optimal views from arbitrary angles based on the geometric features and surface qualities of the model. The output of this process includes the trained recognition model, along with the associated dataset and guide views. This package is imported into Unity and used by VE to provide the desired model-tracking functionality.

4.5.8 Development Challenges

Software system development is always challenging and this work was no exception. The system requirements and complexity, hardware and software issues, and resource constraints all contributed to a variety of challenges that the team overcame.

A fundamental consideration in the success of any collaborative development effort is a reliable software environment. The integration of Unity, MRTK, and Vuforia, along with their requisite packages and settings, was intricate and fragile. This made it difficult to ensure a reliable baseline and consistent results for all developers. Even with version control, too much time was spent chasing bugs and deployment issues rooted in these issues.

Implementing accurate and robust tracking for the AR/MR applications posed additional challenges beyond the selection and integration of another third-party framework (Vuforia

[27] FBX File Type: https://en.wikipedia.org/wiki/FBX

Engine). Area-based tracking was mildly sensitive to the surface properties of some materials and workstation configuration changes, requiring some additional care in setup. In particular, the exterior of the fixture had to be masked with painter's tape to limit the reflectance of its milled aluminum surfaces. These detractions were more than offset by the overall robustness of the method, which all but guaranteed nearly instant acquisition of tracking from the rich set of provided features.

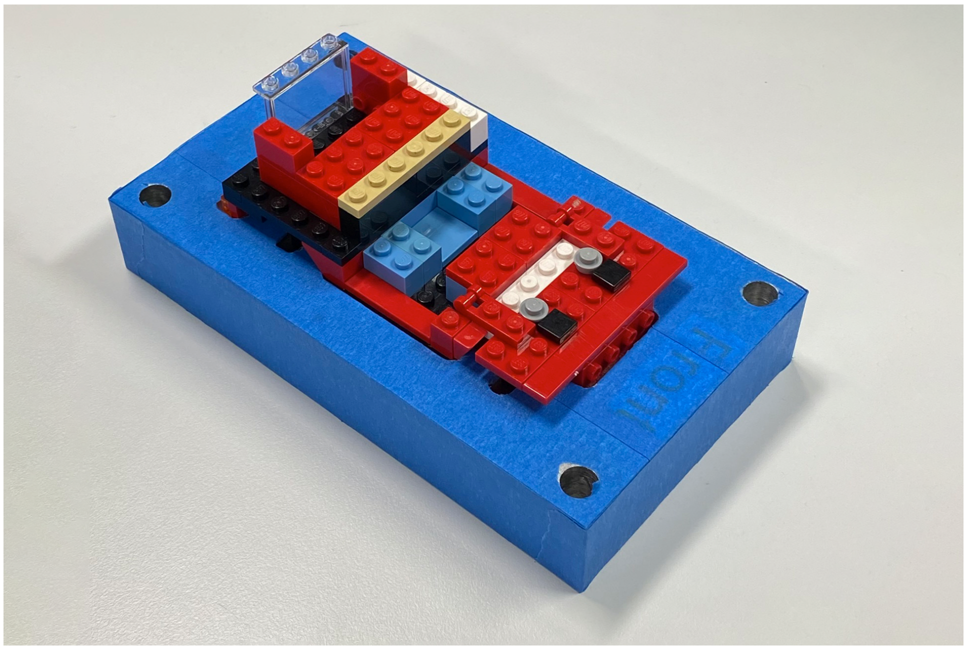


Figure 4.27: Fixture Masked to Reduce Reflectance

The inherent complexity of MR also led to less reliable tracking and increased instability in that treatment. When tracking is interrupted the system is unable to place virtual objects in the scene. The user experiences this as a "drop-out," where everything disappears from their field of view. Once tracking is reacquired, the virtual objects return. In rare cases of extreme instability, this cycle could cause the system to crash or require a reset. As described in Section 4.11.5, the frequency, duration, and impact of these events varied, and was accounted for.

Working with expensive, body-worn hardware created additional complications. Final testing could only be done on the HL2, which requires physical access to the device. The time required to build and deploy an update to the HL2 significantly slows iteration, adding to the disruptive nature of regularly donning/doffing a shared headset. These factors created friction that naturally led developers to prefer testing via simulation or emulation, neither of which provides a complete or accurate view of user experience or system performance.

Developing AR/MR applications for the HL2 requires careful design and optimization. To avoid related discomfort concerns, smooth performance and minimal latency must take priority. Considering the device's computational constraints and the complexity of the required tracking and rendering tasks, this may require tradeoffs in the instructional design. Thankfully, the design of the PWI and "low fidelity" nature of LEGO bricks helped mitigate these challenges.

Finally, reliance on a student-led development effort demanded a flexible and supportive management approach from recruitment to completion. None of the primary contributors had previous experience with the software, hardware, methods, or tools involved. Despite that, they built a successful instrument from scratch. Beyond hard work, skill, and determination, their success owes something to a careful management of project requirements, system capabilities, and resource constraints in the academic context.

A number of other challenges commonly associated with software development projects were largely avoided through careful process planning, iterative development, and ample time allocated to training early in the project.

4.6 Participants

This section will outline all participant recruitment, selection, benefits, and assignment considerations. The onboarding process is also described.

4.6.1 Recruitment and Selection

A convenience sample of participants were recruited from the Auburn University community using digital and printed promotions around campus, the graduate school mailing list, and outreach in various undergraduate engineering classrooms. The latter focused on freshman and sophomore engineering students in Industrial & Systems Engineering, as they are accessible and most likely to meet all requirements. Figure 4.28 exemplifies the recruiting materials, approved copies of which are included in Appendix E, IRB Documentation.

Potential participants in the first investigation were screened for exclusion based on their age (under 19) or a tendency to motion sickness. Additionally, they were screened for experience with head-mounted or projected AR devices using gesture based controls. This does not exclude those having experience with VR headsets like META's Occulus product

Figure 4.28: Digital recruitment flyer distributed on campus monitors.

line, which is relatively common among the target population, but utilize traditional input devices.[28] Finally, any candidate that had previously participated in a manufacturing simulation at the Tiger Motors Lean Education Center was excluded. This controlled for prior experience with the assembly task, as part of INSY 5800/6800 or otherwise.

Due to practical constraints described in Section 4.10.2, the study could accommodate maximum of 100 participants. Realistically, 70-90 were expected based on early response. For the results to exceed the minimum desired statistical power, at least 60 participants were required. Based on these considerations, the recruitment goal was 70+ participants.

4.6.2 Benefits and Compensation

Other than the compensation described below, there were no direct benefits for participants in this study. All were offered an opportunity to interact with projection and/or head-mounted AR hardware and training methods for the first time. This may have lead them to a greater appreciation of the benefits and opportunities these technologies offer.

To incentivize sign-ups, some extra credit and the possibility of financial compensation were offered. In addition, all participants were invited to an event at the end of the study.

[28] This distinction had to be explained to several interested participants, which supports the notion that the differences between AR and VR are not yet well understood. Ultimately, no participants were eliminated by this requirement.

Extra Credit

Any instructor promoting these studies to their students was free to provide extra credit for participation. This was entirely at their discretion. We cautioned all instructors to offer no more than 1% on the final class average, and encouraged them to provide alternative bonuses for students unable to participate.

End of Study Event

At the conclusion of the main study, all participants were invited to return to the lab for an "open house" event. This gave them the opportunity to experience other treatments and related technologies, and learn more about the experiment and lab. Food and drink were provided. In exchange, we asked all attendees to participate in a brief retention experiment. Attendance and participation were voluntary.

Compensation

The possibility of financial compensation was introduced in the final IRB Modification, submitted April 3, 2023. Following its approval on April 10, we began promoting this retroactive benefit. All participants in the main study were eligible for one of three random drawings. Those that attended the end of study open house qualified for additional awards, as outlined in Table 4.3.

Table 4.3: Compensation drawings by study and award category.

Study	Category	Quantity	Amount (each)	Sub-Total
Main	Participation	6	$25	$150
Retention	Performance	1	$50	$50
Retention	Participation	4	$25	$100

For the performance prize in the retention study, eligibility was limited to those who completed the experiment in under one minute without errors. A total of $300 was awarded via email in the form of digital Amazon Gift Cards.

No member of the research team was eligible for any of the financial compensation described, and all payment processing was handled by appropriate members of the ISE staff.

4.6.3 Onboarding

The initial participant onboarding process was conducted manually by the PI. In a phone call with interested participants the PI would (1) briefly explain the investigation, recapping and elaborating on the recruiting materials; (2) discuss the exclusion criteria and identify relevant issues for the candidate; (3) set expectations for participant involvement, including time commitment and tasks; and (4) answer any questions the candidate had regarding participation in the investigation.

If the candidate indicated a willingness to proceed, their information was collected using the Subject Recruitment Data Sheet included in Appendix E. A unique participant ID was logged on the code list and a date and time for data collection were then assigned. As detailed below, the code list and consent form provided the only link between personally identifiable information and experimental data. Afterwards, a confirmation email was sent. A copy of the Informed Consent form was included for their review prior to the appointment.

This process quickly proved impractical, and a self-service web-based alternative was offered. By eliminating the reliance on manual, call-based screening, SignUpGenius[29] streamlined the entire onboarding and scheduling process. This increased the rate for converting interested into scheduled participants and allowed the team to focus on running the experiments.

4.6.4 Random Assignment

Treatment assignment was accomplished through a combination of participant scheduling and treatment ordering. Participants set their own appointments based on availability. Without knowledge of the underlying treatments or their ordering, this was an inherently random process. During the intake process, each was assigned the next available treatment from a randomly ordered list.

Treatment randomization was completed before the onboarding process began. To ensure that all treatments were tested at the start of the experiment, a random sequence of all four treatments started the order. Next, a set of eight treatments, including two of each type, was shuffled to create a randomly ordered batch with an even distribution. This block-wise process was repeated as necessary to cover the maximum number of participants. The batches

[29] SignUpGenius: https://signupgenius.com

were combined, in the order generated, to create the final treatment sequence. This approach ensured random, balanced, and unbiased assignment, regardless of the final number of participants.

The treatment randomization process was implemented with a simple Python function, `gen_treatment_order()`, as seen in Figure 4.29, below. The `random.shuffle`[30] function from the base Python 3.x distribution was used to randomly reorder a group of values. Though this program was only run once (after validating and verifying its output), no random seed was set to ensure that a unique sequence was generated with each use.

```
def gen_treatment_order(n=2):
    '''
    generate random treatment order

    IMTS is a list of the four treatments
    NUM_CYCLES is the number of 4-treat batches
    '''

    # start with random selection of all treats
    trials = []
    first_set = IMTS.copy()
    random.shuffle(first_set)
    trials.extend(first_set)

    # shuffle IMTS in groups of 8 (two cycles)
    for _ in range(NUM_CYCLES // n):
        part_trial = IMTS * n
        random.shuffle(part_trial)
        trials.extend(part_trial)

    return trials
```

Figure 4.29: Python 3.x Code for Treatment Randomization

4.7 Research Compliance

As with any protocol that involves human participants, this study required Institutional Review Board (IRB) approval. This section will describe that process, and detail considerations related to consent, privacy, security, and risks / discomforts identified.

[30] `random.shuffle` Documentation: https://docs.python.org/3/library/random.html#random.shuffle

4.7.1 Institutional Review Board

An initial review of the protocol found that it created minimal risk to participants, did not involve vulnerable populations, invasive methods, or sensitive data, and required informed consent. Furthermore, it involved the "collection of data from voice, video, digital, or image recordings," which is identified by the University as a category of research eligible for expedited review.[31]

The IRB submission and expedited review process began in December of 2022, and final approval was granted on April 10, 2023. During that time four versions of the IRB were approved, and only two rejected. The approval dates and changes are summarized in Table 4.4.

Table 4.4: IRB Version History

Version	Description	Approved
1.1	Original submission	1/30/23
1.1a1	Adjusted protocol to add survey instruments	2/13/23
2.1	Incorporated 2nd investigation	2/23/23
3.0	Added compensation	4/10/23

This process was somewhat complicated by the decision to incorporate two separate but related investigations into a single application. The first investigation (I1) is the focus of this dissertation. I2 (the second investigation) is a separate work that used similar methods to investigate the relationship between I4.0 technologies and Lean Manufacturing systems. There is no connection between I1 and I2 beyond the collaborative relationship between their research teams.

The development of the IRB was a collaborative effort which I spearheaded as the Principle Investigator (PI), authoring the majority of the application, and ensuring alignment with the study's goals and ethical standards. Contributions from other members of the research teams, as specified in Appendix A: Team Contribution Matrix, were essential to its thorough design and timely approval.

[31] Per AU IRB Expedited Category Guidance: https://cws.auburn.edu/shared/files?id=159&filename=AU%20Expedited%20Categories%20Guidance.doc

The final approved version, including all supporting materials, is incorporated as Appendix E: Institutional Review Board Approval. All processes described herein are all based on the approved protocols.

4.7.2 Consent

All participants were provided a copy of the approved informed consent form in advance of their trial. As part of the intake process they were provided a paper copy for further review and encouraged to ask any questions or share any concerns they might have. After a verbal confirmation that the participant had read and is satisfied with the terms of the informed consent agreement, they were asked to sign and date it. The form was then countersigned and placed in a locked filebox.

4.7.3 Privacy and Data Security

A variety of data were collected for this study, including video recordings, performance metrics, demographic information, and survey responses. All data were collected anonymously, referenced only by the unique ID assigned to each participant. The code list, used solely for contacting participants during the ongoing protocol, was securely stored alongside the consent forms in a locked box within a restricted-access location. Notably, the consent forms do not include any reference to the participants' ID numbers. Both the consent forms and the code list are maintained exclusively in paper format to facilitate secure storage and subsequent disposal through shredding. Upon the completion of the protocol, the code list will be destroyed, thus rendering the data completely anonymous. These measures were diligently enforced to protect the privacy and confidentiality of participant data.

All electronic data pertaining to the study are stored on a secure server. Non-identifiable data is available to other members of the research group, for the purposes of approved research, under conditions that ensure continued confidentiality. Access to consent forms and the code list is limited to the PI and, if required, the research committee.

For reasons detailed in the study design, two angles of each trial were recorded on video; a view from the participant's perspective and a side view focused on the work surface. The side view was carefully framed to limit identification of the participant, and later edited to crop out identifying features, ensuring participant privacy. Additionally, all participants were required to wear the HoloLens2 head-mounted display, which further obscured their appearance.

4.7.4 Risks and Discomforts

Beyond the privacy and data concerns previously described, other potential risks and discomforts were identified. Prolonged used of HMD VSTs has been reported to cause mild neck strain, disorientation, and eye strain in some cases. The optical and physical design of these systems can also result in a limited or obscured field of view and degraded acuity, which could increase the participant's risk of trip or impact (UL, 2022).

The HL2's untethered design with wide, unobstructed field of view mitigates many of these concerns, but additional precautions were taken. All participants were screened for a tendency towards motion sickness. The study was intentionally designed to limit each participant's time wearing the HL2, and to ensure that they were generally stationary in an obstruction-free environment. Finally, the Lean Lab was selected in part because it is a organized, safe, and well-lit environment with no history of related hazards.

As these experiments were conducted in the Spring of 2023, the risk of COVID-19 exposure remained a lingering concern. Precautions were implemented during data collection as outlined in the University-provided protocol for studies without high-risk procedures or participants (Category C). All work surfaces and equipment were wiped down before and after each participant, and necessary supplies were made available. All research participants followed the University's guidance on self-screening. Throughout the administration of this study, the CDC's COVID-19 community level for Lee County, Alabama remained LOW, eliminating the need for participant screening. The Shelby Center for Engineering Technology, where this protocol was administered, is assigned the highest level of building readiness due to increased air turn-over and filtration.

This study did not involve any vulnerable populations. Overall, the likelihood and impact of any of the risks outlined above were considered low. Nevertheless, all participant activities were supervised to monitor for likely symptoms or unexpected side-effects. In either event, the experiment would be suspended and the situation assessed. If escalation was deemed necessary, an emergency plan and contact list were available to the research team.

During the post-experiment debriefing all participants were asked about injury and discomfort, and were observed for lingering or delayed effects. Ultimately, only a few mild discomforts were reported, and no significant side effects, injuries, or need for escalation.

4.8 Conduct of First Session

This section will describe key operational details for conducting the first session protocol, including the learning and recall experiments. The conduct of this study was designed to ensure participants felt comfortable and understood the tasks, while also aiming for valid results and thorough data collection. The approach balanced a need for clear procedures and ethical integrity with the flexibility required to adapt to individual participant needs.

4.8.1 Division of Labor

The approach detailed herein was designed for three roles, referred to as the primary, secondary, and tech. The role of the primary investigator (PI) was always played by the author.[32] The secondary investigator (SI) role was typically filled by his counterparts on collaborating studies. Finally, the tech support (TS) role was played by a member of the study's HL2 development team.

The responsibilities of each role vary throughout the session as described below. All three roles were usually filled for each session, allowing for the most comfortable division of labor, but some trials were run effectively with fewer.

4.8.2 Start and End of Day Procedures

Daily preparation for the study involves careful setup across three roles. The PI ensures the schedule and trial documents are ready, workstations and PCs are prepared and operational, and camera equipment is correctly configured and tested. The SI handles the specifics of the trial setup, including treatment slates, car checks, and part inventories, while coordinating with additional support staff. TS focuses on maintaining the software, streaming setup, and hardware sanitation. Together, the team ensures a controlled environment ready for trials.

The PI is also responsible for returning the lab it its original condition at the end of each day. Post-session procedures include resetting all cell two workstations, powering down equipment, cleaning and sanitizing the lab, turning off lights, and securing the premises. Additionally, batteries and memory cards are tended to, and consent forms securely stored.

[32] Except when he was quarantined for a week with COVID, during which PIs from collaborating studies generously substituted. The show must go on!

Figure 4.30: The Battle Wagon

Both setup and tear-down were facilitated by the rolling cart adopted for this study (Figure 4.30), affectionately known as the “battle wagon.”

4.8.3 Workflow and Roles

Workstations seven through nine are configured for the first session as seen in Figure 4.31. The numbered positions indicate the flow of work in process (WIP) through the system. Assemblies begin at (1) and move, sequentially, to ST-8 (2). When a car is completed or retired (due to breakage considered irreparable by the participant), it is moved to either the green (3a) or red tray (3b), respectively. WIP at (3a) and (3b) is promptly collected and moved to (4) for inspection, after which they are moved to the results tray (not represented). Video recording and HL2 performance are monitored via the PC and iPad represented. This figure is not to scale but does effectively convey the approach used.

The flow of WIP is facilitated by the SI, who is responsible for collecting and inspecting finished assemblies, recording results, and related tasks. The PI is focused on interactions with the participant and recording observations about their performance. The TS manages the HL2 system and acts as a secondary observer. In addition to these primary duties, all team members collaborated effectively to ensure the successful conduct of this study. This division of labor is summarized in Table 4.5.

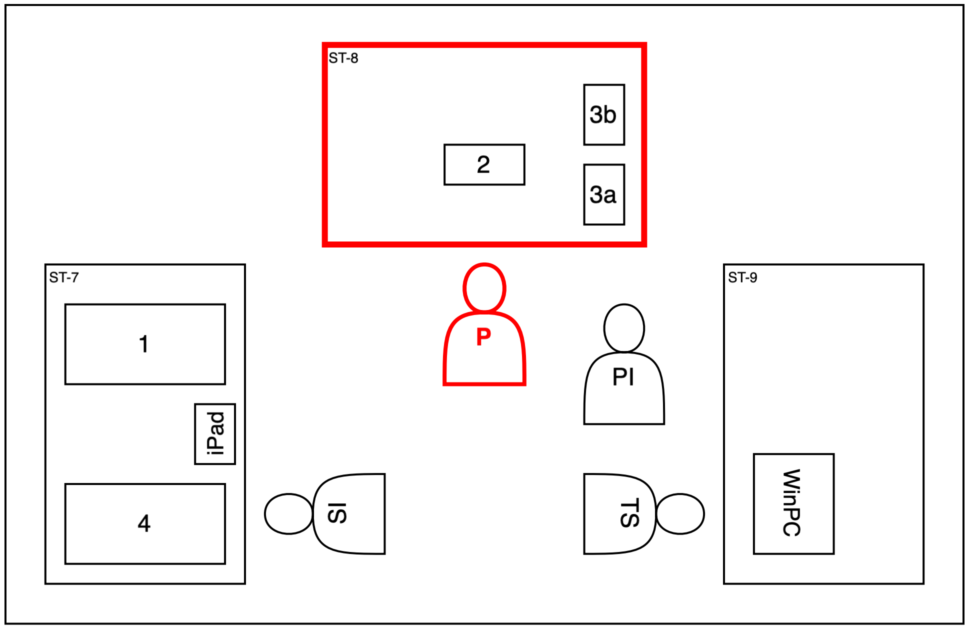


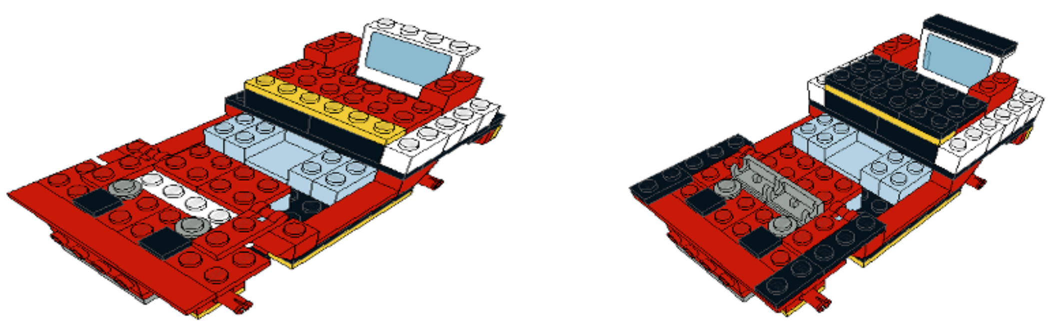
Figure 4.31: Workstation Configuration for Session 1

Table 4.5: Primary Responsibilities, First Session

Responsibility	Description	Role
Workflow	Facilitate the flow of work into and out of ST-8. Ensure that an inventory is available on ST-7 and promptly collect completed or rework items from ST-8.	SI
Data Collection	Review all completed, retired, and incomplete assemblies. Record the number and type of errors for each, along with the steps completed if incomplete.	SI
Documentation	Photograph the starting setup and the final output with the trial card in frame. Get additional photos were required to document unexpected results.	SI
Run HL2	Manage operation of the HL2 for treatments and/or recording. Monitor the stream and recording. Troubleshoot as required.	TS

Responsibility	Description	Role
Interact	Lead all participant interactions.	PI
Observe	Carefully observe the participant without disrupting their work. Intervene if required to correct unexpected behavior. Record interesting observations and insights. Respond to any work-stoppage events.	PI
Recycle	Prepare work area for next experiment or participant. Disassemble ST-8 complete cars to ST-7 state and verify. Maintain bin inventory, reset recording devices and HL2.	SI & TS

At the start of each day, ST-7 through 9 are cleared and configured as described above. Ten pre-built assemblies are positioned to the participant's left, at position (1). The construction of all ST-7 inventory is carefully verified against ST-7 standards, as pictured in Figure 4.32, to ensure an accurate starting point for all ST-8 tasks. A completed ST-8 assembly is also pictured for comparison.



(a) ST-7 Complete Assembly

(b) ST-8 Complete Assembly

Figure 4.32: Start and End Configurations for ST-8 Assemblies

4.8.4 General Policies

Except as otherwise noted, several policies are followed throughout the conduct of each session. Participants are encouraged to ask questions at any time. At each transition point, their understanding and readiness to proceed are confirmed. Discussion is allowed but limited and questions are answered but not embellished upon. Members of the research team are not allowed to prompt or otherwise instruct participants, except as specified. Overall, these policies were established to put participants at ease while maintaining a semi-formal tone and staying “on-script.”

4.8.5 Session Procedure

Following the initial setup, each session would proceed as follows.

1. Intake: welcome the participant, complete initial paperwork and instruments
2. Orientation: introduce the participant to the Lean Lab and their work area
3. Introduction: describe the general workflow at ST-7
4. Demo: demonstrate the assigned treatment at ST-8
5. HoloLens: introduce the participant to the HL2; don it and adjust fitment
6. Practice: have the participant practice using their assigned treatment
7. Learning Experiment: conduct the first experiment
8. Intermission: complete TLX and SUS for first experiment
9. Reset: prepare ST-8 for the second experiment
10. Recall Experiment: conduct the second experiment
11. Debrief: complete TLX and SUS for the second experiment; gather General Feedback
12. Recycle: prepare ST-8 for the next participant

Intake

Each participant is greeted and welcomed into the conference room, where drinks and snacks are offered. After they are settled, the participant is talked through the consent document, which was previously supplied. Once any questions are answered, they are asked to acknowledge their understanding and acceptance of it, initial each page, and sign. The document is countersigned by the primary investigator and placed in a lockbox with other consent forms before proceeding.

In accordance with the experimental procedure outlined in the NASA TLX instructions, all participants were familiarized with that instrument during the intake process. The primary investigator first introduces the TLX as a tool for workload assessment and briefly summarizes its design. The participant is then asked to read the provided *Subject Instructions*, after which their questions are answered. Finally, the participant is asked to complete a mock administration of the TLX for a hypothetical task, including both the *Sources of Workload Evaluation* and *Workload Rating Scales*. For this step they were asked to imagine they had just run a marathon, a task chosen for the high level of workload likely associated by all participants, regardless of running experience.

Next, the participant provides a variety of demographic data on the Participant Intake Form and self-reports behavioral data on the Behavioral Control Survey. Finally, they are briefed on COVID protocols and emergency procedures.

Orientation

The participant is guided to the second work cell while receiving a brief summary of the lab's LEGO-based training methodology, which emphasizes real-world practices and efficient, high quality production. Importantly, they are made aware of the line's 60-second takt time and its implications. This description is read from a script to ensure consistent delivery of the information. Upon arrival at the second work cell, they are introduced to assisting members of the research team.

Introduction

At the second work cell, workstations seven and eight are identified. Before moving to ST-8, where the experiment is conducted, participants are introduced to the general assembly process at ST-7. They are shown how to interpret the paper work instructions and a few assembly steps are demonstrated. The PWI from ST-7 is used to limit exposure to the ST-8 task details.

Then, a few rules and expectations are set related to dropped parts, correcting errors, breakage, and rework. It is also explained that the research team will only intervene in the event of an event that stops work.

Demonstration

The participant is led to ST-8 where they are informed of the assigned treatment. Interventions are briefly demonstrated to all those assigned, while members of the control group proceed to the next step. Each intervention is demonstrated in a manner similar to that used for the PWI at ST-7, with a focus on the treatment's operational details rather than the instructional content. Critically, all system controls, UI elements, and feedback mechanisms are explicitly described.

The HMDAR and HMDMR treatments, whose output is invisible to observers, requires a different approach. One member of the research team performs the demo while another narrates the process. The HL2's output is simultaneously streamed to a nearby laptop on ST-9. This allows participants to observe the HL2 operations as they are described, from both real and virtual perspectives.

HoloLens

All participants are reminded that, for observation and recording purposes, they are required to wear the HL2 during the practice session and all subsequent experiments. Additionally, they are notified of a second camera, positioned to the left of ST-8 and adjusted to only capture the work area.

Streaming is initiated on the HL2 before the participant is advised to don the device. Assistance is provided as necessary to ensure proper fit, after which the participant is asked to look at a reference point. If the streamed video centers on that target, it confirms that the wearer's line of sight aligns with the HL2's field of view.

Practice

Participants are instructed to engage in a brief practice session during which they assemble the first four bricks according to the provided instructions. This task is designed to ensure their operational understanding of the instructional treatment, while deliberately avoiding undesirable task training.

Learning Experiment

The assigned task, timeline, priorities (as outlined in Section 4.4.2), and workflow (as outlined in Section 4.8.3) are described. Recording is then initiated on both cameras and participants are asked to view the treatment slate (see Figure 4.33) while recording and alignment are re-checked. This also serves to mark the start of both videos with essential details that might help avoid confusion later. A 10-minute timer in view of the participant and both cameras is initiated, and they begin building with the aid of their IMT. Data is collected during the experiment as described in Section 4.4.2.

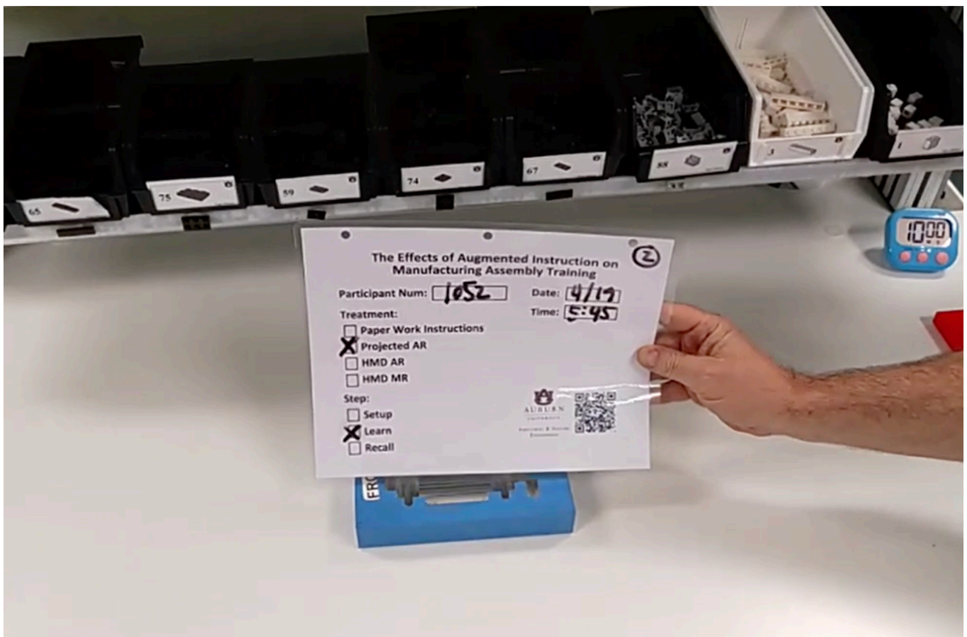
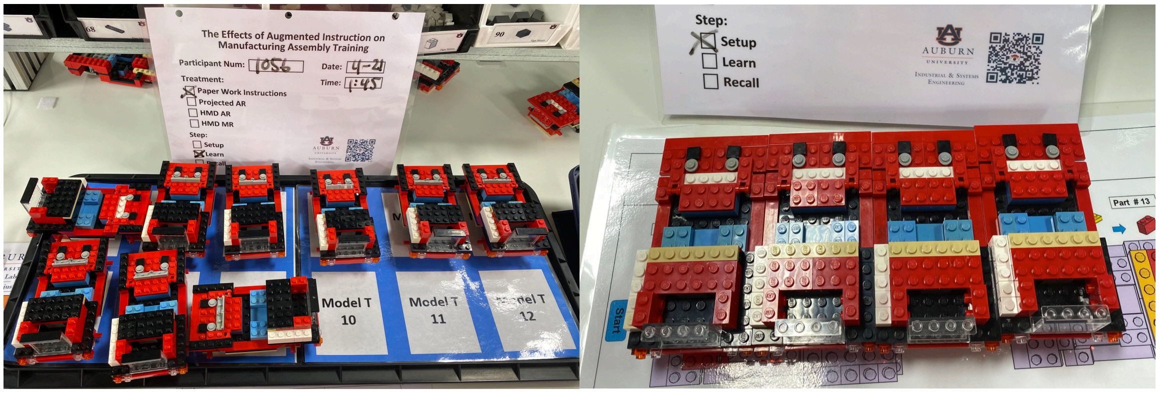


Figure 4.33: Laminated Treatment Slate

Intermission and Reset

At the conclusion of the first experiment recordings are stopped and the participant is asked to remove the HL2. Then they are escorted back to the conference room, where they are asked to complete the TLX and SUS based on their experience.

During this intermission, the research team records the learning results and resets the workstation for the next experiment. Four pre-built assemblies are put into inventory at ST-7, and any models built in the previous experiment are recycled to that standard. Photos of the results tray and reset inventory are taken to document the process and results.



(a) Results Tray

(b) Recall Setup

Figure 4.34: Documenting Results and Setup

Recall Experiment

The second experiment is conducted in the same manner as before. No timer or intervention is used, and the instructions / priorities are altered, per Section 4.4.2. Otherwise, the procedure is identical. This experiment concludes when four cars are completed.

Debrief

Following the second experiment the participant is led back to the conference room for the accompanying round of the TLX and SUS. Finally, the PI solicits any additional feedback the participant is willing to offer. For those that require prompting, the PI can refer to a list of standard topics. All participants are asked if they experienced injury or discomfort during the session. Feedback and responses are recorded on the General Feedback Sheet. Finally, the participant is thanked for their time and escorted to the exit.

Recycle

During the debrief, the research team records the recall results, resets the workstation for the next participant, and documents both with photographs.

4.9 Conduct of Second Session

This section will describe key operational details for the retention assessment, which was conducted during the end of study event. As described in Section 4.4.3, the scope and complexity of these trials was limited by expected traffic at the event. This also necessitated different signup procedures, lab arrangement, workflow, and staffing, all of which are described below.

4.9.1 Signup

The signup process was again managed by SignUpGenius, making it easy for the research team and invitees alike. The system was configured to help even out the flow of arrivals, preventing a backlog and keeping the event well attended throughout the day.

Twelve different start times were offered, one every 15-minutes between noon and 3:15pm. Up to five people could sign up for any start time, allowing for up to 60 total signups. Though we expected fewer, this ensured some flexibility in start times for those interested in attending. Start times were set, but all attendees were free to stay as long as they liked.

4.9.2 Setup and Traffic Flow

Prior to the event, the Lean Lab was arranged as depicted in Figure 4.35. Ongoing demonstrations for attendees would require the LG system at ST-8. Consequently, the retention experiment was relocated to ST-3, a similarly configured workstation in work cell #1. This choice ensured that the experience at ST-3 would most closely mimic ST-8, thus controlling for the effects of the change.

Experiments and demos from the collaborating study were conducted at ST-5 and 10, respectively.

4.9.3 Workflow and Roles

Before entering, attendees would register at (1), where they were given the appropriate data sheet and welcomed inside. Seating was provided at (3) to handle the queue of attendees waiting for their trial. Partitions placed at (2) obscured their view of the ongoing experiments and demos. This provided privacy and prevented re-exposure to the instructional material. As each retention trial concluded participants were led, in order of arrival, to ST-3. Those finishing were thanked and informed of the food and available educational / entertainment options.

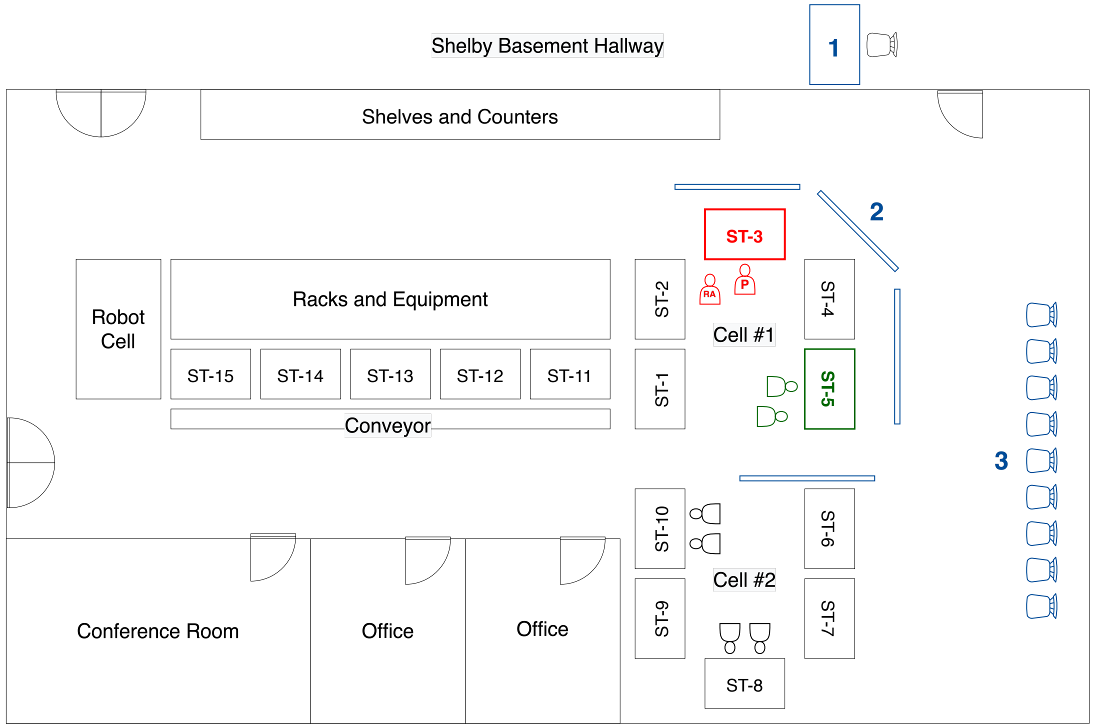


Figure 4.35: Lab Arrangement for Second Session

Volunteers from both studies were responsible for directing attendees through registration process, through the queue to their assigned experiment, and onto the activities that followed. The general policies set forth in the first session were again applied.

At the end of the day the lab was cleaned and restored to its normal operating state.

4.9.4 Session Procedure

Participants were asked to complete the ST-8 task for one car, from memory, without reference (e.g., paper work instructions). Data were efficiently collected with a single video camera, timer, and photos. With this approach, only a single research associate (RA) was required to direct and document the trial. An additional volunteer was responsible for recycling the inventory of assemblies.

4.10 Administration of Protocol

This section will describe the manner in which the study was run, with a focus on administrative details like the team, scheduling, location, safety measures, quality assurance, and challenges.

4.10.1 Personnel and Certifications

The IRB that this research was conducted under approved two related but separate studies. A third study planned to use the data from both for separate analysis. Members of all teams collaborated in the administration of the separate experiments, helping one another provide the coverage and support required to perform over 100 trials in a single semester.

As a result, the final IRB listed 13 key personnel, as summarized in Table 4.6.

Table 4.6: Research Team Breakdown

Type	Description	Count
Study PIs	The Principal Investigators for all collaborating studies.	3
Faculty Members	Members of related committees, for oversight and institutional stewardship.	3
Research Assistants	Graduate (3) and undergraduate (1) student volunteers assisting with the conduct of experiments.	4
Technical Support	Members of the undergraduate research team that developed the HL2 capabilities and assisted with the conduct of experiments.	3

All team members were certified by the Collaborative Institutional Training Initiative[33] (CITI), in accordance with university policy and the team's commitment to professional research. Through this program they received training on the ethical implications and compliance standards of their work.

[33] Collaborative Institutional Training Initiative: https://about.citiprogram.org/

Except for faculty members, members from all groups were directly involved in the various aspects of the conduct of this study, as detailed throughout this document. Volunteers signed up for shifts on SignUpGenius, using a form separate from the participants.

4.10.2 First Session Scheduling

As discussed in Section 4.5.3, the first session experiments were conducted in the Spring of 2023. Specifically, trials were run for 10 weeks, from February 10th to April 27th. A routine schedule of weekly trials was established based on the availability of the Lean Lab and members of the research team.

Between 10 and 12 slots were available each week at a variety of days and times. A total of approximately 100 available slots were offered. Though we planned to run significantly fewer, this would make it easier for interested participants to find a day and time that would work with their schedule. A few Saturday shifts were also offered for the same reason.

Each slot was 75-minutes in duration, which was slightly more than the estimated average treatment duration (60-minutes). This padding allowed us to accommodate late arrivals, unusually long trials, and other unexpected events with little knock-on effect.

SignUpGenius was configured to only show available slots in a rolling two-week window. This was done with the hopes of encouraging signups by creating a sense of scarcity and urgency, and reducing no-shows by preventing interested but forgetful participants from signing up too far in advance. At the start of each week the schedule was reviewed to make adjustments for staff availability (e.g., illness and travel) and notify interested participants of slots that remained open.

4.10.3 Safety Measures

All team members were aware of the study's COVID-related precautions and emergency action plan, and committed to the self-screening and reporting as required by the former. Both documents were readily available to team members, who were trained to reach out to members of the included contact list or escalate to emergency or non-emergency assistance as deemed necessary. Current phone numbers were provided for all cases. This documentation is included with the approved IRB forms in Appendix E.

Additionally, team members were trained to observe all participant activities for dizziness, related vestibular issues, or any other significant but unexpected side-effect. In that event, the experiment would be suspended, HL2 removed, and the participant seated for assessment. In the event assistance was required, the emergency plan would be consulted.

General safety procedures included routine sanitation and work area maintenance. All work surfaces were wiped down before each participant and the HL2 was sanitized between wearers. In keeping with the Lab's 5S[34] plan, all team members worked to ensure the work area remained free of obstructions and trip hazards.

4.10.4 Test Runs

Prior to running the first participant trials, a series of five test runs were conducted to train the research team on its execution and identify any procedural issues. These sessions utilized volunteer members of the research team that were not qualified to participate in the study due to their experience in the Lean Lab and/or with the interventions. Accordingly, no data was collected for subsequent analysis.

Feedback and notes collected during these tests were reviewed by the team and changes proposed. The outcome was used to refine the procedure, streamline its flow, identify details previously overlooked in the checklists and data collection forms, and flesh out the draft script. Though no large scale changes were made, collectively, the improvements had a meaningful effect on the procedure's overall quality. They were incorporated into the second IRB submission as appropriate to ensure it properly reflected the latest protocol.

4.10.5 Quality Assurance

Throughout the design and conduct of this study, every effort was made to ensure its findings were valid and verifiable. Primarily, this was achieved through the deliberate and meticulous design of the protocol, as documented throughout this chapter.

Quality assurance was operationalized through a carefully documented procedure that integrated a script and checklists. The script was used by the PI to ensure consistent interactions with each participant. Checklists were used by the entire team to verify that workstations, assemblies, and equipment were always properly configured.

Standard roles, consistent staffing, and clearly defined responsibilities were all established to help ensure routine trials by reducing variability in their conduct. Test runs performed

[34] 5S is a methodology and mindset for maintaining a work space organized for efficiency and effectiveness. https://en.wikipedia.org/wiki/5S_(methodology)

before the first participant were used to train the team and collectively identify ways to refine the plan. Once trials began, feedback from team members and participants was routinely used to make adjustments that improved the process without impacting the results.

Results recorded by the SI were routinely double-checked by the PI before being photographed for later review. Most data was extracted and verified outside of the hustle of data collection, using video and photo evidence to ensure the accuracy of data post-intervention. Together, these measures helped ensure the quality of the data collected, thereby elevating the integrity of our findings.

4.11 Data Extraction

This section will describe the methods used to extract the data collected in instruments, data sheets, photographs, and video. All extraction processes were performed by the author on a 16" 2019 MacBook Pro (model A2141) with 2.3 GHz 8-Core Intel i9, 32GB of RAM, and an AMD Radeon Pro 5500M with 8GB, running MacOS 13.x. Before being archived, all original paper documentation was digitized using a Fujitsu ScanSnap iX1300 scanner and the included software.

When extracting raw data from PDF sources, the only changes made were to label categorical values as described in the following sections. Otherwise, except where specifically noted, no changes were made to the data during transcription. Where no value was provided, "N/A" was used.

Any data issues encountered were noted and marked for correction during the subsequent cleaning process. This approach allowed us to explicitly document required corrections in the code that made them, improving transparency and reproducibility. Finally, anything more involved than simple labeling, such as more sophisticated encoding (e.g., one-hot encoding) was left to the analysis phase, where computational methods would be employed to reduce the chance of error.

4.11.1 Instruments and Data Collection Sheets

The contents of all digitized results were manually transcribed from their PDF into spreadsheet format. Excel was used at this stage for its human-friendly interface, which facilitated data entry and consolidated all results into a single file. As detailed in the subsequent section, the `XLSX` file format is an open standard easily parsed by most programming languages. Together, these considerations made `XLSX` an ideal interim format for raw data.

Each instrument was given a separate tab, with each row representing a participant record. The results from both Data Collection Sheets were combined into a single tab, where each row recorded the outcome of a single assembly task. This process was straightforward except as described below.

Personal Information Form

Fields in the PIF were a mix of numeric (e.g., age, height), nominal and ordinal categorical (e.g., gender, education level), datetime (e.g., the date and time of scheduled trial), and text values (e.g., notes). One change was made to the raw data for this form. Where participants that were known to be degree-seeking AU students incorrectly marked "High school degree or equivalent," their status was changed to "Some college but no degree".

NASA TLX

For each of the TLX's Sources of Workload comparisons, the response was labeled `1` or `2` for top and bottom choice, respectively. Columns were labeled `S1` through `S15`, corresponding to each of the 15 pairs shown in Figure 4.13, numbered left to right and top to bottom. For example, a value of `2` (bottom choice) for `S2` (Temporal Demand vs Frustration), corresponds to a user response of Frustration.

Workload Ratings were similarly tabulated in columns `R1` through `R6`. They were scored as marked on the 100-point scale, with 5-point graduations. Any mark between graduations was rounded up, per the TLX instructions. The first five participants were incorrectly given a Likert-style scale with values from 1 to 7 and bipolar descriptors "Very Low" and "Very High." These were scored as indicated with a note to correct the values during analysis. The form was corrected for subsequent participants.

Finally, some participants may have mistakenly scored their Performance Rating. While other factors employ a scale progressing from left to right, with descriptors ranging from Very Low to Very High, the Performance scale is labeled with Perfect and Failure at its endpoints, reflecting the inverse correlation between performance and perceived workload. Despite the consistency of a rightward increase in perceived workload for all factors, participants might erroneously associate higher ratings with enhanced performance. This issue is noted in the TLX instructions and explicitly pointed out to participants, but suspicious values appeared during the transcription process. We will investigate potential issues during the data cleaning stage.

System Usability Scale

Numeric responses (1-5) for all ten questions were transcribed as–is in columns `Q1` through `Q10`. Participant #1058 realized they had scored themselves using a reversed scale during the first administration of the SUS. This was noted in the data, which was corrected during transcription.

Data Collection Sheets

Each row of the "Outcomes" tab contained the manually-recorded results for each attempted assembly task, i.e., a unique combination of participant number, experiment number, and car number. The number of uncorrected errors (UCE) was determined by inspection of each final assembly, verifying that the brick for each step was correctly selected, placed, and oriented. The quality of attachment was not considered. For each UCE encountered, its type was recorded as a combination of selection, position, and rotation errors. Brief descriptions of the nature of each error, e.g., "front 59 swapped", were also included to aid interpretation.

Results recorded by the research team during the experiments were naturally error-prone. During transcription, all were verified against the corresponding photographs and, if necessary, video recordings. Corrections were made as required. This approach differs from the handling of self-reported data, where the integrity of participants' personal perceptions must be preserved as provided.

A final step transcribes the recorded results into a concise format that encodes a detailed, contextualized description of each error. For example, `F59LRP (2)` indicates that, at the front of the car (`F`), part `59` attached to the left and right (`LR`), were incorrectly positioned (`P`), accounting for `(2)` errors. This encoding scheme was developed for the study to describe error types in a contextually rich manner that may better relate to the corresponding human error than the discrete elements it comprises. This approach is designed to facilitate pattern discovery and analysis, support predictive modeling and simulations, and inform more targeted interventions.

4.11.2 Qualitative Feedback

In addition to the quantitative data from surveys, observed results, and annotated performances, a number of qualitative observations were recorded by the research team during the conduct of each session and in debriefs that followed. These were manually transcribed from various sources, including the Data Collection Sheets, General Feedback Forms, and the PI's notes. When transcribing notes from the debriefing, which were often terse, care was taken to accurately represent the participant's original feedback.

The results were collected in a separate Markdown[35] file (`MD`) for each participant, e.g., (`1001.md`). Markdown is a lightweight plain-text formatting syntax proposed by John Gruber in 2004[36]. Originally intended as a tool for HTML generation, Markdown is now widely used to create text documents that have structure and are easily rendered in a variety of styles.

```
# 1048

## Script Notes
- Tall, head-looker
- Area tracking drop tally: 3
- Button issues - inconsistent, sometimes first tap, others 5-10

## Participant Feedback
- Next button inconsistent - sometimes didn't work, double-clicked
- Memorized by c2-3 in Learn
- Broke down assembly steps by section of car: front, back, middle
- Going back to previous part (eg 59?) problematic
```

Figure 4.36: Sample of Collected Feedback, in Markdown Format

Figure 4.36 exemplifies both collected feedback and Markdown formatting, using data from participant #1048.

4.11.3 Video

The bulk of collected performance data was extracted from the video recordings, a tedious and time-consuming process. Given the estimated 12-18 hours of total footage, the need for efficient, accurate results demanded a strategy that allowed close inspection from multiple

35 Markdown: https://en.wikipedia.org/wiki/Markdown

36 Original Markdown Project Home: https://daringfireball.net/projects/markdown/

synchronized views along with rapid tagging and annotation. Careful consideration was given to the selection of tools used to achieve that.

Tooling

When considering tools for the video annotation workflow, priority was given to robust and reliable software with modest hardware requirements, user-friendly interfaces, and affordability. Specifically, the annotation workflow must be capable of exporting machine-readable outputs. Finally, it was essential to avoid altering the original videos. A non-destructive editing approach would dismiss any concerns regarding the veracity of the recordings, by ensuring a reliable and traceable data source.

A search was conducted, and a variety of tools were tested. Some notes from that process are included in Appendix B. In the end, three complementary tools were selected that best fit the requirements identified:

1. Filmora[37] v12.x was selected to combine the raw video into a split screen presentation with synced action. This facilitated annotation by providing two angles of the action and a zoomed view. Filmora is an affordable, non-destructive editor that offers a good balance of capabilities and ease of use.

2. Handbrake[38] v1.7.x was selected to downsize and compress Filmora output, improving performance during annotation. Handbrake did this much faster and with better final quality than Filmora. This popular and highly-regarded tool is free and open-source.

3. Kyno[39] v1.8.x was selected to add markers for instantaneous events (e.g., breakage occurs) and sub-clips for events with duration (e.g., task completion times). Notes can be added to either as desired. The metadata is exported in eXtensible Markup Language (XML, World Wide Web Consortium (W3C), 2008), a flexible, text-based language that is used to structure, store, and transport data. Kyno is specifically designed to accelerate this task in a production environment and is generously priced for the market.

[37] Filmora: https://filmora.wondershare.com/
[38] Handbrake: https://handbrake.fr/
[39] Kyno: https://lesspain.software/kyno/

Processing

A general description of video processing follows. For a step-by-step treatment, see Appendix C for the original instructions.

For each participant, two composite videos are created, one from footage of their learning experiment and another for recall. Raw video files are first renamed in a standardized format based on participant number, phase, and camera angle, then organized into specific project folders.

For each participant experiment, a single three-pane video was created in Filmora, as demonstrated by Figure 4.37. The two stacked views on the left provide the first-person view captured by the HL2 (top) and a zoomed-in view from the side camera, centered on the fixture (bottom). The right-hand pane provides a wider angle version of the same side camera footage. All three views are manually synchronized by tagging corresponding events and sliding the clips to align those markers, matching the observed action. Before exporting the result, the Filmora project is saved.

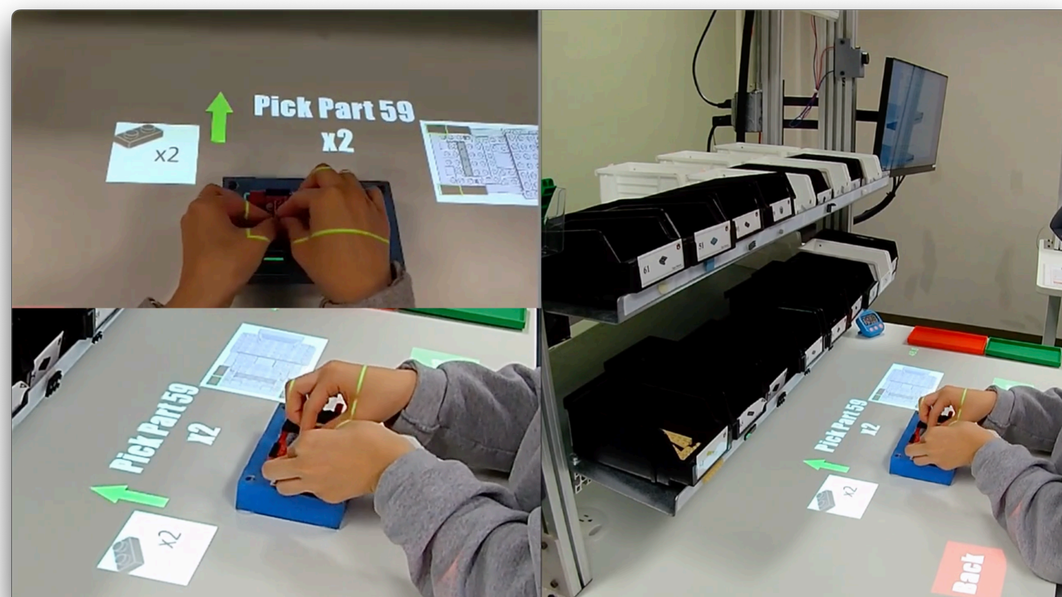


Figure 4.37: Frame of Composite Video from Learning Trial, PAR Treatment

The output was rendered by Filmora in the MP4 video format with full HD resolution (1920x1080) at 30 frames per second, using H.264 compression. This was then resized to 720p (1280x720) by Handbrake. The final file size is reduced to about 20mb / minute from multiple gigabytes of source material. This greatly increases the responsiveness the annotation experience without significantly degrading visual quality.

The compressed 720p video is loaded in Kyno where two types of events were annotated. Instantaneous events (e.g., car breaks during assembly) are identified by placing a marker at the moment of occurrence. Events with a duration (e.g., the start and end of each assembly task) are assigned to sub-clips. Both markers and sub-clips are given standard names that denote the event type, and descriptions where additional detail is called for. The standard list of names is included in Appendix C, along with other important details about this process.

Finally, the Kyno project file is saved and annotation data is exported in XML format. A sample of that output is provided in Figure 4.38, based on participant #1051's learning experiment. This is confirmed by line (2), which shows the data is associated with the file `1051-Learn.mp4`. Lines (10) and (12) denote the start and end times of the "Car 1" subclip, each given in 1/90,000ths of a second.

```
<context-info>
  <url>file:/Volumes/.../1051-Learn.mp4</url>
  <size>173341872</size>
  ...
</context-info>
<title>1051-Learn</title>
<description>side camera views are...</description>
<marker>
  <id>6436e1f6-b9a2-4a58-800a-12a13ed89ab7</id>
  <timestamp time-base="1/90000">895175</timestamp>
  <type>subclip</type>
  <duration time-base="1/90000">17932825</duration>
  <title>Car 1</title>
  <description>First car duration</description>
</marker>
```

Figure 4.38: XML Export of Annotation Data from Kyno

4.11.4 XML Processing

A Python (v3.11) script was used to create a CSV comprised of performance data extracted data from all Kyno `XML` files. The process is outlined with pseudocode in Figure 4.39.

While the XML data includes annotations for both instantaneous events (e.g., breakage occurs or defect encountered) and those with duration (e.g., task completion time), only the latter are included in the resulting CSV. While instant events can provide important

```
For each xmlFile:
    Extract file metadata (url, size, etc.)

    For each marker in xmlFile:
        Extract marker details (id, timestamp, type, title, descript.)
        If duration element present, add duration to marker
        Convert timestamp and duration to seconds
        Add marker to markers list

    Categorize markers into subclips and other
    Add data to xmlData with key (participant, phase)

For each participant, phase pair in xmlData:
    Extract subclip markers for the participant and phase
    For each subclip marker:
        Format data (participant, phase, marker details) as row
        Add row to csvData

Write csvData to a CSV file
```

Figure 4.39: Pseudocode for XML Data Extraction (`process_data.py`)

context for understanding performance, the focus of our analysis will be on task outcomes, which involve a duration.

The result includes columns for participant number, experiment number, and the event name, start time, duration, and description. For example, the data extracted for the first participant's learn experiment is summarized in Table 4.7.

Table 4.7: Learning Event Data Extracted for Participant #1001

Part	**Exp**	**Event**	**Start**	**Dur**	**Description**
1001	1	Car 1	6.28	97.933	N/A
1001	1	Car 2	108.367	66.7	N/A
1001	1	Break 3	178.367	76.033	failed attempt to repair
1001	1	Car 4	260.933	68.9	N/A
1001	1	Car 5	333.467	60.9	N/A
1001	1	Car 6	397.333	66.933	N/A
1001	1	Defect 7	467.867	27.333	prebuilt missing piece
1001	1	Car 8	498.4	53.7	corrected
1001	1	Car 9	556.067	47.833	finishes at the buzzer

This shows they attempted nine cars and completed seven. Breakage occurred during the third assembly and a repair was attempted, but the car was ultimately retired. Car 7 was also retired when a defect was noticed in the prebuilt. The last car was completed in only 47.8 seconds, just before the 10-minute time limit.

The output of this script was saved as `i1_times.csv` and carefully validated against the XML data, video recordings, and reports described next.

4.11.5 System Availability

Due to challenges associated with area and model-based tracking, described elsewhere, participants using the HL2 would experience system down time that we referred to as "drop-outs." This was the result of a loss of tracking that caused the user interface to deactivate. All drop-outs were marked as subclips during video annotation process so the lost time could be accounted for.

Each drop-out event was later reviewed to assess how much impact it had on the current task. In some cases the drop-out occurred between cars or was disregarded by the participant. In others it caused worked to stop until tracking was reacquired. For each drop-out a value between 0.0 and 1.0 was assigned, based on the assessed impact.

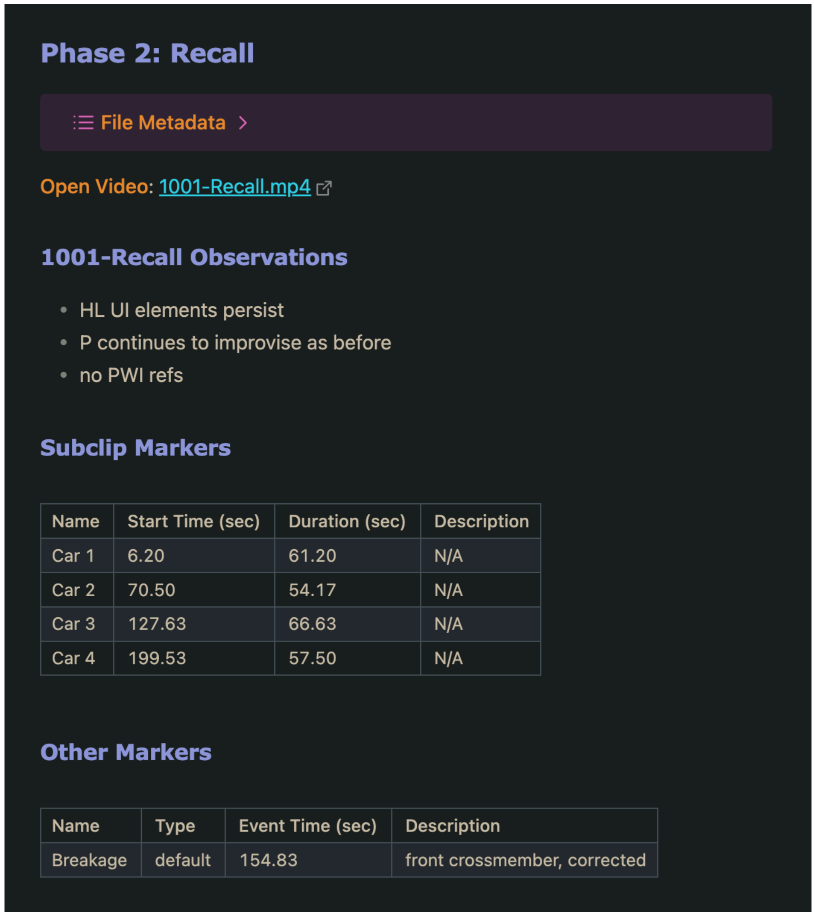
The resulting times and weights could be used later to scale the drop-out duration accordingly, either for the overall analysis or to compare results with and without drop-out effect. Alternatively, they could be treated as system availability for OEE calculations.

This process was done manually and based on the PI's best judgement after all other annotation work was complete. It was recorded in a separate Excel sheet, `adjusted_drop_events.xlsx`.

4.11.6 Report Generation

To assist with data validation and better understand factors that contribute to individual performance, a detailed report was generated for each participant. This was done as part of the same XML extraction Python script described above.

Each report aggregates participant demographics and car outcomes from observed and self-reported data, qualitative feedback from transcribed Markdown files, and quantitative event data from XML annotations. The result paints a comprehensive picture of the participant's experience in a single Markdown report. This is demonstrated by Figure 4.40, a screenshot of the recall portion of #1001's report.



Phase 2: Recall

File Metadata >

Open Video: 1001-Recall.mp4

1001-Recall Observations

- HL UI elements persist
- P continues to improvise as before
- no PWI refs

Subclip Markers

Name	Start Time (sec)	Duration (sec)	Description
Car 1	6.20	61.20	N/A
Car 2	70.50	54.17	N/A
Car 3	127.63	66.63	N/A
Car 4	199.53	57.50	N/A

Other Markers

Name	Type	Event Time (sec)	Description
Breakage	default	154.83	front crossmember, corrected

Figure 4.40: Portion of Participant #1001's Report, Obsidian Screenshot

From the top, this report includes metadata (collapsed for brevity), clickable link to the video, list of PI's observations, and summaries of performance (subclip) and event (other) markers. The instantaneous event markers included here provide valuable added context. For example, we can see that a breakage was corrected during the third car, likely contributing to its increased duration.

The formatting seen in this image was rendered by Obsidian[40], a powerful tool for making and organizing Markdown documentation. This approach greatly facilitated the review of participant outcomes.

[40] Obsidian: https://obsidian.md/

4.11.7 Data and Code Management

Like all academic work, the success of this study hinged on the credibility, transparency, and reproducibility of the research, data, and results. As such, data management was a critical consideration in the development of the data extraction and processing workflow. The plan carefully accounted for data storage, version management, and the file formats used. Those details aligned with an access plan balancing accessibility and portability with privacy and security.

Anonymous data was stored locally in three separate locations, all on secure hardware accessible only to the PI: an internal laptop hard drive, an external RAID5 storage array, and an external NVMe backup drive. Additionally, GitHub[41] and BOX[42] cloud services were employed to provide more redundancy and secure access to select team members. A detailed account of this plan, including the hardware, software, folder structures, and naming conventions used, are provided in Appendix D.

Key criteria for software selection included multiplatform compatibility, open-source licensing, and the ability to generate files using open, preferably text-based, standards. These characteristics were crucial to ensuring data portability and the reproducibility of results.

Wherever appropriate, working files, including all data analysis in R and Python, and manuscript development in Quarto[43], were placed under version control using Git[44], and synchronized to GitHub. HL2 source code was treated separately, as discussed in Section 4.5.5.

Isolated virtual environments were used with both Python and R projects to manage package dependencies. This setup guards against software conflicts and maintains consistent computational environments that are crucial for reliable and reproducible research outcomes.

The entire process was designed to ensure traceability, allowing every result to be reliably linked back to its original source data, bolstering the integrity of our research findings. The priorities outlined here were carried on throughout the analysis process, described next.

[41] GitHub: https://github.com/
[42] Box: https://www.box.com/
[43] Quarto: https://quarto.org/
[44] Git: https://git-scm.com/

4.11.8 Data Extraction Challenges

Data extraction was a significant effort that required the bulk of the author's time over the summer of 2023. A number of challenges were encountered along the way, predominantly related to the collection of data from recorded videos.

Four videos were recorded for each participant, two each from the learning recall phases. Each pair of videos were manually synchronized, composited, edited, and compressed, and annotated into a single output for each participant-phase. This process took approximately 1.5 hours per participant, but ranged between 45 minutes and over 4 hours, based on rough notes of progress. Of this, the video synchronization and annotation processes described in Section 4.11.3 took the most time.

Manually synchronizing videos was done by offsetting them to match movements and align the timelines. Tricky in the best of circumstances, occasional HL2 issues, including dropped video, crashes, and unsteady frame rates significantly elevated the challenge. Dropped video, a crash, or reset required piece-by-piece reconstruction of the session, with additional synchronizations. Variable frame rates, on the other hand, resulted in HL2 recordings with dropped frames and slightly non-linear playback. Over time, this leads to drift in the sync between sources, which cannot be corrected without destructive edits that would alter the timing data. In most cases the drift was insignificant, but in others it required a workaround. To address it we adopted the convention of using the side camera view for all annotations related to the workpiece, and the first-person view for any related to PWI consultations and HL2 issues. Consistently annotating events in this manner helped avoid including sync offset in a measured duration.

The time required to annotate each video varied with participant performance and behavior. Uneventful trials mostly involved marking the start and end time for each workpiece. In most cases occasional additions were made for various other events. In situations where the HMDMR tracking was problematic, or where PWI consultations were extensive, the required time and effort increased substantially. One trial included over 100 PWI consultations during the Recall phase, each with manually located start and end points. This took several hours to do accurately.

Otherwise, data extraction was relatively straightforward, with only two other noteworthy issues. While we had originally intended to count corrected errors, it proved impractical to consistently differentiate corrected errors from other participant behaviors. In many cases this required too much interpretation on the part of the scorer. Finally, we relied on handwritten notes taken by the PI for the final debrief and participant feedback. Some were terse and difficult to interpret, suggesting that important details may have been lost. In the future it would be better to record and digitally transcribe these sessions if acceptable.

Overall this process was time consuming but well planned and carefully executed for the given inputs. Future studies of this sort would greatly benefit from a synchronized multi-camera recording setup capable of directly rendering the desired screen layout, saving hours of editing time and eliminating the drift issues.

4.12 Data Cleaning and Analysis

This section will detail the procedures for data preparation, then summarize the intended analytical approach. Analysis will be implemented and more thoroughly detailed in the Results chapter to follow.

4.12.1 Tools and Methodology

R was used for all data cleaning and analysis. A mix of R Markdown (`RMD`) and Quarto (`QMD`) notebooks were created in the RStudio integrated development environment (IDE). The notebook format allows users to intermingle Markdown-formatted text, with code and output in a way that is very well suited to the exploratory nature of this work. This best approximates the idea of *Literate Programming* originally described by Donald Knuth[45] as a narrative approach that interleaves code and writing in a way that promotes reader understanding (Knuth, 1984).

4.12.2 Cleaning and Transformation

The primary outputs of the prior extraction effort were:

1. `i1_raw_data.xlsx` containing the self-reported data and observed results, manually transcribed. Each tab contains a different data set.
2. `i1_times.csv` containing event times extracted from the video annotations. Each row corresponds to a participant, experiment, car number combination.

[45] Donald E. Knuth is an esteemed American computer scientist, best known for "The Art of Computer Programming," a multi-volume series widely regarded as one of the most comprehensive texts on algorithms. He also developed the TeX typesetting system, and made numerous other contributions that significantly shaped the landscape of computer science.

3. `adjusted_drop_times.xlsx` listing all drop out events and the assigned weights.
4. Notes for each participant in `MD` format, compiled from observations and feedback.

Prior to analysis, additional cleaning and transformation was required. This included identifying parent-child relationships among events, scoring the TLX and SUS, correcting errors in the data, enforcing standard naming conventions and data types, and collecting the results into a single `XLSX` file.

Data was cleaned and transformed by R code found in the notebook `forms_data.rmd`. The process was iterative, with each step detailed in the following sections.

First Pass

The initial pass of cleaning and transformation can be summarized as follows:

1. Process Demographics: Combine date and time into a single datetime column, cleans column names, and converts most columns to factors.
2. Process Car Outcomes: Replace flag values with meaningful categories and assign sequence numbers.
3. Process Car Times: Clean names, extract event types, assign sequence numbers, and categorize markers into parent and child events based on event types and times
4. Correct Errors: Implement various changes to correct for previously identified errors in the data. Each is documented and justified in the code.
5. Join and Save: Add unique IDs (UIDs) for events and outcomes, reorder essential columns, join table data, and save as CSV.

Second Pass

The first pass output was carefully reviewed, during which an improved categorization scheme was defined for all events. The new scheme, which simplified and standardized the ad-hoc labels generated during annotation, is summarized in Table 4.8.

Table 4.8: Event Categorization Scheme

Type	Category	Description
Parent	Car	Assembly completed during the time allowed.
Parent	Breakage	Assembly retired due to breakage.

Type	Category	Description
Parent	Defect	Assembly retired due to defect in prebuild.
Parent	Incomplete	Assembly incomplete when time expired.
Child	Drop	Time lost due to HL2 drop-out.
Child	PWI	Time lost to PWI reference (recall only).
Child	Repair	Time lost as repairs are made to assembly.
Child	System	Time lost due to other system related issues.

The second pass of changes started by applying the updated categories. Again, each change is explicitly stated in the code. Adjusted drop times were then joined with the resulting data, and two new columns were generated for each type of child event. These totaled the number of events of each type for each parent, and their durations. Finally, those results were verified with automated tests and the results were written to `CSV`.

TLX and SUS Scoring

These instruments were scored as described by their providers. As noted in Section 4.11.1, TLX responses for the first five participants were corrected by rescaling to the standard 100-point system, rounding up. TLX and SUS results were written to separate `CSV` files and reviewed for correctness.

Final Output

Finally, the outputs described above were combined and saved as `combined_results.xlsx`, again with one tab for each: demographics, car outcomes, car times, car results, system usability scores, and tlx scores. For clarity, the car tabs differ as follows:

- **Outcomes** is the final version of the observed result and errors for each car.
- **Times** is the final version of the data extracted from video annotations, including child events, with event type and category.
- **Results** combines the outcomes and times tables, with one row per assembly and aggregating all time lost to child events.

Most analysis will focus on the Results tab, but the others are retained for traceability.

4.12.3 Analysis

Based on the primary and secondary research questions identified herein, several types of analysis will be required. Essential methods are summarized in the following list.

1. Descriptive Statistics
   - Calculate means, standard deviations, and other descriptive measures for the dependent variables (e.g., task completion time, error rates, OEE) across different treatment groups.
   - Present summary statistics for participant demographics and prior experience.
2. Hypothesis Testing
   - One-way ANOVA: Use to compare means of dependent variables (e.g., average time per car, average error count per car, OEE) across the four instructional methods (PWI, PAR, HMDAR, HMDMR).
   - Repeated measures ANOVA: Calculate to analyze changes in performance over time (e.g., learning rates, change in OEE) within and between treatment groups.
   - Post-hoc tests (e.g., Tukey's HSD, Bonferroni correction): Apply these to determine which specific treatment groups differ significantly from each other, if the ANOVA results are significant.
3. Effect Size Estimation
   - Partial eta-squared (η^2) or omega-squared (ω^2): Calculate these to assess the magnitude of the treatment effect on the dependent variables.
   - Cohen's d: Used to compare the effect sizes between specific treatment groups, if post-hoc tests reveal significant differences.
4. Regression Analysis
   - Multiple regression: Examine the relationship between operator characteristics (e.g., prior experience) and performance outcomes, while controlling for other relevant variables.
5. Correlation Analysis

- Pearson's or Spearman's correlation: Use these to investigate the relationships between perceived workload, usability, user satisfaction, and performance outcomes.

6. Qualitative Analysis and Visualization

   - Thematic analysis: Apply to the open-ended exit interviews to identify common themes and patterns in participants' experiences and perceptions of the different instructional methods.
   - Data visualization: Create graphs, charts, and tables to present the results of the above analyses effectively, such as bar charts for comparing means, line graphs for displaying learning curves, and scatterplots for showing correlations.

4.13 Limitations of Study Design

Despite the careful design of this study it has its limitations. Those identified before the first trial are described below. Except as described therein, we consider each of these unlikely to influence the results and then only limited in their effect, making them very low risk overall.

The participant recruitment, sampling, and selection processs had a number of practical limitations. As is often the case with graduate research, participants were recruited from the university community and the sample was dominated by undergraduate engineering students. This "convenience sampling" approach limits our ability to infer from it the expected performance of manufacturing assembly operators, or the factors that influence it. That said, the ecological validity of this study is deliberately higher than most similar studies due to its realistic surroundings and validated task instructions.

Also, this study employs a method for assigning treatments that is as random as possible given the constraints. Strictly speaking, the method described in Section 4.6.4 may only be considered pseudo-random by some. In that case, the validity of statistical tests that assume random assignment may be compromised.

Additionally, the decision to offer the possibility of compensation as described in Section 4.6.2 was approved shortly after the trials commenced. The first five trials had been run and other participants had been recruited without knowledge of these perks. Due to the timing of this change and a delay between the pilot study and subsequent participants, the impact of this change is considered negligible but could potentially change the type and motivations of participants recruited and their results.

From a task design point of view, we must accept that LEGO cars are not, in fact, real cars. This does reduce task validity, but is again offset by the surroundings and validated instructions. Furthermore, the complexity of the task itself may be insufficient to fully assess learning, recall, and retention. On the other hand, one could argue that LEGO assembly is, in fact, more error prone than many automotive manufacturing tasks which are designed to prevent the possibility of errors. This error-proofing process is known as Poka-Yoke. It is a key component of Lean Manufacturing principles and considered superior to any error checking methods, especially those done by humans. Regardless, these limitations are reported in other studies and are likely the most impactful identified. We will use analytical methods to look for signs of insufficient difficulty in the Results chapter.

As detailed in Section 4.2.2, treatments were carefully designed to control for most recognized confounders, but some limitations have been identified. First, the HMDAR and HMDMR treatments are not exact replications of the PAR instructions. Even allowing for the inherent differences of these treatments, the user experience in each deviates in minor ways from the PAR implementation. We believe these deviations are small and have no material difference, but recognize that they may introduce uncontrolled differences.

Learning effect is another source of concern. The study's between-groups design controls for it by preventing participants from carrying learning from one treatment to the next. However, the participant orientation process may create a minor dissimilarities in baseline task knowledge across treatment groups. It is introduced by an additional training step given only to the PAR, HMDAR, and HMDMR groups. This training introduces each participant to the operation of their assigned device so that the learning experiment that follows measures how they learned to perform the task, not how to operate the device. This is also a tradeoff, seeking to put all participants on equal footing for the first experiment, without any advantage in either task or treatment understanding. We can look for evidence of an effect during the analysis to follow.

Finally, this study did not assess the general cognitive or spatial skills of participants and therefore cannot balance the groups accordingly or use the results as independent variables during analysis. These measures were deliberately excluded to limit the overall time required for each trial. The instruments typically used would nearly double the expected duration, which would severely limit data collection. This aspect of the design is left as future work.

4.14 Summary

This chapter provided a comprehensive overview of the research methodology employed in this study, which aims to investigate the impact of different AR/MR instructional methods on operator learning, recall, and retention in a manufacturing assembly training context. The study's design is grounded in an innovative, affordance-based framework that systematically compares the effects of specific AR/MR features, such as hands-free interaction, spatial registration, and user-centric displays, on training outcomes.

A key strength of this research lies in its rigorous and ecologically valid approach. The study is situated within an authentic manufacturing training environment, utilizing validated assembly tasks and instructional materials that closely resemble real-world conditions. This enhances the generalizability of the findings to industry settings and ensures their relevance for informing the practical implementation of AR/MR technologies in manufacturing training.

The study employs a multi-phase, mixed-methods design that combines quantitative performance measures with qualitative user feedback to provide a comprehensive assessment of the effectiveness of different instructional media types (IMTs). The inclusion of a traditional paper-based control group allows for a direct comparison of AR/MR interventions against standard training methods, while the multiple AR/MR treatment groups enable a nuanced examination of the relative benefits of specific technological affordances.

Another distinguishing feature of this research is its emphasis on both immediate learning outcomes and long-term retention. By assessing operator performance at multiple time points, including a delayed retention test several weeks after the initial training, the study provides valuable insights into the durability of the learning effects associated with different IMTs. This longitudinal perspective is crucial for understanding the practical value of AR/MR technologies in supporting sustained improvements in operator performance.

The chapter also highlighted the study's meticulous attention to data collection and analysis procedures. The use of video recordings, photographs, and standardized performance metrics ensures a rich and reliable dataset for evaluating the impact of different IMTs on learning, recall, and retention outcomes. The application of appropriate statistical techniques, such as ANOVAs, effect size estimates, and regression analyses, enables a robust and nuanced examination of the research questions.

Furthermore, the study's compliance with ethical guidelines and the steps taken to ensure participant safety and confidentiality demonstrate a strong commitment to research integrity. The transparent reporting of the study's limitations, such as the convenience sampling approach and the potential influence of participant compensation, enhances the trustworthiness of the findings and provides important context for their interpretation.

In conclusion, the methods described in this chapter provide a solid foundation for the execution of this study and the subsequent presentation and interpretation of its results. The affordance-based framework, ecological validity, comprehensive assessment strategy, and rigorous data collection and analysis procedures set this research apart from previous work in the field. By addressing key gaps in the existing literature and employing a robust methodological approach, this study is well-positioned to make significant contributions to our understanding of how AR/MR technologies can be effectively leveraged to enhance manufacturing assembly training outcomes.

5.1 Study Administration and Participation

This study was administered during the Spring of 2023. Ultimately, 62 eligible participants were recruited. All completed the Learning and Recall experiments of the first session beginning February 10th. Progress of the conduct of these trials is illustrated by Figure 5.1.

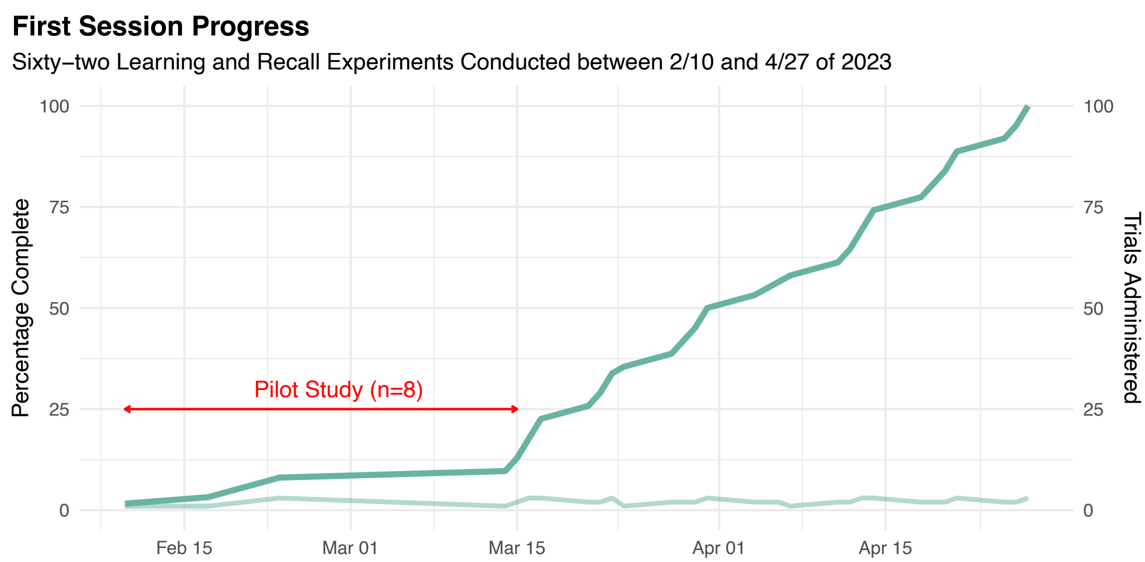


Figure 5.1: Daily Trial Administration and Completion Percentage

Only eight trials were run before March 16th. That period was treated as a pilot study, during which the research team was trained, administration methods were refined, the self-service onboarding system was deployed, and version 2.1 of our IRB application was approved. Most significantly, the intake process was adjusted to collect more demographic data, increase the role of the TLX and SUS instruments, and gather BCS data for a companion study.

Otherwise, only minor tweaks were made during the pilot to streamline the experiment's administration. Although the changes outlined above were not believed to have a material effect on the results, the data collected during the pilot were omitted from further analysis. Following the pilot, steady progress was made until the completion of the study's first session on April 27th. The second session was conducted during the end of study event on April 29th. Twenty-four of the study's original participants volunteered to attend and complete the Retention experiment.

5.2 Analysis of Participant Demographics

5.2.1 Univariate Analysis

Numeric Demographic Data

After excluding data from the pilot study, 54 participants (Mean age = 22.6, SD = 6.5, range: [19, 47], 3.7% missing) were included in this analysis. Other than age, height (Mean height = 68.6, SD = 4.2, range: [62, 76], 5.6% missing) was the only numeric demographic data collected. As discussed in Section 5.7, height was first recorded following the pilot study, when its impact on user behavior and results was first observed. Both numeric characteristics are visualized with a combined box and scatter plot in Figure 5.2. Where present, the notches of these box plots show the 95% confidence interval around the median.

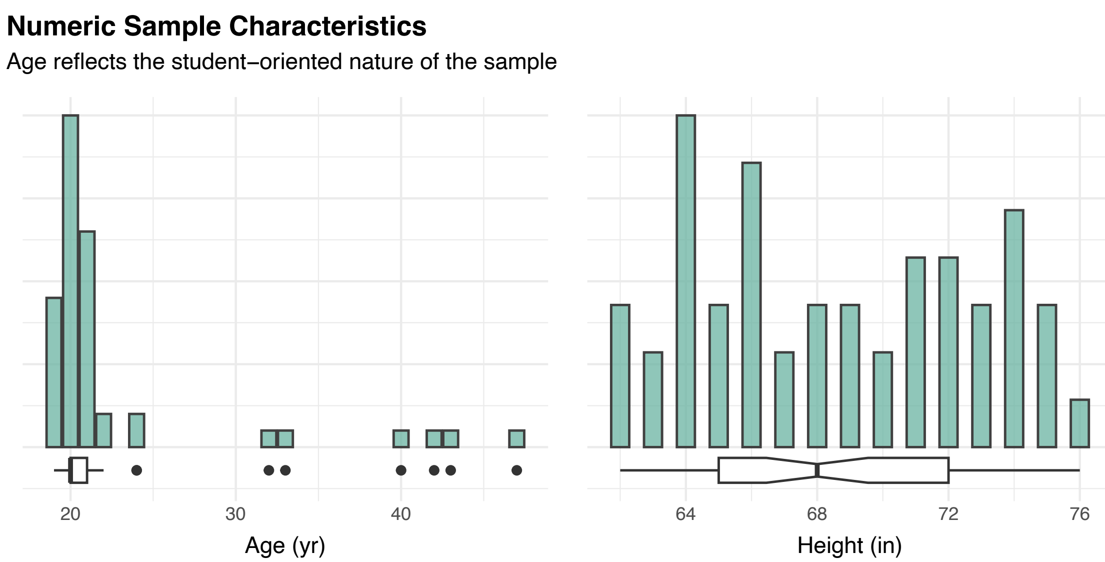


Figure 5.2: Distribution of Participant Age and Height

Categorical Demographic Data

Of the sample's 54 participants, 89% (n=48) reported that they mainly speak English at home. As shown in Figure 5.3, the gender composition was relatively balanced, with slightly more males (n=27, 50%) than females (n=26, 48%). The racial makeup of the sample was predominantly White (n=45, 83%) and Asian (n=7, 13%), with smaller representations from two other groups. Fully 98% of participants (n=52) reported a non-Hispanic or Latino ethnicity. The majority of participants (n=41, 77%) were from the United States, though

Table 5.1: Vision Correction Table

	Worn During Session			
	Yes	No	NA	**Total**
Lenses, n (%)				
Glasses	7 (47%)	5 (33%)	3 (20%)	15 (100%)
Contacts	13 (100%)	0 (0%)	0 (0%)	13 (100%)
Total, n (%)	20 (71%)	5 (18%)	3 (11%)	28 (100%)

nine other countries were represented, including S Korea (n=3) and Poland (n=2). Saudi Arabia, Germany, UK, Indonesia, Australia, India, and China were also listed as countries of origin by one participant each (n=1). All of this aligns with the study's recruitment focus.

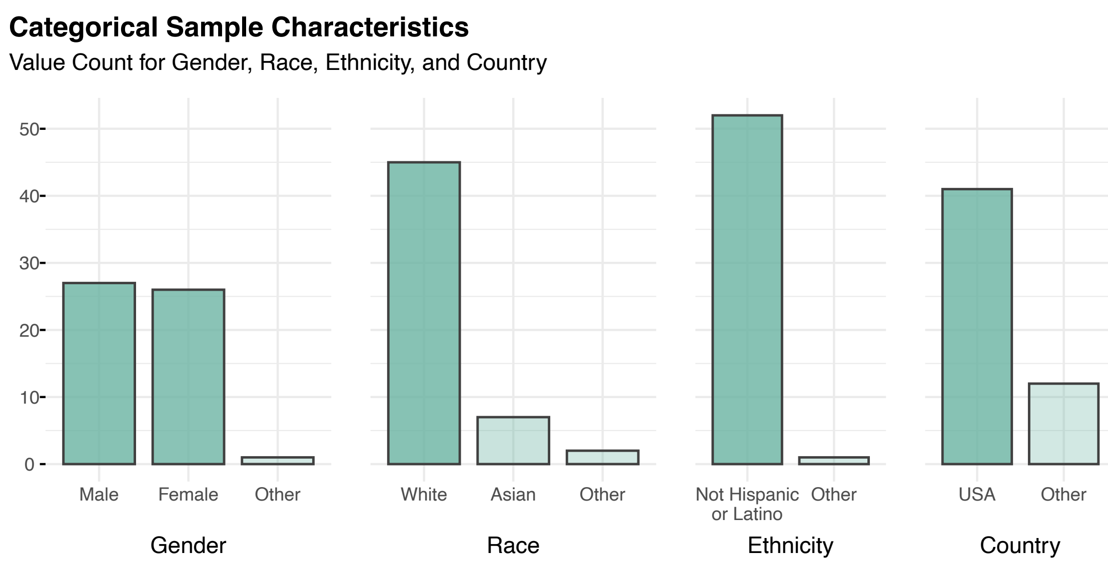


Figure 5.3: Distribution of Categorical Characteristics

Of the 54 participants, 52% (n=28) reported having a vision prescription. Of those, 13 (100%) were for contact lenses , but 15 (100%) had glasses. Whereas all contact wearers reported they would wear them during the session, among glasses wearers only seven (47%) planned the same. Five (33%) did not intend to do so, and three (20%) gave no indication. This breakdown is summarized by Table 5.1. Two participants also reported color-blindness (n=1) or other vision conditions (n=1). No participants indicated any other condition that might affect their performance.

Following the primary hypothesis testing, additional analyses may explore the differences in key outcomes across these demographic subgroups. That effort will aim to provide additional context and insights into potential underlying factors influencing the findings.

Ordinal Demographic Data

Unlike the data above, the categories analyzed in this section are inherently ordered. Responses for each participant's current Education level (Median = 3.0; Mode = 3, CND; IQR = 0.0) and their experience with Lego (Median = 2.0; Mode = 2, Some experience; IQR = 1.8) and Manufacturing (Median = 1.0; Mode = 1, No experience; IQR = 1.0) are visualized in increasing order by Figure 5.4. The results are mostly in line with expectations based on the recruitment, though it is disappointing to see that 33% of participants (n=18) claimed to have little to no experience with Lego. The great majority were undergraduate students working towards a degree (n=44, 81%), with no experience in manufacturing (n=34, 63%).

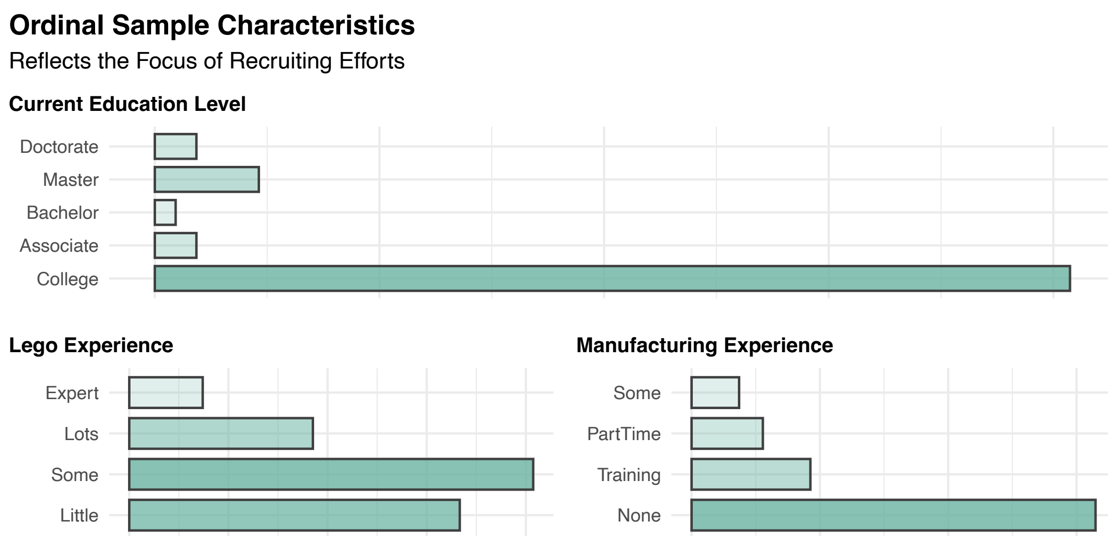


Figure 5.4: Distribution of Ordinal Characteristics

5.2.2 Multivariate Demographic Analysis

In order to assess the equivalence of treatment groups and validate the random assignment that is assumed by most statistical analysis, Age, Gender, Lego experience, and Education were compared across treatment groups. These variables were selected for their meaningful variation and potential relevance to the outcomes of interest. Other notable variables, including Ethnicity and Country of Origin were considered less critical due to their skewed distributions and small subgroup sizes.

Table 5.2 shows the result of this analysis. The Kruskal-Wallis rank sum test was utilized

Table 5.2: Participant Demographics by Treatment Group

Characteristic	**PWI** N = 14 (26%)	**PAR** N = 12 (22%)	**AR** N = 14 (26%)	**MR** N = 14 (26%)	**p-val**
Age	22 (5)	21 (4)	23 (8)	24 (8)	0.2
Gender					0.4
Female	7 (50%)	7 (58%)	8 (57%)	4 (29%)	
Male	6 (43%)	5 (42%)	6 (43%)	10 (71%)	
Non-Binary	1 (7.1%)	0 (0%)	0 (0%)	0 (0%)	
Lego					0.5
Little	4 (29%)	6 (50%)	5 (36%)	3 (21%)	
Some	4 (29%)	4 (33%)	8 (57%)	6 (43%)	
Lots	4 (29%)	2 (17%)	1 (7.1%)	3 (21%)	
Expert	2 (14%)	0 (0%)	0 (0%)	2 (14%)	
Education					0.8
College	10 (71%)	11 (92%)	11 (79%)	12 (86%)	
Associate	1 (7.1%)	0 (0%)	1 (7.1%)	0 (0%)	
Bachelor	1 (7.1%)	0 (0%)	0 (0%)	0 (0%)	
Master	2 (14%)	0 (0%)	1 (7.1%)	2 (14%)	
Doctorate	0 (0%)	1 (8.3%)	1 (7.1%)	0 (0%)	

for the numeric variable (Age) to compare distributions across groups, while Fisher's exact test was applied to the categorical and ordinal variables (Gender, Education, and Lego experience) to assess the independence of distributions from treatment assignments.

The high p-values for Age, p=0.2; Gender, p=0.4; Lego, p=0.5; and Education, p=0.5 indicate that these observed characteristics are not strongly associated with treatment groupings. This result validates the group randomization effort, mitigating the influence of confounding variables, and reducing bias in the data. It is essential to the validity of most of the statistical tests employed by this study, which assume sample independence.

5.3 H_1: Learning Phase Analysis

The results of the first phase of the experiment are tested with three hypotheses, each of which explore the affect of instructional method (treatment) on key measures of learning: average task completion time (TCT), rate of change of TCT, and the average number of uncorrected errors (UCE). Only completed tasks, the primary result of interest, are considered in this analysis. While excluding incomplete tasks may limit the generalizability of the findings to scenarios where time constraints are common, it ensures the analysis is based on objective, directly observed data and avoids introducing assumptions or biases that could arise from predicting results for unfinished tasks.

One final step of data preparation was taken to account for drop-outs and other system events, along with time spent making repairs. This lost time was deducted from measured task times to give a more accurate record of participant performance. Forty-six of 272 completed tasks (16.9%) were affected by this adjustment.

5.3.1 Descriptive Statistics for Granular Data

The data contains 272 observations of the following 2 variables:

- Task Completion Time (sec): n = 272, Mean = 97.94, SD = 44.18, Median = 88.10, MAD = 37.06, range: [36.23, 288.13], Skewness = 1.37, Kurtosis = 2.42, 0% missing
- Uncorrected Error Count: n = 272, Mean = 2.65, SD = 3.64, Median = 0.00, MAD = 0.00, range: [0, 13], Skewness = 1.34, Kurtosis = 0.69, 0% missing

These statistics provide a high-level summary of performance across all tasks and participants, where each observation corresponds to a single repetition of the Learning task. Both Task Completion Time (TCT) and Uncorrected Error Count (UCE) show considerable variance, as depicted in Figure 5.5.

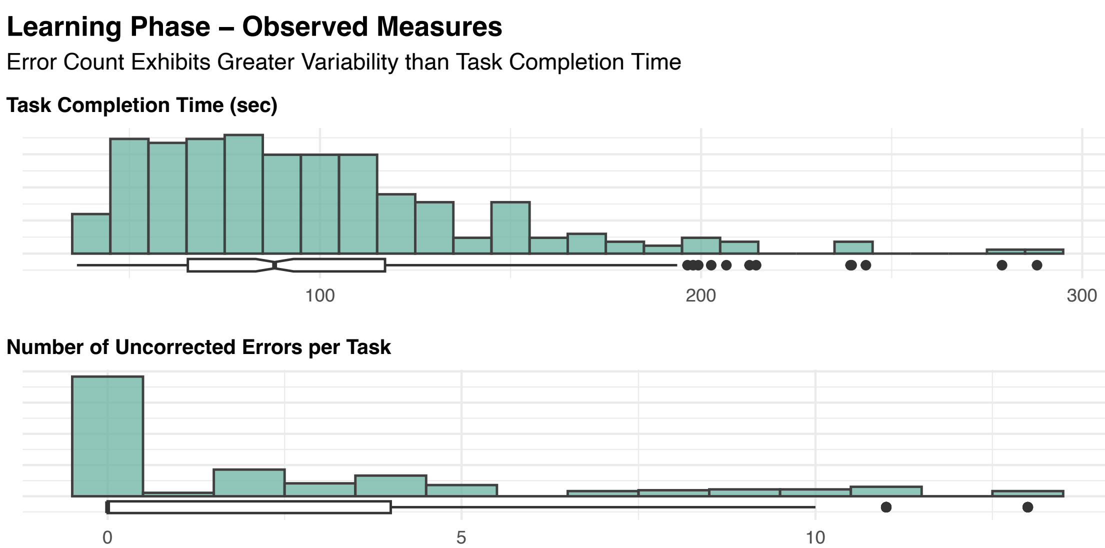


Figure 5.5: Distribution of Task Completion Time and Error Rates in Learning

Compared to TCT, with a coefficient of variation (CV) of 0.45, UCE shows substantially greater spread (CV=1.37) and the presence of a floor effect. The latter could potentially

Table 5.3: Avg. Participant Performance Data Summarized by Treatment Group

Characteristic	Overall n = 54 (100%)	PWI n = 14 (26%)	PAR n = 12 (22%)	AR n = 14 (26%)	MR n = 14 (26%)
Number of Tasks Completed					
Mean (SD)	5.0 (2.1)	7.1 (2.3)	5.9 (1.2)	3.7 (1.0)	3.6 (1.0)
Median (IQR)	5.0 (3.0, 6.0)	7.5 (5.3, 8.0)	6.0 (5.0, 7.0)	3.5 (3.0, 4.8)	3.0 (3.0, 4.0)
Avg TCT per Task (sec)					
Mean (SD)	113.2 (42.8)	79.8 (30.1)	89.6 (21.1)	142.0 (37.0)	137.9 (39.4)
Median (IQR)	107.7 (77.3, 140.6)	68.5 (65.0, 99.5)	83.8 (75.4, 104.4)	141.7 (111.8, 153.5)	127.9 (114.4, 156.4)
Avg UCE per Task					
Mean (SD)	2.2 (3.1)	6.1 (3.5)	0.8 (1.3)	0.8 (1.3)	0.7 (1.3)
Median (IQR)	0.3 (0.0, 3.5)	4.9 (4.0, 8.8)	0.3 (0.0, 0.8)	0.0 (0.0, 1.1)	0.0 (0.0, 0.5)

suggest limitations in the task's ability to capture a full range of participant performance. Further investigation is required to determine if this is due to task complexity or other factors.

5.3.2 Aggregated Data

Table 5.3 presents participant-level metrics, including the number of tasks completed along with the average TCT and UCE per task. In addition to the overall results, the data is decomposed by assigned treatment group to allow for comparisons across conditions. Note that the statistics presented are the overall (aka "grand") mean and median values. That is, they are the average or middle values of all individual participant averages. This provides overall summary statistics for each treatment group that reflect the central tendency and variability among the average outcomes reported by participants within each group.

See Figure 5.6 for a depiction of the grand means for all variables, with 95% confidence intervals for each. Observe that the previously noted variability and positive skew are also present in the grouped data, though both are less pronounced due to averaging effects. It appears that the PWI treatment group may have prioritized speed over accuracy.

Assess outliers

Given the level of variation and spread seen above, the presence of outliers was analyzed. Participant-level outliers within treatment groups were the focus, rather than individual data points across the entire dataset. This method identifies annomolous participant level contributions to treatment variability.

Four different tests were used to assess each participant as an outlier in their treatment group: (1) Tukey's Fences at 1.5 times the IQR, (2) 2.5 times the Z-score, (3) outside the

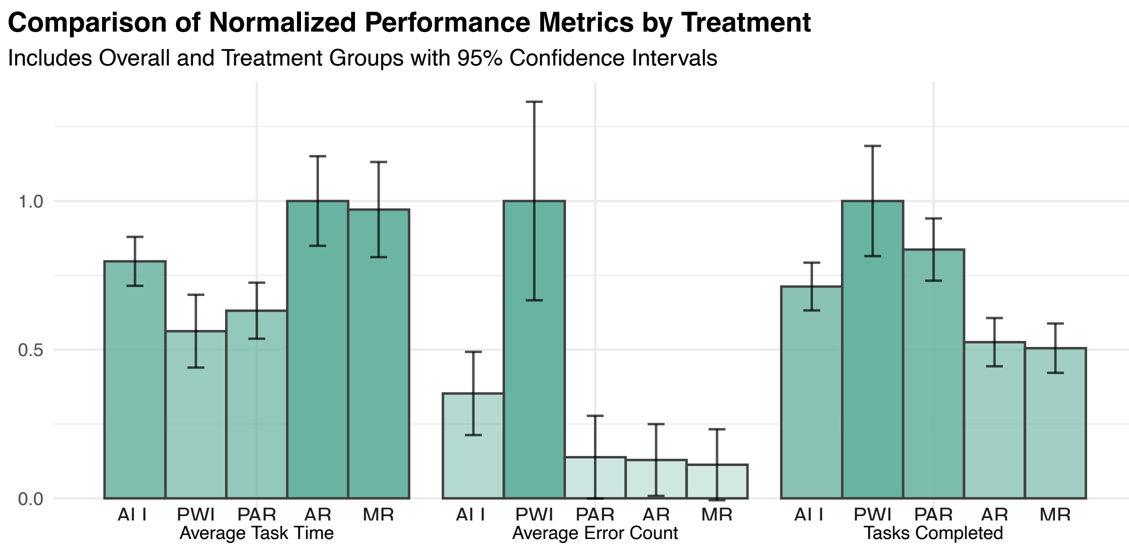


Figure 5.6: Aggregated Learning Performance Data by Treatment

5th and 95th percentiles, and (4) using the Mahalanobis distance. TCT and UCE were each tested separately using the first three methods. The fourth method is a multivariate test that considers TCT and UCE together. The results are summarized in Table 5.4, for all participants that were identified by two or more tests.

Table 5.4: Summary of Participants Identified by Two or More Outlier Tests

participant	treatment	tot_out	tot_cars	avg_tct_adj	avg_uce
1037	MR	4	2	242.2830	3.000000
1063	AR	4	2	236.3440	0.500000
1051	PAR	3	4	139.0902	4.000000
1053	AR	2	3	154.3443	3.666667

Further investigation of these four participants showed that only #1063 experienced systematic problems that might warrant excluding their performance. The others were just poor performers, with unusually high overall TCT and/or UCE. Given the limited amount of data collected per treatment, it was decided to retain all but 1063's data for hypothesis testing, where additional outlier detection steps may be taken.

5.3.3 H_{1a}: Average Time per Car

The first hypothesis of the learning phase:

$$H_{1a}\text{: Average task completion time varies with treatment}$$

was tested by comparing the average TCT of each treatment group to understand the magnitude, direction, and significance of differences.

Applicable Statistical Methods

Methods based on one-way ANOVA, including Fisher's or Welch's parametric methods and their non-parametric equivalent, Kruskal-Wallis' test by ranks, are commonly used for such comparisons. All assume observations are independent and treatment effects are additive. The use of aggregated data reduces each participant to a single observation, addressing independence. The experimental design generally ensures the treatments themselves are additive, with independent effects on the response.

Parametric methods also require a normal distribution of data *within each treatment group*. As noted in the original assessment of the learning data, the overall TCT dataset exhibits skewness and kurtosis, suggesting deviations from a normal distribution. This must be revisited with group wise testing of the aggregated data. By averaging the observations and eliminating an outlier, the grouped data may better resemble a normal distribution.

Check Model Assumptions

A combination of quantitative and qualitative analysis is required to accurately assess normality of each treatment group. Statistical tests provide a formal but imperfect measure of the data's shape. Visual confirmation of the characteristic quantile-quantile (Q-Q) plot provides further support for the claim. Table 5.5 tabulates the results of the D'Agostino skewness test, Anscombe-Glynn kurtosis test, and the Shapiro-Wilk normality test. In each case, the null hypothesis (H_0) for these tests assumes that the data follows the characteristics of a normal distribution. Therefore, low p-values indicate evidence against H_0, suggesting that the data deviates from normality in the tested aspect.

Table 5.5: Summary of TCT Test Statistics, p-Values for Normality Assessment

Group	Skewness	Kurtosis	Normality
ALL	0.68, 0.037	3.63, 0.201	0.95, 0.044
MR	1.28, 0.021	4.52, 0.052	0.89, 0.084
AR	0.01, 0.978	1.86, 0.294	0.93, 0.382
PWI	1.02, 0.058	2.95, 0.505	0.85, 0.021
PAR	1.01, 0.068	3.45, 0.221	0.9, 0.151

From this we see that only PWI fails the normality test (p=0.021), despite showing marginally non-significant skewness (p=0.058). Similarly, MR exhibits significant positive skew (p=0.021) but the Shapiro-Wilk test gives evidence of normality (p=0.084). Both AR and PAR are consistent in confirming normality, though PAR's slightly positive skew is marginally non-significant (p=0.068).

To reconcile these findings for MR and PWI, and confirm those for AR and PAR, the set of four Q-Q plots in Figure 5.7 were generated. Observations for normal data should fall along or near the line, within the shaded confidence interval. These plots show that PAR and AR both meet the expectations of normality. The MR group mostly aligns with the line, except for a single point in the right tail. This aligns with the skewness and kurtosis test results and provides insufficient evidence to reject normality. On the other hand, the PWI group diverges substantially from the line, with several points leaving the confidence interval. This supports the statistical evidence that the PWI data is not normally distributed.

Given these deviations from normality, particularly for the PWI group, a non-parametric approach was adopted. This provides a reliable and straightforward approach for hypothesis testing for data that are non-normal or heteroscedastic (have unequal variances). This decision also ensures robust statistical results in the scale of the original data, easing interpretation.

Analysis

Figure 5.8 shows the resulting comparison of average TCT within each treatment group. It combines elements of violin, scatter, and box plots to provide a comprehensive portrayal of the data. Treatment differences were analyzed using the Kruskal-Wallis test by ranks. Its test statistic, H, approximately follows a χ^2_{k-1} distribution, where k is the number of

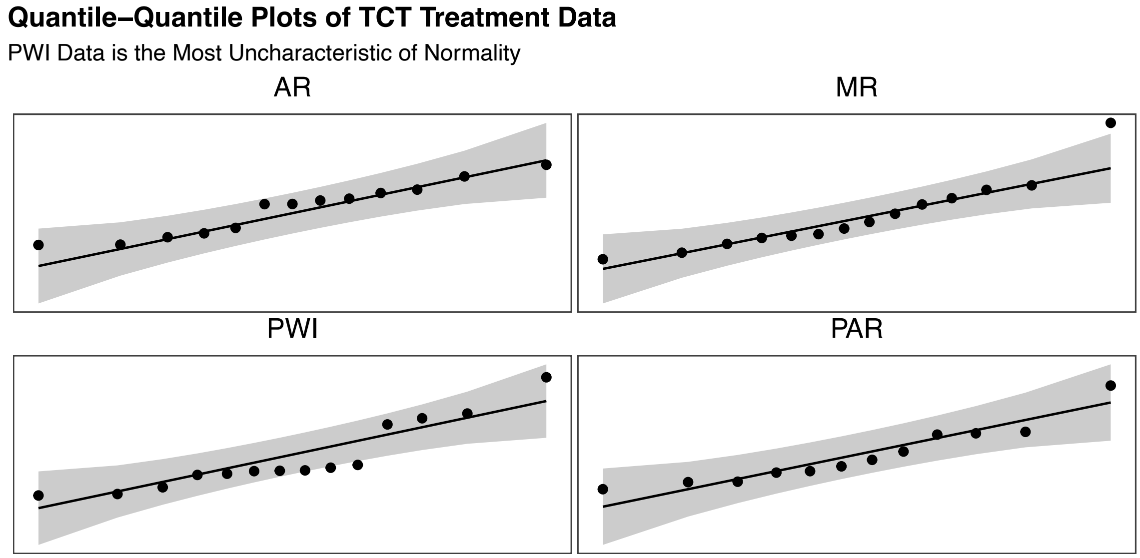


Figure 5.7: Q-Q Plots of Learning Times by Treatment

groups. Here, $\chi^2_{Kruskal-Wallis}(3) = 25.9$, with a p-value of less than 0.001. These results indicate a statistically significant difference between groups. The effect size of this difference is $\hat{\varepsilon}^2_{rank} = 0.5$, with $CI_{95\%}$ $[0.4, 1.0]$, indicating a moderate to large treatment effect where 50% of the variance in TCT can be explained by the difference in groups. Together, these results provide strong evidence of overall differences in TCT that are both statistically and practically significant.

Pairwise comparisons were conducted post-hoc using Dunn's non-parametric test for Kruskal-type ranked data. Holm's adjustment for multiple comparisons was preferred over the often cited Bonferroni method, which tends to be too conservative. The results show that PWI and PAR are both significantly faster than AR and MR, with all p-values less than 0.01.

A post-hoc simulation was run to validate these results by estimating the statistical power. Power analysis measures the effectiveness of a statistical test in detecting true differences when they exist. High power (typically 0.8 or above) suggests that the test is appropriate for the given data and experimental conditions, therefore supporting factual decision-making. The simulation repeatedly applies the Kruskal-Wallis test to data generated based on the observed mean and effect size, and resulting power is the proportion of tests that correctly reject the null hypothesis. The outcome of 1 (100% power), while uncommon, suggests that, for the conditions of this experiment (n=53, an average of 13.25 samples in each of the four groups, $\hat{\varepsilon}^2 = 0.5$, and $\alpha = 0.05$), the test is extremely effective at detecting the observed

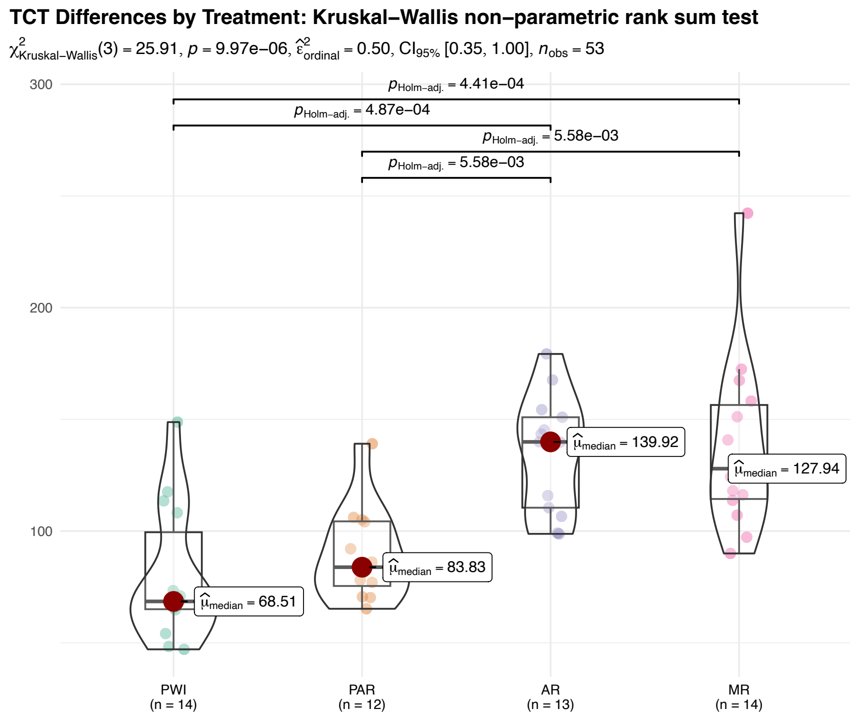


Figure 5.8: Task Completion Time Differences by Treatment

differences. The combination of a large effect size and a reasonable sample size is largely attributable to this value, which offers great confidence in the validity of the test results.

H_{1a} Result

Based on analysis of the data and validation of the methods, it is determined that sufficient statistical evidence exists to accept H_{1a}: average task completion time varies by treatment. Specifically, the PWI and PAR instructional methods are both equivalent and faster than AR and MR, which are also equivalent to one another.

5.3.4 H_{1b}: Learning Rates

The second hypothesis of the learning phase:

$$H_{1b}\text{: Learning rates vary with treatment}$$

will be tested by comparing how TCT changes with each repetition of the task based on treatment. Decreased TCT is used as a measure of the learning effect, so a change in TCT over time is a proxy for learning rate.

Applicable Statistical Methods

It is tempting to assess this using aggregated data as before. For example, the average change in TCT between consecutive tasks completed might give suitable measure. For clarity, this can be expressed as follows for each participant, j:

$$\overline{\Delta TCT_j} = \frac{\sum_{i=1}^{N-1}(TCT_{i+1} - TCT_i)}{N-1} \tag{5.1}$$

where N is the total number of tasks completed by participant j, and i indexes each of the $N-1$ differences in TCT. This metric can be compared by group using the approach employed by H_{1a}, as seen in Figure 5.9. The results show that there is a significant difference in groups (p = 0.002), with small to moderate effect size $\hat{\varepsilon}^2_{rank} = 0.29$. Significant pairwise differences were identified between PWI and both AR (p=0.027) and MR (p=0.039) as well as PAR and the same (AR p=0.015, MR p=0.025).

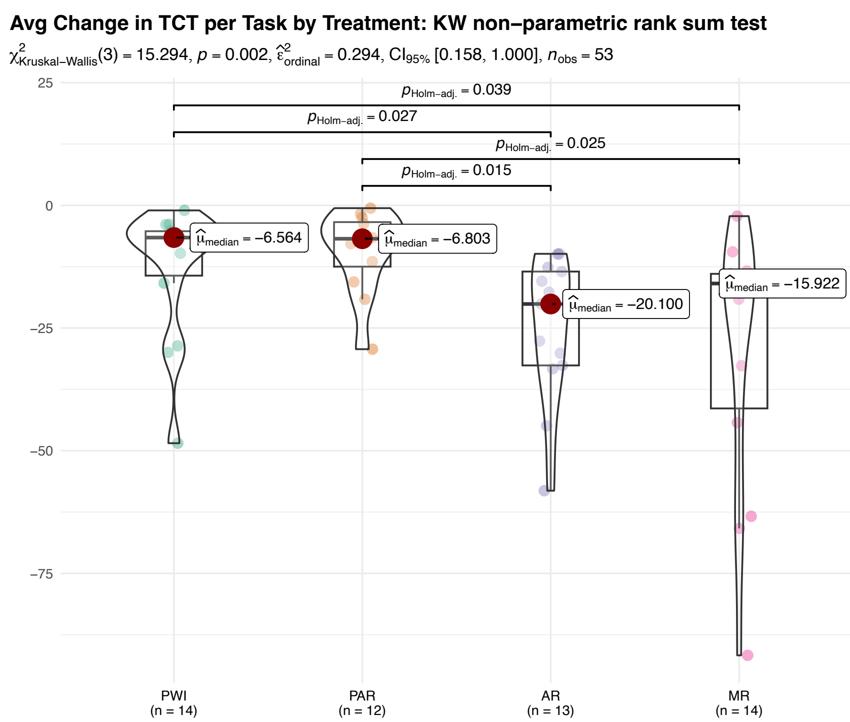


Figure 5.9: Naive method

While this method offers initial insights and highlights significant differences in learning rates across treatment groups, it has crucial limitations. Aggregating data by participant and within treatments masks both between and within-participant variability, discarding important nuances in learning rates. By ignoring the repeated measures nature of the data it fails to capture how learning rate changes over time, and how that differs between individuals. It also does not account for imbalanced repeated measures, which may lead to biased estimates. By reducing the number of observations, it may also limit the statistical power of the test. Finally, this method does not account for the non-linear nature of change in TCT over time, further limiting the validity of these findings. These simplifying assumptions may be reasonable for a gross estimate of overall differences, e.g., average TCT, but they are not appropriate for an in-depth analysis of the rate and change of learning in the presence of individual and treatment effects.

For analysis of this sort, particularly where repeated measures are nested within each group and participant variation is an important consideration, mixed-effects models are often used. These models provide a thorough and accurate analysis by accounting for individual differences in learning rates and leveraging all available repeated measures data. This approach preserves data granularity, explicitly models the dynamic aspect of learning, and is specifically designed to handle imbalanced data. It increases the statistical power and flexibility to detect effects and interactions in a complex data structure, and it can also account for non-linear response data. Overall, mixed-effects models provide a robust framework for examining the interaction between treatment and learning effect over time.

The general form of a suitable linear mixed-effects model that expresses TCT as a function of task sequence, ignoring treatment effects, is:

$$TCT_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(SEQ_{ij}) + \varepsilon_{ij} \tag{5.2}$$

where:

- i and j are the indices for each observation and participant, respectively.
- β_0 is the overall intercept, a baseline value of TCT.
- U_{0j} is the participant variation on the intercept.
- β_1 is the overall slope.
- U_{1j} is the participant variation on the slope.
- SEQ_{ij} is the time value associated with each observation i of participant j, i.e., the task number.

- ε_{ij} represents the residuals for each observation—variation not accounted for by the model.

This formulation allows for random intercepts, $(\beta_0 + U_{0j})$, and slopes $(\beta_1 + U_{1j})$, each based on individual differences per participants U_j. In the context of this analysis those can be interpreted as individual starting points (intercept) and learning rates (slope). While Equation 5.2 makes it easy to understand how participant-level variation contributes to both slope and intercept, it is more commonly formulated as:

$$TCT_{ij} = (\beta_0 + \beta_1 SEQ_{ij}) + (U_{0j} + U_{1j} SEQ_{ij}) + \varepsilon_{ij} \tag{5.3}$$

where the terms are rearranged to differentiate *fixed* and *random* effects. The first term corresponds to group-level fixed-effects (β_i) while the second captures the random-effects of between- and within-participant variation (U_j). This is visualized in Figure 5.10, where the completion time for each task iteration is shown in all treatment groups. Black lines indicate the best fit within each group, the fixed-effect term, and colored lines show the same for each participant, reflecting the random-effect term.

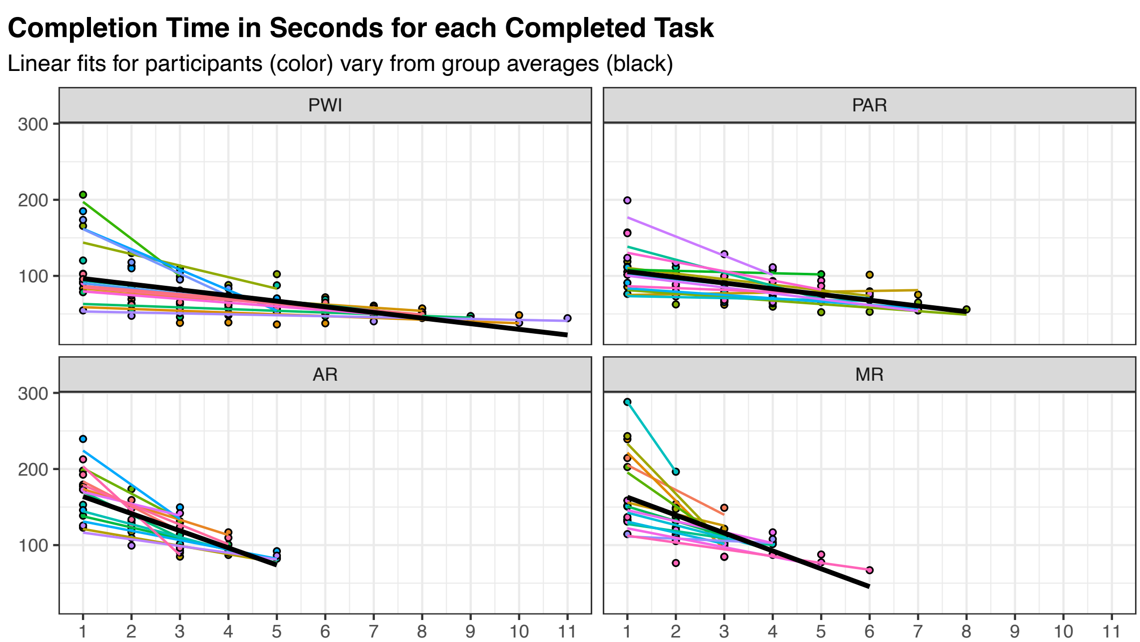


Figure 5.10: Random slopes and intercepts model visualized.

Using linear mixed effect models (LMEs) to assess treatment effect is a two-step process. First, a model using task number as the only predictor is fit to the data, including both fixed and random effects. Because this model is independent of the predictor of interest (treatment), it is referred to as the *unconditioned* model. Once selected and validated, this acts as a baseline for comparison with more complex models that incorporate the treatment effects. These so-called *conditioned* models are compared with the baseline to see which, if any, improve the model fit. Finally, the parameters of the model that best explains the data are interpreted to quantify the relationships between time, treatment, and the response. The following sections will elaborate on each step of this process.

Unconditional Models of Time

To find an unconditional model that best fits the data, several options of increasing complexity were considered. In the model selection process, it is important to consider that learning is an inherently non-linear process. Typically, TCT starts high and decreases with each training iteration, but the rate of decrease is not constant. The pattern often resembles exponential decay, with rapid initial improvement followed by more gradual changes. When working with an LME, quadratic terms can be used to create a curvilinear approximation of this effect. This is depicted in Figure 5.11, where each line is fit using a quadratic term for sequence: $TCT \approx \beta_0 + \beta_1 SEQ + \beta_2 SEQ^2$. The result appears provide an overall improvement to fit at both participant and treatment levels compared to the previous linear formulation.

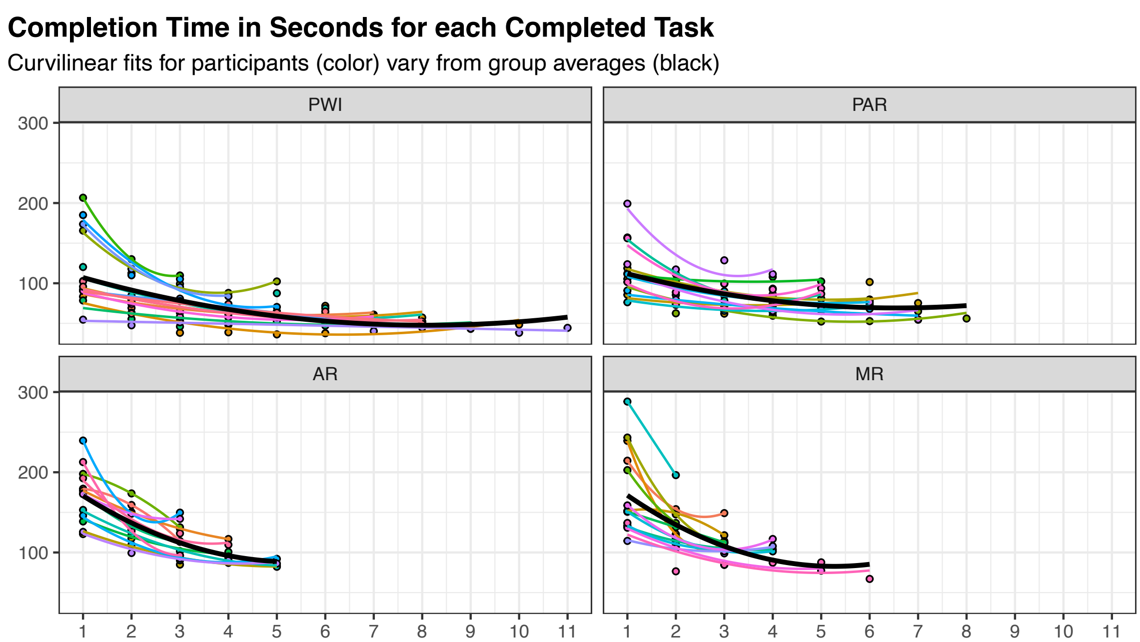


Figure 5.11: Quadratic random slopes and intercepts model visualized.

Seven models of increasing complexity were fit, as summarized in Table 5.6. This table adopts the model formulation notation used by many R modeling packages, including `lmer`, which is used here. A fixed overall intercept is specified for all models with the first term (`1`). Random effects are described within the parentheses of the last term, where terms to the right of the bar (`|`) are grouping variables, and those to its left specify the random intercept and coefficients. Terms between the intercept and random effects, outside the parentheses, are fixed effects with constant coefficients. For brevity, Sequence and Participant terms have been abbreviated in this table as `s` and `p`, respectively. The "XF" column describes the non-linear transformation employed (quadratic or log), and "SE" identifies the sequence effect. SE is "Fixed" where SEQ is included only as a fixed effect, and "Both" where it also appears in the random effect.

Table 5.6: Model Summary

#	**Model Formulation in `{lmer}` Notation**	**XF**	**SE**
1	1 + (1 | p)	None	None
2	1 + s + (1 | p)	None	Fixed
3	1 + s + (1 + s | p)	None	Both
4	1 + s + s^2 + (1 + s + s^2 | p)	Quad	Both
5	1 + log(s + 1) + (1 | p)	Log	Fixed
6	1 + log(s + 1) + (1 + log(s + 1) | p)	Log	Both
7	1 + s + log(s + 1) + (1 + s + log(s + 1) | p)	Log	Both

For example, model 4 has a fixed overall intercept (`1`) with additional fixed effects for `s` and `s^2`. Its random effects (`1 + s + s^2 \| p`) can be interpreted as random intercepts and slopes for `s` and `s^2` for each participant. This can be expressed in the previous mathematical notation as:

$$\begin{aligned} TCT_{ij} = (&\beta_0 + \beta_1 SEQ_{ij} + \beta_2 SEQ_{ij}^2) \\ &+ (U_{0j} + U_{1j} SEQ_{ij} + U_{2j} SEQ_{ij}^2) + \varepsilon_{ij} \end{aligned} \tag{5.4}$$

Prior to fitting, the sequence predictor was transformed from discrete integers to continuous values in the range [0, 1]. This is common practice to prevent numerical issues, aid

Table 5.7: Observations per Treatment for all recorded task iterations

	1	2	3	4	5	6	7	8	9	10	11
PWI	14	14	14	13	12	10	9	7	3	2	1
PAR	12	12	12	12	11	7	4	1	0	0	0
AR	13	13	13	7	4	0	0	0	0	0	0
MR	14	14	13	6	2	1	0	0	0	0	0

model convergence, and to make the results easier to interpret and generalize. Additionally, The treatment variable was converted to an unordered factor to remove any implied ranking among treatments that could bias the models. Treatment was originally ordered to ensure the PWI, PAR, AR, MR presentation of results in charts and tables. It has no predictive implications.

Finally, only a subset of the data is used. As seen in Figure 5.10 and Figure 5.11, the number of task iterations completed in the allotted time varied substantially by treatment. Numerous attempts to fit the full dataset using various modeling techniques all resulted in significant overfitting, particularly after the fifth iteration where available observations decline rapidly and treatments become imbalanced (see Table 5.7). The overfit models produced poor extrapolations that were not reflected in performance diagnostics, as these measures only compare observed and predicted values. To address this, it was decided to use data only from the first five task iterations, where all treatments are still reasonably represented. This cutoff was chosen as the best trade-off between available data and extrapolation quality, after considering earlier (four) and later (six) alternatives.

Following those data transformations, the models were fit using maximum likelihood estimation (rather than the default restricted maximum likelihood) to enable valid comparisons between models with different fixed effects structures. Fitting was accomplished using the overloaded version of `lme4::lmer` from `{lmerTest}`, which adds p-values for fixed effects using Satterthwaite's method. The results were compared using `performance` from `{compare_performance}`, the output of which is summarized in Table 5.8.

Columns include Akaike (AIC) and Bayesian (BIC) Information Criterion, conditional and marginal R^2, Interclass Correlation Coefficient (ICC), and Root Mean Squared Error (RMSE). AIC and BIC are, respectively, frequentist and Bayesian measures of the amount of information lost by a model when approximating the true data generating process.

Both account for the number and relevance of predictors, and lower scores are better. R^2_{cond} represents the variance explained by both fixed and random effects, while R^2_{marg} only accounts for the fixed effects. ICC measures the proportion of total variance that is accounted for by the grouping of data. In this context, high ICC values suggest that a larger portion of the variability in TCT is due to differences *between* participants, not within (i.e., across task iterations). Finally, RMSE is a measure of the average magnitude of the model's prediction error (residuals).

Table 5.8: Model Performance Results

Name	AIC	BIC	R2 (cond.)	R2 (marg.)	ICC	RMSE
m1	2242.07	2252.32	0.61	0.00	0.61	24.56
m2	2127.64	2141.31	0.75	0.19	0.69	18.17
m3	2060.66	2081.16	0.84	0.30	0.77	14.51
m4	1960.11	1994.27	0.94	0.28	0.91	7.93
m5	2105.54	2119.21	0.78	0.21	0.72	17.00
m6	2027.19	2047.68	0.88	0.32	0.82	12.23
m7	1943.30	1977.47	0.95	0.28	0.93	7.10

The result of this comparison shows that m7 outperforms the other options in most measures, with the lowest AIC, BIC, and RMSE. Approximately 95% of the variance in TCT can be explained by this model (R^2_{cond}), of which only 28% is due to its fixed effects (R^2_{marg}). The high ICC (0.93) indicates that a large proportion of the total variance is due to differences between participants, suggesting that individual variation is a crucial factor in the model. The resulting predictions are plotted in Figure 5.12.

Even with the subset of data, we see that the log-based model begins to exhibit a slight upward trend after the fourth iteration, which defies learning theory, expectations, and the limited data from subsequent iterations. Model fits based on exponential and hyperbolic forms, which better align with theoretical learning curves, led to more promising predictions but were plagued by convergence issues, especially as additional terms were added to account for treatment and its interactions. This is a limitation of the study that originates primarily from the imbalanced experimental design employed by the learning phase.

Inspecting the fixed effects parameters for m7, summarized in Table 5.9, shows that all terms are significant. The intercept of 145.5 seconds establishes the baseline average performance for all participants on the first iteration. Estimated coefficients for the linear and

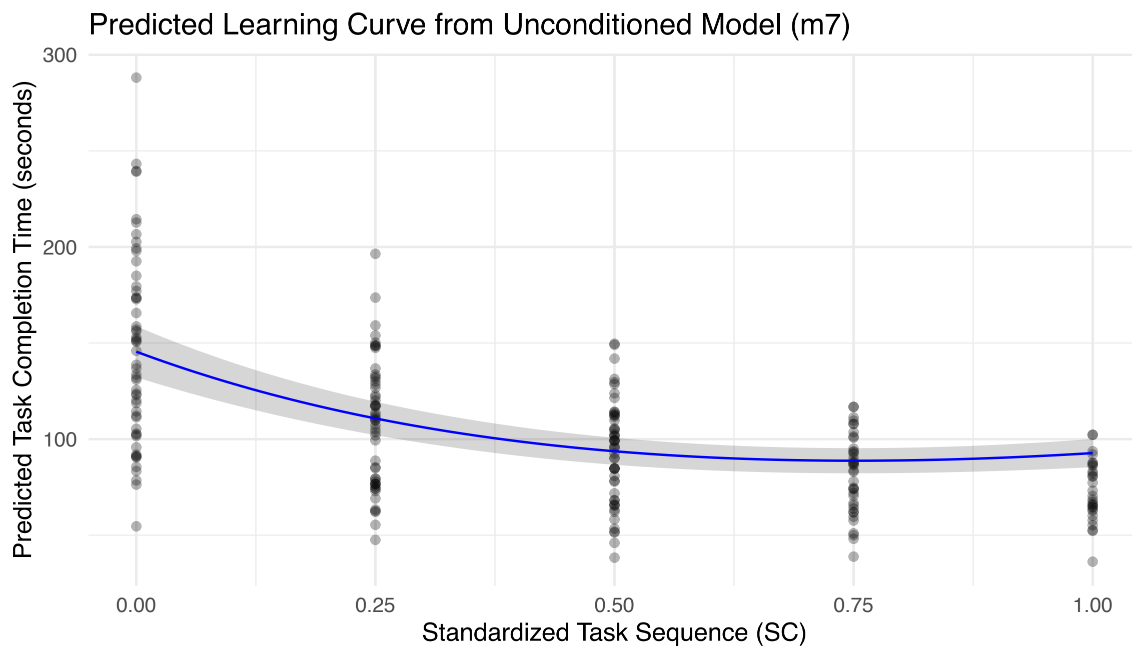


Figure 5.12: Learning Curve for Unconditioned Model

log terms of sequence show that these terms compete for dominance, with the linear term increasing TCT at a rate of 246.5 seconds per unit of seq.fp, the continuous transformation of task iteration number. This is offset by the log term, which decreases TCT by 431.7 seconds per unit of $\log(\text{seq.fp} + 1)$. These coefficients represent the change in TCT for a one-unit change in their respective transformed sequence terms, resulting in a non-linear learning curve when combined. The strong negative correlation of -0.75 between the log term and intercept indicates that participants with higher initial TCT experience a stronger effect from the log term, which drives rapid initial improvement. In learning theory, this phenomenon is known as the *power law of practice*. It formalizes the idea that, all other things constant, the farther a learner starts from expected proficiency, the faster they will initially improve.

Table 5.9: Fixed Effects

Parameter	Coefficient	SE	95% CI	t(215)	p	Sig
(Intercept)	145.496	6.776	(132.14, 158.85)	21.472	< .001	***
seq.fp	246.481	28.833	(189.65, 303.31)	8.549	< .001	***
log(seq.fp + 1)	-431.712	45.465	(-521.33, -342.10)	-9.495	< .001	***

Validation of m7 confirmed that the residuals exhibit equal variance (p = 0.16), but are not normally distributed (Shapiro-Wilk test, p = 0.004). Traditional QQ plots of the raw residuals do not properly account for the hierarchical structure and non-independence of mixed effect models, leading to overly optimistic results. The `{DHARMa}` package was specifically designed to address this problem by comparing observed residuals to a large number of simulations from the fitted model. This technique provides a more robust and accurate representation of model accuracy. Diagnostics produced by `testResiduals` indicate deviation from the expected distribution of residuals (Kolmogorov-Smirnov test, p = 0.03) without significant dispersion issues (p = 0.86) or outliers (p = 0.43). Visual inspection of residual plots confirmed these findings (Figure 5.13).

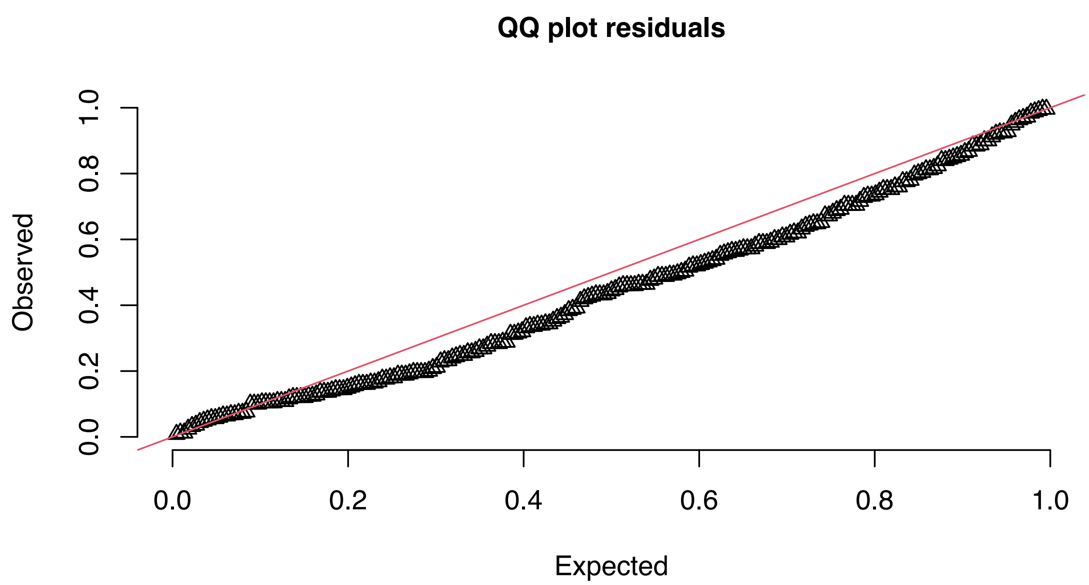


Figure 5.13: Residual Plots

Given these minor deviations from model assumptions, the model was refit using a robust linear mixed model designed to mitigate those issues and ensure reliable inference. Robust LMs (RLMs) use techniques such as M-estimation or MM-estimation to reduce the influence of high-leverage observations, and weighted estimation to equalize variance. This approach makes RLMs less sensitive to violations of normality assumptions in both fixed and random effects, and more resistant to departures from other common assumptions of linear mixed models. As a result, RLMs typically provide more stable and accurate parameter estimates when dealing with non-ideal data conditions.

The result from `robustlmm::rlmer` is compared with the original fit in Table 5.10. Robust fits for two additional variations of m7 are also included. Both add the initial TCT as a predictor, first as a covariate, and then including interactions with the linear and log terms in the second. This was done to account for the impact of baseline performance on the model, as discussed above. As seen in Table 5.10, both models incorporating initial TCT (m7r_ti and m7r_ta) show much higher R^2_{marg} values, indicating that initial performance explains a large portion of the fixed effects variance. The version with interactions also exhibits the best ICC and RMSE scores, suggesting it is the best choice for further development.

Table 5.10: Model Performance Results

Name	R2 (cond.)	R2 (marg.)	ICC	RMSE
m7	0.95	0.28	0.93	7.10
m7r	0.96	0.29	0.95	6.97
m7r_ta	0.99	0.78	0.96	7.50
m7r_ti	0.99	0.89	0.90	6.16

These results, particularly the decrease in RMSE (6.16), indicate meaningful performance gains were achieved by the robust fit. However, the real benefit comes from improved model diagnostics. The robust model reduces heteroscedasticity (p = 0.31) and down-weights extreme values as seen in Figure 5.14. Fourty-one residuals and six random effects were affected, addressing potential outliers or overly influential observations. The low residual error achieved (SD = 4.26) aligns with the high R^2_{cond} previously noted. While the Shapiro-Wilk test still indicates non-normality of residuals ($p < 0.001$), this is less concerning for robust models, which are designed to perform well even when normality assumptions are violated.

As a final model validation, Figure 5.15 shows the predictions generated by the robust model for the first five participants in each treatment. In this figure, the black circles and lines represent observed TCT values for each participant, while the colored curves represent the RLM's predictions. Each row corresponds to a different treatment group (AR, MR, PAR, PWI), and each subplot shows data for an individual participant.

This visualization allows for a comparison between observed data and model predictions across different treatments and participants, and helps confirm that a good fit is obtained for the specified range of observation counts. The variation in TCT is evident here, with participants assigned to the PWI and PAR groups completing substantially more task iterations than their AR / MR peers.

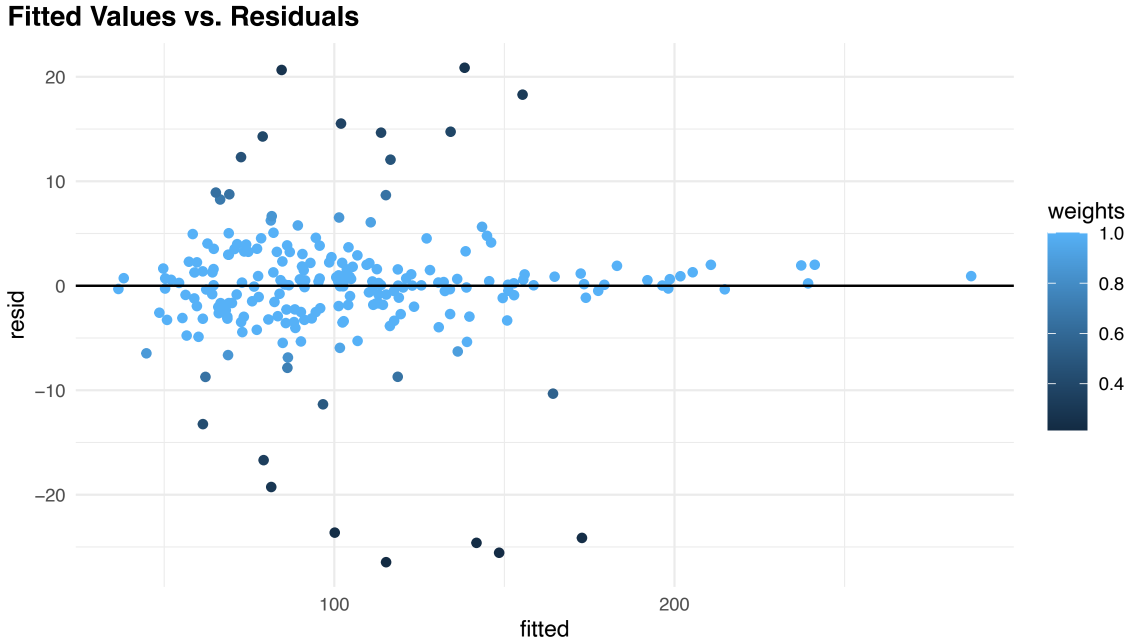


Figure 5.14: Extreme values are assigned lower weights in robust model

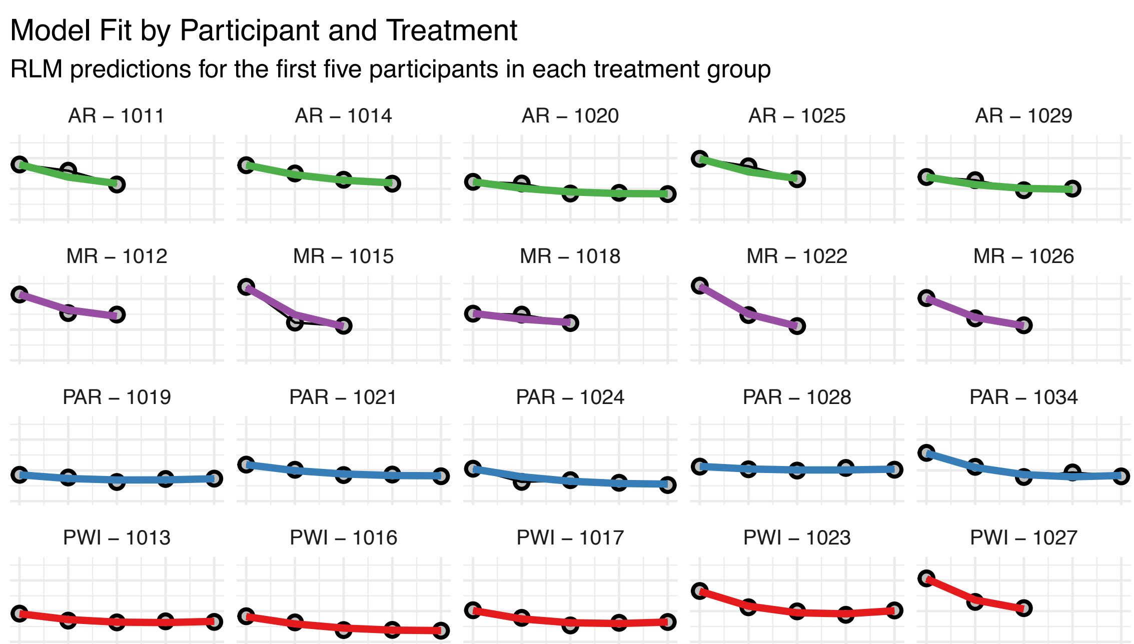


Figure 5.15: Visualizing RLM model fit

Table 5.11: Fixed Effects Parameters for the Unconditioned RLM including initial TCT interactions (m7r_ti)

Table 5.11: Fixed Effects

Parameter	Coefficient	SE	CI	t	p	Sig
(Intercept)	0.189	1.832	(-3.40, 3.78)	0.103	0.918	
seq.fp	-224.576	60.753	(-343.65, -105.50)	-3.697	< .001	***
initial_tct	0.996	0.012	(0.97, 1.02)	83.539	< .001	***
log(seq.fp + 1)	354.028	85.711	(186.04, 522.02)	4.130	< .001	***
seq.fp:initial_tct	3.359	0.435	(2.51, 4.21)	7.715	< .001	***
initial_tct:log(seq.fp + 1)	-5.539	0.599	(-6.71, -4.36)	-9.244	< .001	***

Table 5.11 summarizes the fixed effects parameters for m7r_ti, revealing a complex interplay of factors influencing task completion time (TCT). The intercept term is near-zero (0.189) and not likely significant[1] (t = 0.10). This counter-intuitive result is a direct consequence of including initial TCT as an explanatory variable, centering predictions around each participant's starting point and calculating deviations from initial performance rather than absolute TCT values. The initial_tct coefficient (0.996, t = 83.54) is highly significant and nearly 1, indicating that initial performance is the dominating predictor of subsequent performance, a role played by the intercept in prior formulations. The linear (seq.fp: -224.6, t = -3.70) and logarithmic (log(seq.fp + 1): 354.0, t = 4.13) terms for sequence are both significant and compete for dominance, with signs reversed from the m7 model. This sign reversal is another consequence of including initial TCT. The larger magnitude of the logarithmic term suggests a strong non-linear component to learning, aligning with theory, expectations, and observed values. Significant interactions between initial TCT and both the linear (3.36, t = 7.72) and logarithmic (-5.54, t = -9.24) terms reveal how baseline performance affects TCT changes over time. These interactions suggest that participants with higher initial TCT tend to have a slower linear decrease but stronger logarithmic decrease in TCT over time.

Finally, the model's random effects structure allows for participant-level variations in both

[1] For this interpretation, t values with a magnitude greater than 2.0 are taken as likely significant.

the intercept and slope for the linear and log terms. Together, this sophisticated design incorporates the power law of practice in a nuanced, participant-specific manner. The combination of linear and logarithmic terms, along with their interactions with initial performance, allows it to represent complex, dynamic learning patterns that vary based on sequence, starting proficiency, and participant idiosyncrasies.

To ease its interpretation, this model can be expressed mathematically as:

$$\begin{aligned}\text{TCT}_{ij} = {} & 0.189 + (0.996 \times \text{TCT}_{0i}) - (224.6 \times \text{SC}_{ij}) + \big(354.0 \times \log(\text{SC}_{ij} + 1)\big) \\ & + (3.36 \times \text{SC}_{ij} \times \text{TCT}_{0i}) - \big(5.54 \times \log(\text{SC}_{ij} + 1) \times \text{TCT}_{0i}\big) \\ & + \text{RE}_{ij} + \varepsilon_{ij}\end{aligned} \tag{5.5}$$

Where:

- i and j are indices for participant and task iteration
- TCT_{ij} is the Task Completion Time in seconds
- TCT_{0i} is the initial Task Completion Time
- SC_{ij} is the standardized sequence count, $\text{SC} = (\text{SEQ} - 1)/(\max(\text{SEQ}) - 1)$
- RE_{ij} represents the random effects
- ε_{ij} is the residual error term

Note that RE_{ij} can be further broken down as:

$$\text{RE}_{ij} = U_{INTi} + (U_{LINi} \times \text{SC}_{ij}) + \big(U_{LOGi} \times \log(\text{SC}_{ij} + 1)\big) \tag{5.6}$$

Where the participant-specific random effects for intercept, linear, and log terms are:

- $U_{INTi} \sim \mathcal{N}(0, \sigma^2_{INT})$ and $\sigma_{INT} = 0.749$
- $U_{LINi} \sim \mathcal{N}(0, \sigma^2_{LIN})$ and $\sigma_{LIN} = 108.34$
- $U_{LOGi} \sim \mathcal{N}(0, \sigma^2_{LOG})$ and $\sigma_{LOG} = 160.37$

These terms are participant-specific (as noted by the subscript i), and have a complex correlation structure. Notably, there is a strong negative correlation (-0.99) between the random effects for seq.fp and log(seq.fp + 1), indicating these terms strongly counteract each other at the participant level. The intercept has a moderate negative correlation with seq.fp (-0.68) and a moderate positive correlation with log(seq.fp + 1) (0.79). Finally, the residual error, $\varepsilon_{ij} \sim \mathcal{N}(0, 4.256^2)$. Together, these specifications provide a complete picture of the model's structure and variability.

Conditional Models and Hypothesis Testing

To understand the treatment effects and test H_{1b}, a model conditioned on treatment must be fit. Using m7r_ti as a base, three models of increasing complexity were again fit. The first includes treatment as a covariate predictor, the second includes interactions with the linear term, and the final model adds interactions with the log term. As before, these were fit with `robustlmm::rlmer`, and performance of the resulting models is compared in Table 5.12.

Table 5.12: Comparison of Model Performance

Name	R2 (cond.)	R2 (marg.)	ICC	RMSE
m7r	0.96	0.29	0.95	6.97
m7r_ti_c1	0.98	0.83	0.89	7.51
m7r_ti_c2	0.99	0.93	0.86	5.93
m7r_ti_c3	0.99	0.65	0.97	7.07

All models incorporating treatment as a fixed effect show a substantial increase in R^2_{marg}, suggesting it improves the model's explanatory power. Model m7r_ti_c2, which includes treatment interactions with the linear term, performs better than the alternatives in every metric (R^2_{marg} = 0.93, ICC = 0.85, RMSE = 5.93) except R^2_{cond}, where it matches the most complex model (0.99). Furthermore, where the other models produced singularity warnings during the fit, c2 did not. These warnings did not invalidate the results for c1 or c3, as `robustlmm` automatically negotiated to an alternate optimizer. But the added stability provides additional comfort.

Figure 5.16 plots the predicted learning curves for each treatment based on their mean TCT_0. The black line shows the predictions from the prior model, without treatment effects. All treatments show a general non-linear decrease in TCT with sequence, demonstrating learning. PWI and PAR perform better than average, while AR and MR are slower. This ranking persists through all iterations, despite clear differences in initial performance. To quantify these trends in statistical terms, Table 5.13 summarizes the parameters for the model's fixed effects.

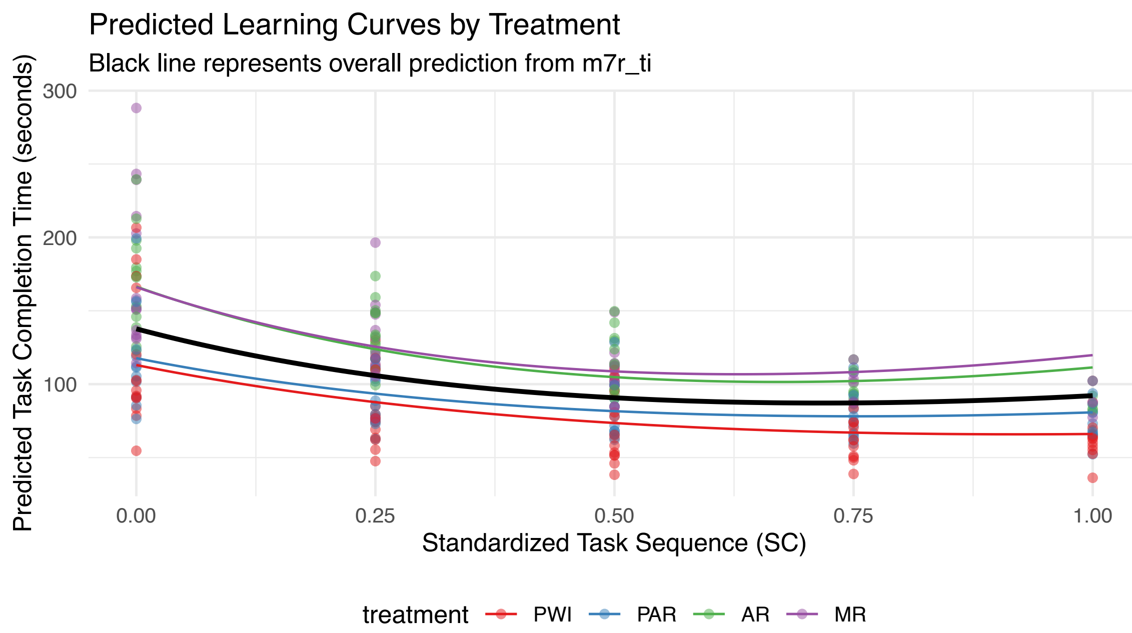


Figure 5.16: Learning Curves by Treatment

Table 5.13: Fixed Effects

Parameter	Coefficient	SE	CI	t	p	Sig
(Intercept)	0.553	1.903	(-3.18, 4.28)	0.291	0.771	
seq.fp	-260.217	81.759	(-420.46, -99.97)	-3.183	0.001	**
initial_tct	0.989	0.013	(0.96, 1.01)	73.906	< .001	***
treatmentPAR	0.006	1.544	(-3.02, 3.03)	0.004	0.997	
treatmentAR	1.842	1.672	(-1.43, 5.12)	1.102	0.271	
treatmentMR	1.206	1.665	(-2.06, 4.47)	0.725	0.469	
log(seq.fp + 1)	399.331	111.641	(180.52, 618.14)	3.577	< .001	***
seq.fp:initial_tct	3.529	0.562	(2.43, 4.63)	6.278	< .001	***
seq.fp:treatmentPAR	12.707	5.396	(2.13, 23.28)	2.355	0.019	*
seq.fp:treatmentAR	21.055	6.513	(8.29, 33.82)	3.233	0.001	**
seq.fp:treatmentMR	29.953	6.529	(17.16, 42.75)	4.588	< .001	***
initial_tct:log(seq.fp + 1)	-5.897	0.757	(-7.38, -4.41)	-7.787	< .001	***

These results are similar to those for the unconditioned model (Table 5.11). All terms previously found significant remain so, including TCT_0, the linear and log terms, and their interactions with TCT_0. The relevant interpretations stand, though the magnitude of the effects vary slightly. Importantly, all interactions between treatment and seq.fp are identified as significant. This directly addresses H_{1b}, confirming that treatment varies TCT over time, when TCT_0 varies. To test if these findings hold when all participants begin with equal proficiency (equalizing the effects of the power law of practice), a second set of predictions is generated using equal values of TCT_0 for all treatments. The resulting learning curves are plotted in Figure 5.17.

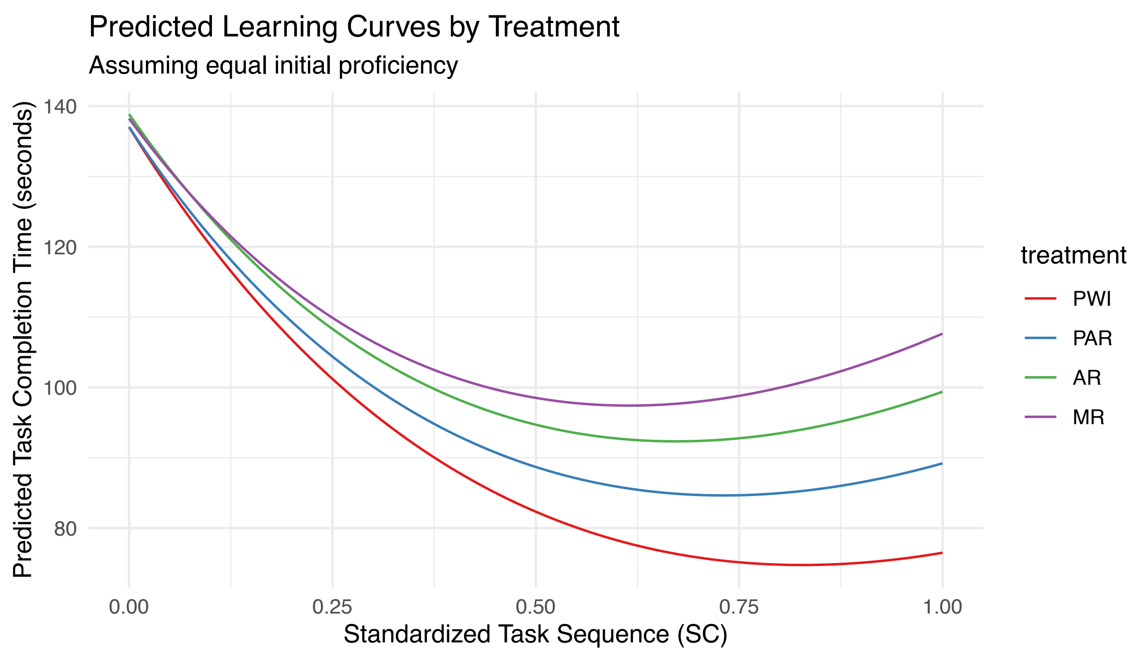


Figure 5.17: Compare Learning Curves with Equal Initial Proficiency

This plot shows that, even when the effects of TCT_0 are standardized, the instructional method influences the learning rate. To quantify this effect, `emtrends` from the `emmeans` package is used to calculate the rate of change in the response variable with respect to specified predictors. In the context of this analysis, it is used to calculate the slope of TCT with respect to sequence for each treatment while holding TCT_0 constant. Pairwise comparisons of the resulting slopes allow for statistical tests, as seen in Table 5.14. These results show significant differences between the learning rates for PWI and both AR ($p = 0.007$) and MR ($p < 0.001$) as well as PAR and MR ($p = 0.04$). The overall ranking of effectiveness can be expressed as PWI ≈ PAR, PWI > (AR ≈ MR), and PAR > MR.

While Figure 5.17 clearly illustrates the differences in learning curves, and Table 5.14

Table 5.14: Average Slope of TCT over Task Sequence by Treatment

contrast	estimate	SE	df	z.ratio	p.value
PWI - PAR	-12.71	5.40	Inf	-2.355	0.0861
PWI - AR	-21.05	6.51	Inf	-3.233	0.0067
PWI - MR	-29.95	6.53	Inf	-4.588	<.0001
PAR - AR	-8.35	6.48	Inf	-1.289	0.5701
PAR - MR	-17.25	6.52	Inf	-2.643	0.0409
AR - MR	-8.90	6.62	Inf	-1.343	0.5351

P value adjustment: tukey method for comparing a family of 4 estimates

provides compelling statistical evidence of these effects, both approaches focus on absolute changes in task completion time. As a complementary perspective, Figure 5.18 expresses progress as cumulative percentage improvements. This approach offers some advantages:

1. It allows for comparison of relative progress across treatments, which can be informative when participants start at different performance levels.
2. Percentage improvements can sometimes highlight subtle differences in learning patterns that are less apparent in absolute time measures.
3. It provides a way to visualize learning gains that is independent of the scale of the original task completion times.

While the result may appear similar to the inverse of earlier plots, it offers a standardized view of learning progress across treatments that aligns closely with how learning is often conceptualized in cognitive and educational theories. This representation helps confirm and clarify the trends observed in our previous analyses, reinforcing our understanding of how each treatment impacts learning over time.

H_{1b} Results

Based on comprehensive analysis of learning curves and statistical modeling, there is strong evidence to accept H_{1b}: the rate of improvement in task completion time varies significantly by treatment. Specifically, PWI demonstrates the steepest learning curve, followed closely by PAR. Both AR and MR show slower rates of improvement. The relationship can be summarized as: PWI ≈ PAR, PWI > (AR ≈ MR), and PAR > MR. This pattern holds true even when controlling for initial task completion time, suggesting that the treatment effects on

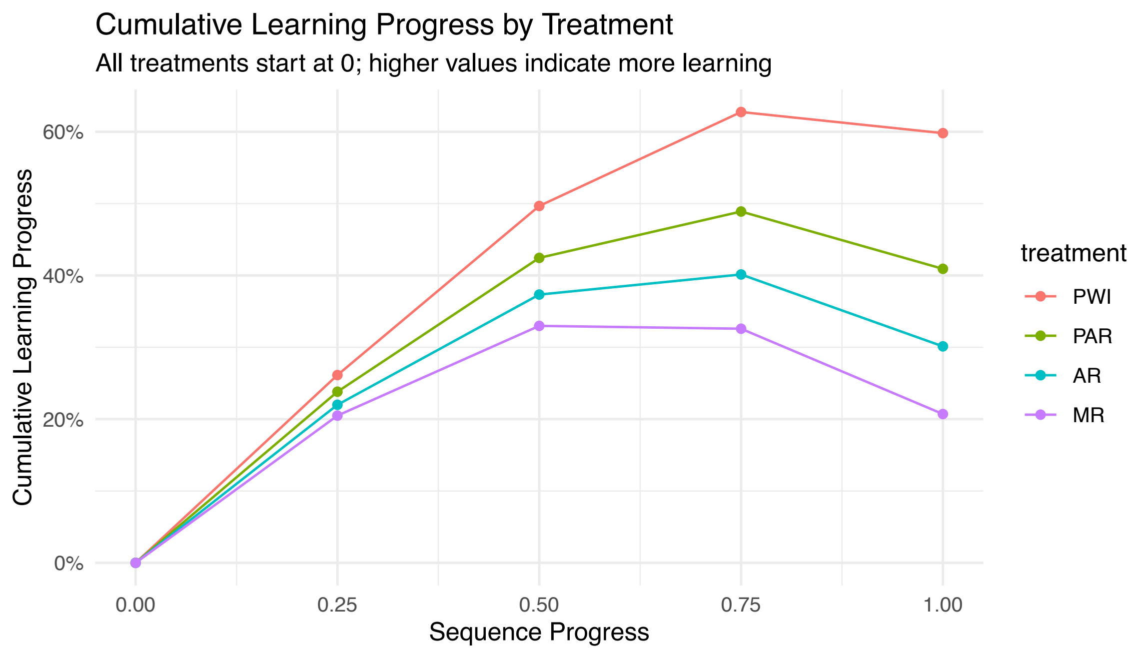


Figure 5.18: Cumulative Learning Progress by Treatment

learning rates are robust. The differences in learning rates are most pronounced in the early stages of task repetition and tend to converge over time, highlighting the importance of instructional method particularly in the initial phases of skill acquisition. However, these findings should be interpreted with the described limitations in mind, primarily those related to the limited extrapolation capabilities resulting from an imbalanced data set and statistical model formulation. In light of these considerations, great effort was invested to ensure the robust and reliable findings outlined above.

5.3.5 H_{1c}: Average Error Count per Car

The third and final hypothesis of the learning phase tests the third commonly assessed aspect of learning:

$$H_{1c}\text{: Average error count per car varies with treatment}$$

Here again, we are interested in finding statistically significant differences in the average error count under different treatment conditions to identify which instructional method leads to the lowest defect rate. Lower error rates can indicate a higher quality of learning outcomes. Combined with the previous measures of efficiency (TCT) and learning rate

Table 5.15: Summary Statistic for Dependent Variables, H1c Quality

	Mean	SD	Min	Median	Max
Average Uncorrected Errors	2.19	3.15	0.00	0.33	12.00

(change in TCT per task), these provide a balanced and comprehensive view of training performance.

Descriptive Statistics

As with previous measures, the uncorrected error count (UCE) is neither independent nor balanced. Additionally, it is a discrete (counted) variable. All of these have implications for the statistical method used, but where average counts are used, many are addressed. The standard five number summary for UCE is presented in Table 5.15.

As this data is aggregated by participant, there is only one observation for each (n = 53). As is common in this study, variation is high (SD = 3.15) compared to the mean value (2.19). Many participants made no errors, but one averaged 12 errors per task iteration, resulting in an overall median value of 0.33. The distribution of UCE is illustrated by Figure 5.19. As seen before, the overall average error count data exhibits floor effects that contribute to a strongly positive skew. Obvious non-normality (Shapiro-Wilk, $p < 0.001$), skew (D'Agostino, $p < 0.001$), and kurtosis (Anscombe-Glynn, $p = 0.02$) are all confirmed in the overall dataset using the appropriate tests.

Finally, Shapiro-Wilk tests are performed for each treatment group, revealing strong evidence that PAR ($p < 0.001$), AR ($p < 0.001$), and MR ($p < 0.001$) are not normally distributed, but PWI is ($p = 0.61$). These findings are supported by the Q–Q plots in Figure 5.20. Furthermore, it is important to note that this is counted data. Despite the fact that means / grand means may appear continuous in this aggregated form, the underlying distribution remains discrete. Traditional parametric tests are not recommended in this case.

Analysis

The Kruskal-Wallis non-parametric test is used to compare average UCE by treatment, the results of which are shown in Figure 5.21. Large differences in the overall mean are

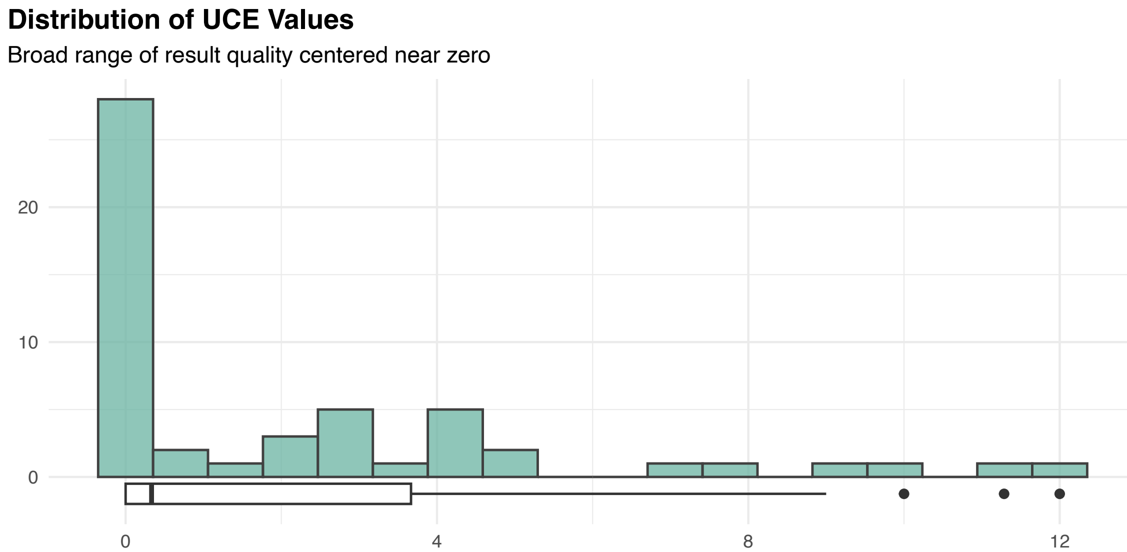


Figure 5.19: Histogram of Average Error Count in Learning

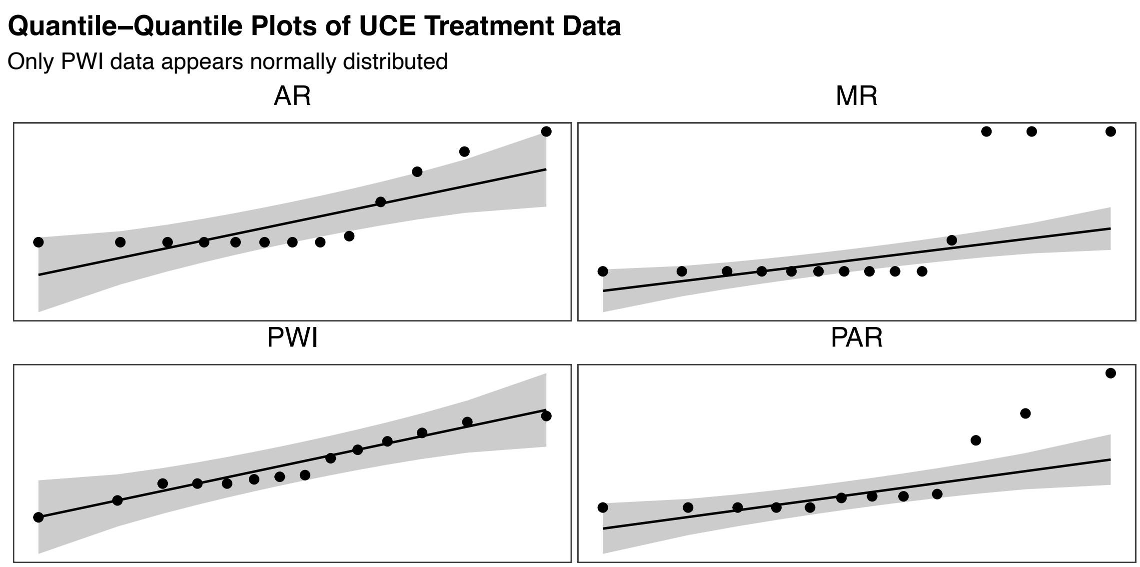


Figure 5.20: Q-Q Plots of UCE by Treatment

observed, along with some floor effects in the augmented treatments. The test statistic $\chi^2_{KW}(3) = 25.5$, with a p-value < 0.001, indicates a statistically significant difference between groups. The effect is $\hat{\epsilon}^2_{rank} = 0.49$, with a $CI_{95\%}$ $[0.32, 1.00]$, reflecting the data's variability. The results of Dunn's Holm-adjusted pairwise comparisons show statistically significant differences between PWI and all other treatments (p « 0.01 in all cases), while PAR, AR, and MR treatments are statistically similar. These results formally quantify the obvious reduction in error count experienced with any form of augmented instruction.

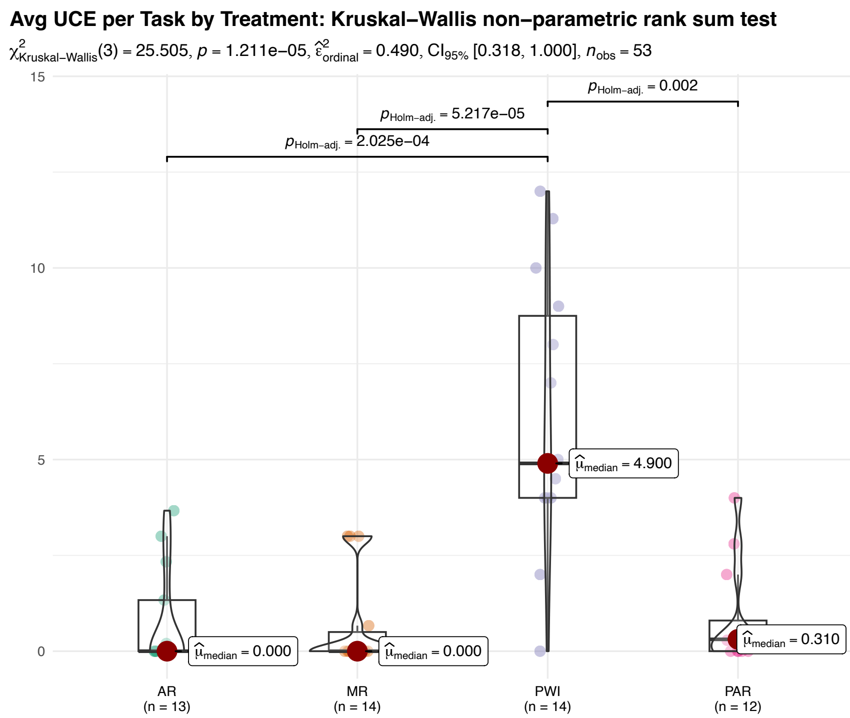


Figure 5.21: Differences in Average Uncorrected Error per Task by Treatment

KW is a robust test that does not assume normality or homogeneity of variances. It can be sensitive to the presence of outliers, three of which were detected in the UCE data based on a combination of IQR and Z-score. Participants 1016, 1017, and 1056 were all assigned to the PWI treatment and averaged 12, 10, and 11.3 errors per task, respectively. A second Kruskal-Wallis test of the data without those outliers showed the difference remained significant (p < 0.001). This confirms that the outliers did not impacting the outcome of the test.

H_{1c} Results

H_{1c} is accepted - all augmented instruction methods result in error rates that are drastically lower in magnitude than PWI, the statistical significance of which is confirmed. Floor effects may imply that the task complexity was insuffcent to capture the full range of participant performance. Further analysis using a generalized linear mixed model with poisson or negative binomial distribution for count data could be used to better account for participant variation but that seems unnecessary given the strong observed differences. While this method was deemed insufficient for capturing the nuances of learning rate analysis, it is entirely appropriate for comparing average error counts.

5.4 H_2: Recall Phase Analysis

As with the learning phase results, only completed tasks are considered for recall analysis. Paper work instructions were available for reference, and participants wore the HL2 only for recording purposes. This almost entirely eliminated system-related interruptions, though participant 1057 inadvertently triggered the HL2 menu twice during their trial. The time lost to those events was deducted from measured task times as before. 212 observations were recorded during this phase, as expected for 53 participants, each with four iterations.

5.4.1 H_{2a}: Overall Equipment Effectiveness

The first hypothesis of the recall phase:

$$H_{2a}\text{: OEE varies with treatment}$$

is designed to determine if the instructional method has an impact on overall performance after the initial training period. As described in Section 4.3.1, OEE is the product of performance (task rate relative to takt time), quality (percentage of units produced without defects), and availability (percentage of system up time). For the purposes of this analysis, 100% availability is assumed. Takt time is set to 60 seconds in accordance with the line policy and instructional design.

Descriptive Statistics

The calculated measures are summarized below:

- Productivity: n = 53, Mean = 0.84, SD = 0.19, Median = 0.82, MAD = 0.18, range: [0.41, 1.47], Skewness = 0.20, Kurtosis = 1.28
- Quality: n = 53, Mean = 0.56, SD = 0.49, Median = 1.00, MAD = 0.00, range: [0, 1], Skewness = -0.25, Kurtosis = -1.96
- Overall Equipment Effectiveness: n = 53, Mean = 0.48, SD = 0.43, Median = 0.62, MAD = 0.58, range: [0, 1.14], Skewness = -0.07, Kurtosis = -1.77

As this is aggregated data, there is one observation for each participant, and no missing values. Productivity and quality are slightly skewed right (0.20) and left (-0.25), respectively. Kurtosis measures suggest that productivity is close to normal (1.28), but quality has a flatter distribution (-1.96). Productivity exhibits relatively low variability and similar mean (0.84) and median (0.82), suggesting consistency among participants. For quality, the median (1.00) and its absolute deviation (MAD = 0.00) show that many participants achieved perfect quality during recall. However, the range (0, 1) and high standard deviation (0.49) relative to its mean (0.56) indicates considerable variability. In fact, several participants produced *only* defective assemblies. Note that quality is essentially a discrete variable, as it can only take on values of $n/4$, where $n = 0, 1, \ldots, 4$, corresponding to each of the four task iterations.

OEE shows a notably asymmetric distribution with a mean of 0.48 and a higher median of 0.62, indicating a significant number of low values dragging down the mean. A high concentration of zeros are observed in the OEE histogram, Figure 5.22, suggesting that many participants experience zero effectiveness due to high defect rates. The standard deviation (sd = 0.43) and median absolute deviation (MAD = 0.58) reflect substantial variability in effectiveness across participants. Despite the approximately symmetric skewness (-0.07), the range (0 to 1.14) indicates diverse levels of effectiveness, from complete failures to high performance. These findings highlight the variability in participants' abilities to achieve consistent and effective output.

Results from the Shapiro-Wilk test provide further evidence against the normality hypothesis for OEE, both overall (p < 0.001) and by treatment (p < 0.05 in all cases). Q-Q plots in Figure 5.23 suggest that OEE is close to normally distributed for all augmented treatments, but PWI is clearly not. No outliers were detected in the OEE data based on the Z-score test. Finally, Levene's test indicated no significant heterogeneity of variance between treatment groups (p = 0.43). These findings again dictate a non-parametric approach to testing H_{2a}.

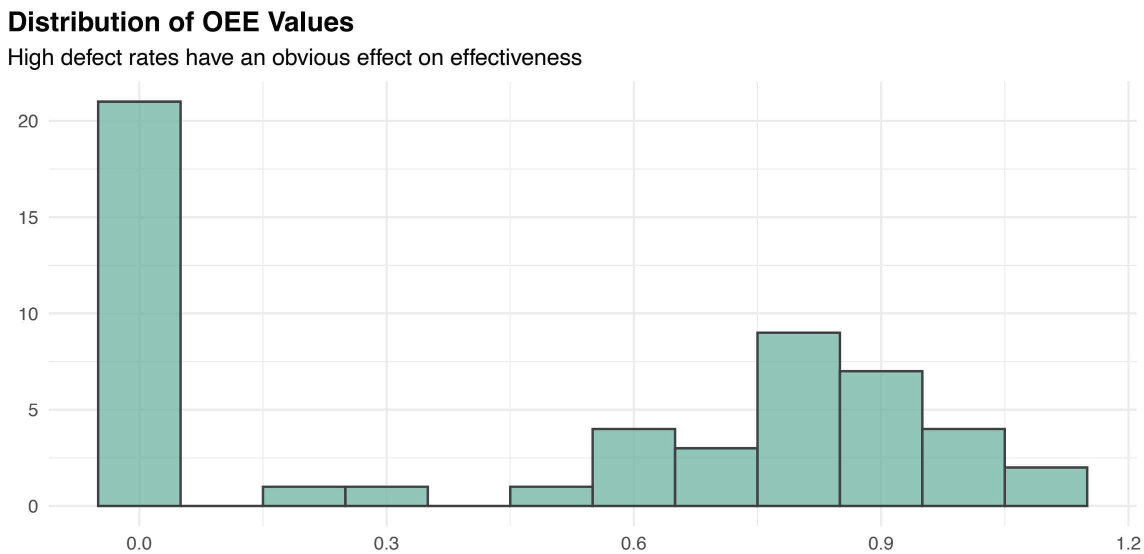


Figure 5.22: Histogram of Overall Equipment Effectiveness in Recall

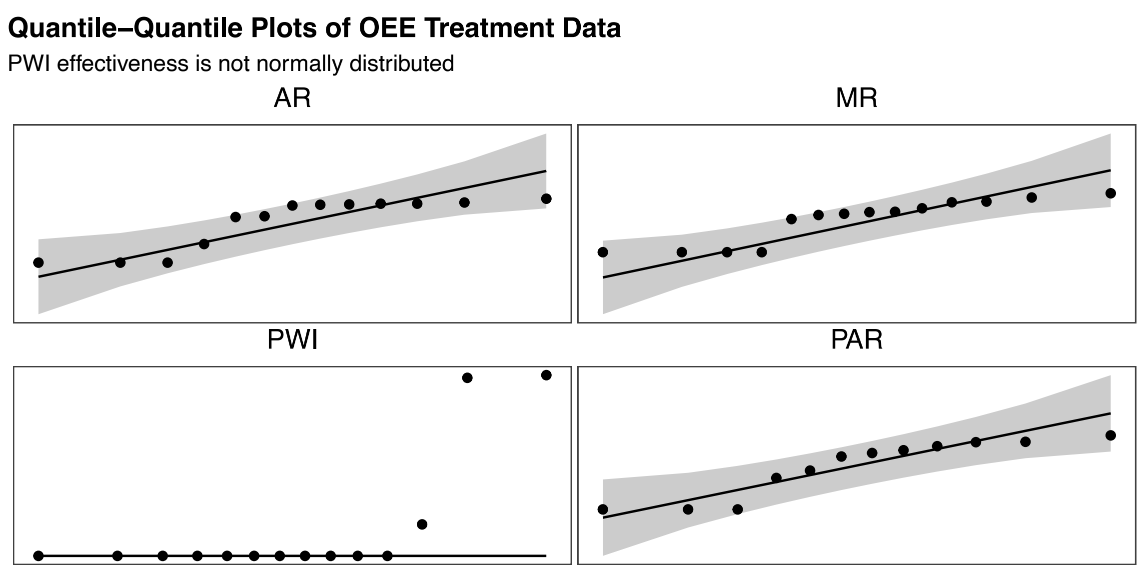


Figure 5.23: Q-Q Plots of OEE by Treatment

Analysis

In Figure 5.24, group-wise significant differences are found (p = 0.02), with with small to moderate effect size $\hat{\epsilon}^2_{rank} = 0.19$ and a broad $CI_{95\%}$ $[0.08, 1.0]$. Pair-wise tests show only PWI and PAR are significantly different (also p = 0.02). A post-hoc simulation was performed to assess the power of the analysis. The result of 85.6% indicates the analysis had sufficient power to detect significant differences, which lends further support to the reliability of this outcome.

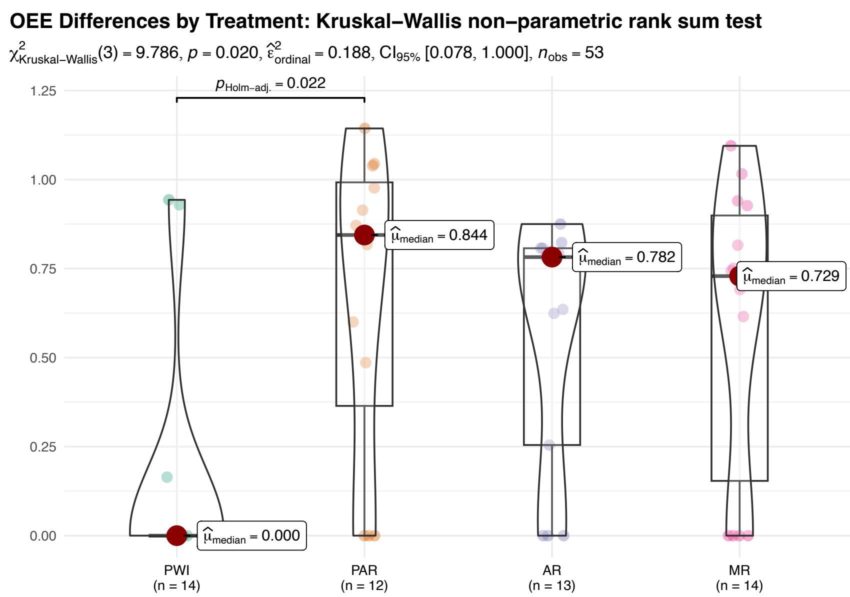


Figure 5.24: H_{2a} Test - OEE Differences by Treatment

This unexpected result indicates that, despite *substantial observed differences*, the PWI, AR, and MR treatments have *statistically equivalent* OEE. High within-group variability likely contributes to this finding by obscuring genuine differences in group means. Kruskal-Wallis test is somewhat vulnerable to this effect due to its rank-based design and underlying assumptions. The ratio of interquartile range to overall range for PAR (0.55), AR (0.63), and MR (0.68) show that the central 50% of observations in these treatments is spread over more than half of its overall range. This wide variation is easily observed in the box plots of Figure 5.24.

To further investigate the observed variability and validate these unexpected findings, bootstrap resampling analysis was employed. This flexible technique samples with replacement to construct robust estimates of the distribution for the statistics of interest (e.g., mean or median) and their confidence intervals. Bootstrapping relies on the empirical distribution of the data to accurately capture variability with few assumptions. Additionally, the replication process effectively increases the sample size and statistical power of the analysis, enhancing the accuracy and reliability of its results.

A bootstrap analysis was used to estimate pairwise differences in three measures of centrality for OEE–mean, median, and the Hodges-Lehmann estimator. The latter is a non-parametric approach to estimate population median shifts that is robust to outliers. It provides a consistent estimate of the true shift in medians, even when the underlying distributions are dissimilar, by considering the differences between all possible observation pairs. This process was replicated five times, each with 10,000 resamples, and the results were averaged to determine grand means and related statistics. The significant results of this complementary analysis, where the 95% confidence interval does not include zero, are tabulated in Table 5.16.

Table 5.16: H_{2a} Bootstrap Analysis of Pairwise OEE Differences

Significant OEE Differences in Centrality: Bootstrapped Estimates

Five Replications of 10,000 Resamples for Mean, Median, and the Hodges-Lehmann Estimator

Comparison	Stat	Mean	95% CI	Bias	Std Error
PWI - PAR	Mean	-0.512	$[-0.792, -0.204]$	-0.001 (0.000)	0.149 (0.000)
PWI - AR	Mean	-0.408	$[-0.643, -0.140]$	-0.001 (0.000)	0.128 (0.000)
PWI - MR	Mean	-0.448	$[-0.700, -0.167]$	-0.001 (0.000)	0.136 (0.000)
PWI - PAR	Median	-0.844	$[-1.008, -0.243]$	0.075 (0.000)	0.194 (0.000)
PWI - AR	Median	-0.782	$[-0.807, -0.255]$	0.088 (0.000)	0.167 (0.000)
PWI - PAR	HL_Est	-0.704	$[-0.955, -0.271]$	0.047 (0.000)	0.185 (0.000)
PWI - AR	HL_Est	-0.565	$[-0.798, -0.232]$	0.022 (0.000)	0.170 (0.000)
PWI - MR	HL_Est	-0.664	$[-0.842, -0.238]$	0.065 (0.000)	0.170 (0.000)

Significance indicated by confidence intervals that do not include zero.

Compared to the initial findings, bootstrapping identifies a number of significant pairwise differences. Even when testing OEE median values, as done with Kruskal-Wallis, this method finds both PWI-PAR and PWI-AR significant. The mean tests, which are more sensitive to variance and outliers, also include PWI-MR. Hodges-Lehmann matches the mean findings, despite its reduced sensitivity.

This result supports the previous findings of significant overall differences in OEE, as well as between the PWI and PAR conditions for all measures of centrality. In addition, the bootstrap analysis, with its robust estimation methods, provided strong evidence of significant differences between PWI and both the AR (for all measures) and MR groups (except for median). These results show that the OEE of PWI is substantially lower than all other treatments, with average estimated differences of 0.687 (PAR), 0.585 (AR), and 0.307 (MR). The negligible bias values and low standard error reported across all replicates reinforce the reliability of these findings.

H_{2a} Result

As both approaches identified a significant treatment effect on OEE, there is ample evidence to accept H_{2a}: OEE does vary with instructional method. However, the conflicting pairwise analysis results must be reconciled. While the Kruskal-Wallis test only identified significant differences between the PWI and PAR treatments, subsequent bootstrapped analysis revealed additional differences between PWI and the AR and MR treatments using a variety of measures. Considering the bootstrap method's advantages when faced with high between-group variance, limited sample size, and non-normally distributed data, the consistent differences and robust confidence intervals it produced provide compelling evidence that the observed differences are reliable. Ultimately, it is determined that PAR, AR, and MR treatments all result in statistically significant and practically meaningful improvements in OEE compared to PWI, which was hampered by a high defect rate.

5.4.2 H_{2b}: Paper Work Instruction References

The second hypothesis of the recall phase assesses an alternative measure of training effectiveness:

$$H_{2b}\text{: PWI reliance varies with treatment}$$

Reliance is quantified by the number of times each participant refers to the paper work instructions (PWI count), and the duration of each event (PWI time). PWI count is considered a measure of the *frequency* of the participant's task uncertainty, while PWI time is associated with the *degree* of uncertainty experienced.

Descriptive Statistics

Overall statistics for the measures of interest (n = 212) are summarized below:

- PWI Count per Task: Mean = 1.47, SD = 4.12, Median = 0.00, MAD = 0.00, range: [0, 30], Skewness = 4.51, Kurtosis = 23.27, 0% missing
- Total PWI Time per Task: Mean = 2.86, SD = 9.43, Median = 0.00, MAD = 0.00, range: [0, 77.18], Skewness = 5.64, Kurtosis = 36.73, 0% missing

Despite low mean values for count and total time (1.47 and 2.86, respectively), both exhibit substantial variability, with standard deviations of 4.12 and 9.43. High skewness (4.51 for count, 5.64 for time) and kurtosis (23.27, 36.73) confirm their strong right skew and heavy tails. Together with a median and MAD of zero for both measures, these statistics suggest that many participants rarely referred to the instructions, but a small number relied on them heavily.

The statistical similarity between count and time suggests a high correlation, confirmed by Spearman's rank correlation test (rho = 0.99, $p < 0.001$). This is natural and expected for two dependent measures, but must be accounted for. To create a stable measure of overall reliance for each participant, a composite measure was created by averaging the time per PWI reference across all four tasks. This approach reduces the impact of task-specific factors and eliminates the correlation between reference count and duration without diminishing either aspect of task uncertainty. Using a single, composite measure simplifies the analysis and interpretation of the results by reducing dimensionality. It also enables a traditional univariate approach to the comparison, which benefits from reductions to both within– and between–participant variance due to averaging. The distribution of average time per PWI reference is visualized in Figure 5.25.

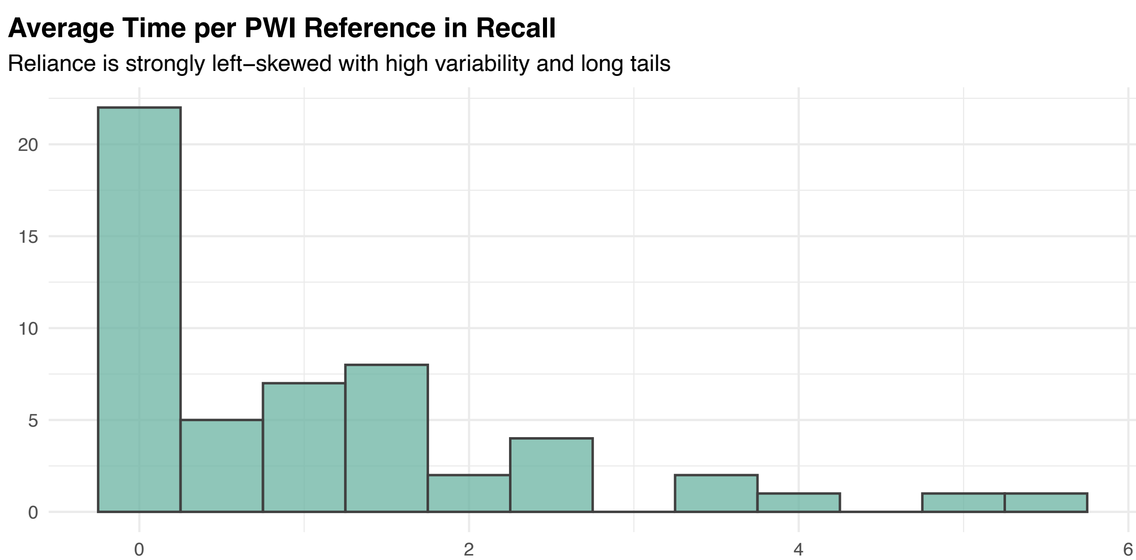


Figure 5.25: Histogram of Composite PWI Measure in Recall

From the descriptive statistics and visualization, this data does not appear normal. To confirm, the Shapiro-Wilk test was administered for the overall data ($p < 0.001$) and within each treatment group. The results show that neither PWI ($p < 0.001$) nor PAR ($p < 0.001$) are normally distributed, but AR is ($p = 0.16$), and MR shows marginal evidence ($p = 0.05$) of non-normality. Visual inspection of the Q-Q plots for each treatment, provided in Figure 5.26, supports these findings. Levene's test ($p = 0.95$) confirms equal variance by treatment.

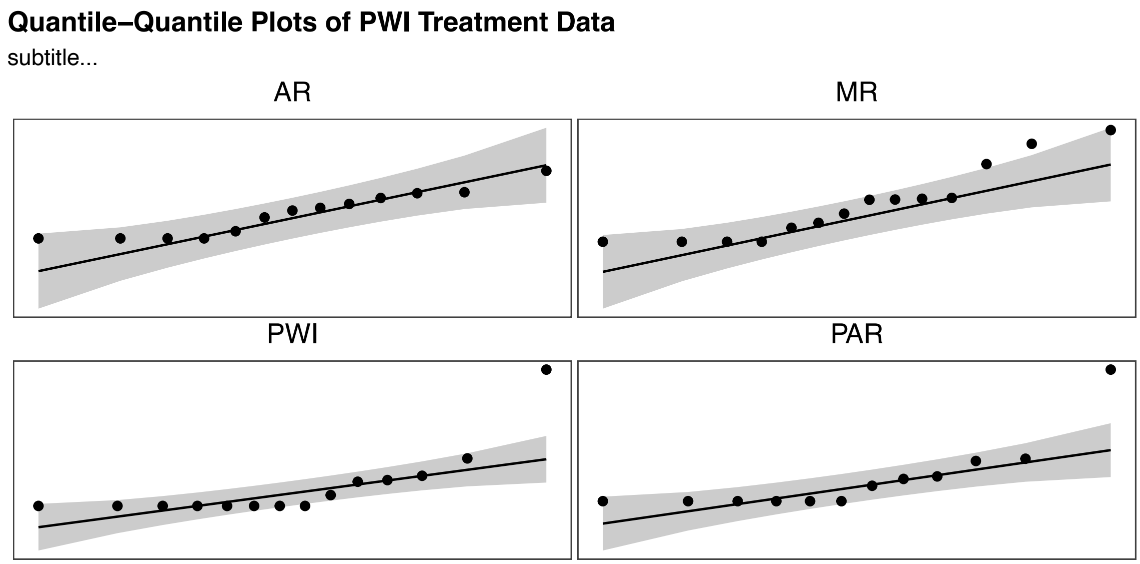


Figure 5.26: Q-Q Plots of PWI Reliance by Treatment

Outliers were identified within each treatment group for all measures of task uncertainty. Data for five participants with Z-scores of 2.5 or higher are summarized in Table 5.17. These participants were removed from the primary analysis to ensure the data represents typical participant behavior and to maintain the integrity of the comparative analysis.

Table 5.17: Outliers in Measures of Reliance

|Z-scores| for Reliance Metrics of Participant Outliers

Participants with |Z-score| >= 2.5 in reliance components

Participant	Treatment	Count	Time	Reliance
1021	PAR	2.74	1.67	0.51
1028	PAR	0.87	2.50	2.87
1031	PWI	0.89	2.98	3.16
1045	PWI	3.08	1.43	0.28

1053	AR	3.11	2.86	0.27

Several checks were then repeated to assess the impact of outlier removal. The Shapiro-Wilk test showed no change in significance for data overall or by treatment group. While Levene's test found significant difference in variance between groups (p = 0.01) after the change, Kruskall-Wallis does not assume otherwise. A Wilcoxon signed-rank test was applied to compare overall and group means with and without outliers. No significant differences were found in average PWI time per reference in the the full data set or by treatment group (p > 0.40 for all cases). It is determined that, despite a reduction in the overall mean value from 1.08 to 0.87, the character of the underlying data is not significantly changed by the outlier removal. Given that most treatments are not normally distributed, and variance differs between groups, a non-parametric approach is chosen for testing H_{2b}. The total number of observations (n = 48) and group sizes (12 PWI, 10 PAR, 12 AR, and 14 MR) remain sufficient for these methods, though statistical power is reduced.

Analysis

Per Figure 5.27, despite a nominal p-value (0.05) the Kruskal-Wallis test fails to identify significant between-group differences for the average PWI reference duration. A small effect size $\hat{\epsilon}^2_{rank} = 0.17$ with broad $CI_{95\%}$ $[0.07, 1.0]$, along with a post-hoc power simulation (80.3%) all support this result.

Given the marginal p-value and high variance data, bootstrap analysis was again performed to provide complementary insights. As before, five replications of 10,000 resamples were conducted for all pairs using mean, median, and the Hodges-Lehmann estimator. Significant results are summarized in Table 5.18. As with the Kurskal-Wallis test, no significant differences were identified using median, but the mean and Hodges-Lehmann estimator both identified 95% CI based differences in PWI-AR, PWI-MR, and PAR-AR pairs. A significant difference in mean was also indicated for PAR-MR.

Table 5.18: H_{2b} Bootstrap Analysis of Pairwise PWI Differences

Significant Reliance Differences in Centrality: Bootstrapped Estimates
Five Replications of 10,000 Resamples for Mean, Median, and the Hodges-Lehmann Estimator

Comparison	Stat	Mean	95% CI	Bias	Std Error
PWI - AR	Mean	−0.993	[−1.776, −0.223]	0.000 (0.000)	0.397 (0.000)

PWI - MR	Mean	-0.880	$[-1.622, -0.196]$	-0.007 (0.000)	0.362 (0.000)
PAR - AR	Mean	-0.954	$[-1.748, -0.170]$	-0.002 (0.000)	0.401 (0.000)
PAR - MR	Mean	-0.841	$[-1.577, -0.171]$	-0.008 (0.000)	0.358 (0.000)
PWI - AR	HL_Est	-1.064	$[-1.933, -0.105]$	-0.013 (0.000)	0.477 (0.000)
PWI - MR	HL_Est	-0.956	$[-1.816, -0.117]$	0.042 (0.000)	0.445 (0.000)
PAR - AR	HL_Est	-0.847	$[-1.933, -0.005]$	-0.166 (0.000)	0.499 (0.000)

Significance indicated by confidence intervals that do not include zero.

On average, AR and MR reduce reliance by 0.686 and 0.612 seconds per reference, respectively, relative to the PWI instructional method. This finding is statistically and practically meaningful, given that the overall average reference duration is 0.87 seconds. In summary, the bootstrap analysis finds AR and MR are equivalent to one another but reduce reliance more than than PAR and PWI, which are themselves equivalent. The negligible bias values and low standard error reported across all replicates once again reinforce the reliability of these findings.

H_{2b} Result

The marginal p-value returned by the Kruskall-Wallis comparison of median reliance, together with consistent and robust estimates of the differences provided by bootstrap analysis give sufficient evidence to accept H_{2b}: reliance does vary with instructional method. However, the conflicting results of pairwise comparisons must be interpreted with caution. Like other results in this study, they suggest that AR/MR are superior to PWI/PAR but the limited data and high variation make this somewhat inconclusive. Ultimately, it is determined that statistically significant differences in reliance exist between treatments, but specific pair-wise differences are not clear enough to declare.

5.5 H_3: Retention Phase Analysis

The final phase of this study is designed to test the durability of the training received. Twenty-four participants volunteered to return to the lab on April 29th to complete the assembly task one last time, without further instruction. All participants were asked to complete the task correctly, quickly, accurately, and entirely from memory. Though a

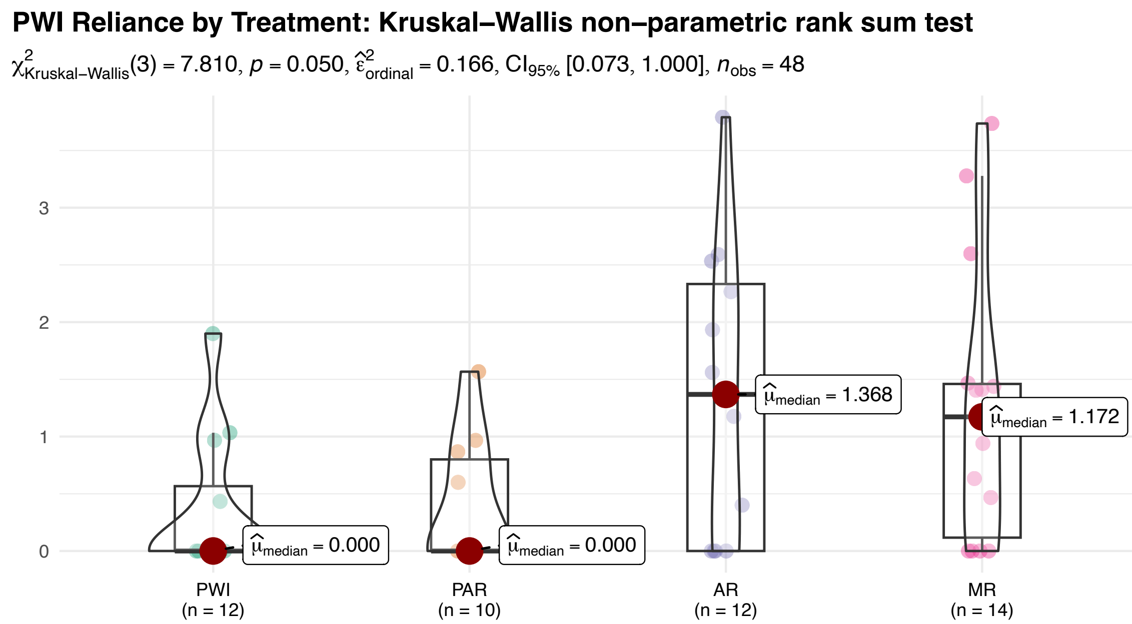


Figure 5.27: H_{2b} Test - Average Time per PWI Reference by Treatment

target time of 60 seconds was established, each participant was given the time they needed to declare their effort complete, with a hard stop at 3 minutes[2].

5.5.1 H_3: Change in TCT and UCE since Recall

The one hypothesis tested during the retention phase:

$$H_3\text{: Learning retention varies with treatment}$$

is designed to determine if the instructional methods have an impact on the durability of training for up to two months.

When originally planned, this analysis would utilize the change in OEE from recall to retention. It was later realized that the way in which OEE assesses task quality as pass / fail based on a single defect is problematic. Each participant only performed one replication of the retention task, where errors are expected. However, minor errors may not indicate a complete lack of retention. The combination of these factors leads to unstable binary outcomes that make it impractical to draw reliable conclusions about retention when using OEE. Instead, retention is assessed by studying the relationships between treatment and gap (the

[2] This limit was never reached.

Table 5.19: Summary Statistics for Dependent Variables, H3 Retention

	Mean	SD	Min	Median	Max
Task Completion Time (s)	101.88	28.10	55.00	106.50	154.00
Uncorrected Errors (n)	6.38	4.45	0.00	7.50	14.00
Increase in TCT (s)	33.25	27.91	−12.65	30.67	107.22
Increase in UCE (n)	4.60	4.78	0.00	3.75	14.00

number of days since the original training), and the increase in TCT and UCE during that time. These are treated as H_{3a} and H_{3b} for the increases in TCT and UCE, respectively.

Descriptive Statistics

Each of the 24 participants completed one task, for which the two measures were observed. The four treatments were relatively evenly represented (8 PWI, 5 PAR, 6 AR, and 5 MR) and the gap between recall and retention ranged from 3 to 44 days (mean 21.58, SD 13.06, median 20.5). Table 5.19 summarizes the descriptive statistics for TCT, UCE, and their increases.

A variety of measures were considered to establish the baseline of performance during recall, including the mean, weighted average, 75th percentile, average of last two tasks, and last value for TCT and UCE. Ultimately, it was decided that an average of TCT / UCE over the last two task results was the most suitable measure of "typical" performance. Given the limited number of replications during recall, this approach balances recency with stability, reflecting the most recent performance trends while limiting the influence of a single result.

As seen in Figure 5.28, the median increases in TCT appear similar across all treatments, with PWI and AR slightly higher than PAR and MR. Considerable variation exists within each treatment group, with MR exhibiting the most consistent values, while PWI and AR have greater variability and potential outliers. The increase in UCE is more clearly differentiated by treatment, with median steadily decreasing from AR to PAR, PWI, and MR. Variance is similar in the PWI-PAR and AR-MR pairs, with PWI and PAR both exhibiting significantly less than AR and MR. The overlapping IQR of increase in TCT suggests a lack of significant treatment effect. Overlap is less prominent in UCE increase, giving some evidence of treatment effect. Note that positive increases were consistent for both TCT and UCE. The lone exception is the increase in UCE for MR, with a median of zero.

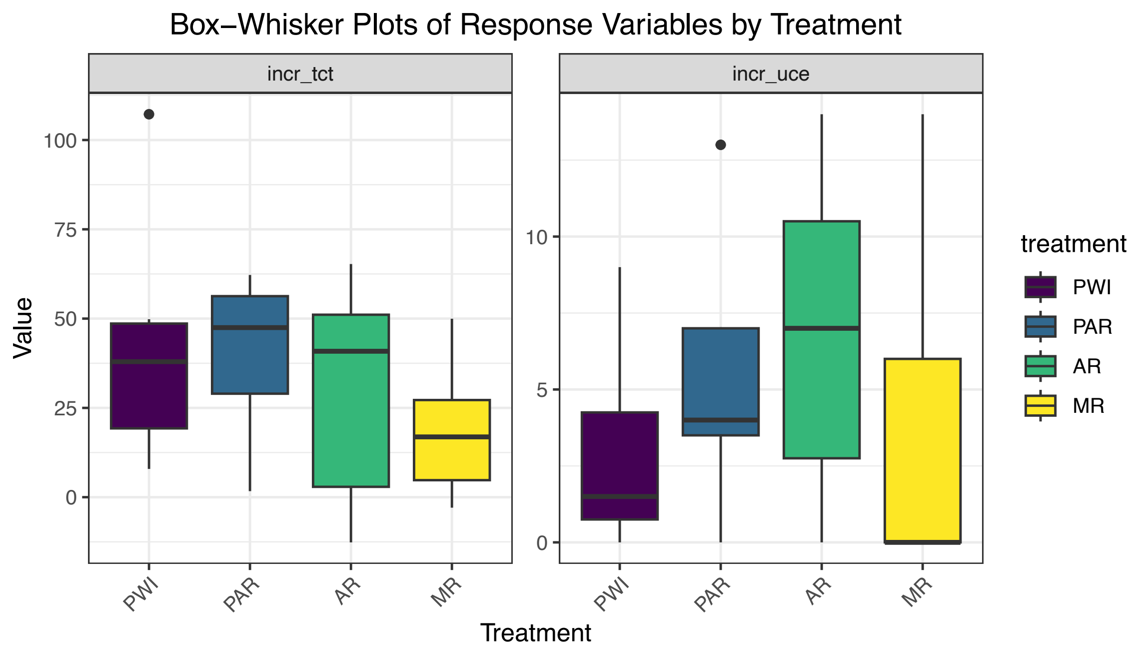


Figure 5.28: Variation of Response Variables by Treatment

Check Model Assumptions

The observed values for the increase in TCT (iTCT) and UCE (iUCE) are all independent. Each response will be analyzed separately, as a function of treatment and gap. A simple comparison of means like Kruskal-Wallis is insufficient for two predictors, and model-based methods are required. When testing continuous variables like iTCT, a linear regression model (LM) of the form `iTCT ~ treatment +/* gap` is appropriate. iUCE, on the other hand, is a discrete variable (error count) with insufficient sample size to approximate a normal distribution. A generalized linear model (GLM) with similar form and Poisson or Negative Binomial distribution is commonly used in this situation.

The assumptions of a GLM, including linearity, mean-variance relationship, and zero-inflation, are assessed after fitting the model. Meanwhile, to confirm the iTCT data meets the normal requirements of a LM, normality, homoscedasticity, and linearity must be checked. A Shapiro-Wilk test suggests the data is normal overall (p = 0.30) and within treatment groups (p >= 0.109 for all). The Q-Q plots of Figure 5.29 fail to provide strong counter-evidence, validating the normality requirement. Furthermore, the Levene test finds no evidence of heteroscedasticity (p = 0.77).

To test the linearity assumption, a graph of iTCT vs gap was constructed for each treatment. The scatter plots in Figure 5.30 show some support for linearity, though the limited data

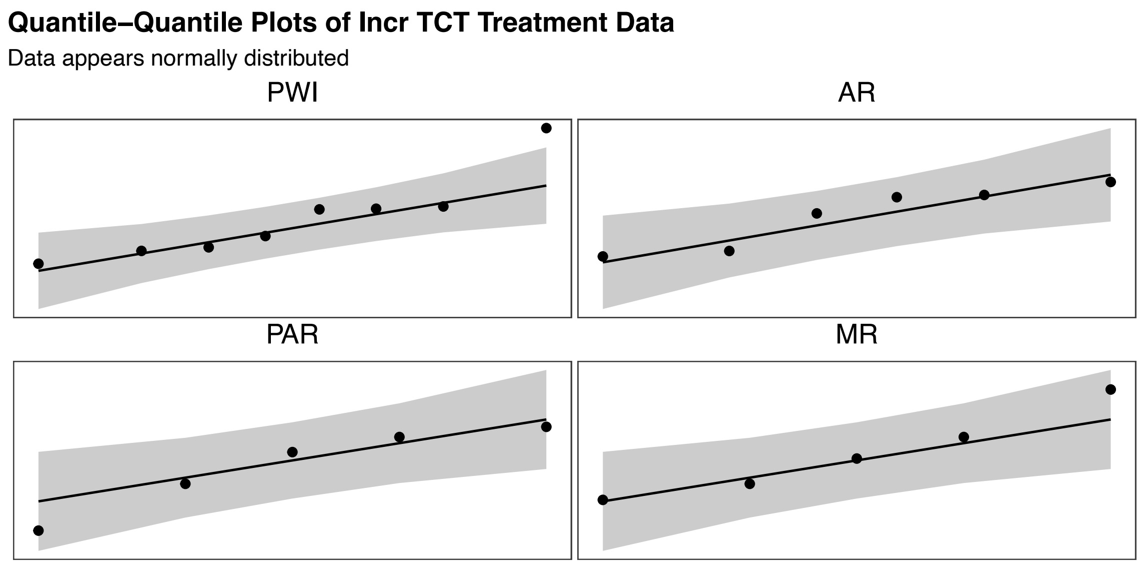


Figure 5.29: Q-Q Plots of Increase in TCT by Treatment

makes it difficult to judge. Additional evidence may come from the model residual plots. Until then, this is sufficient. In summary, the model assumptions are met: iTCT is continuous and linear, with independent observations, and all treatment groups are normal, with homogeneous variance.

As a final step prior to analysis, both iTCT and iUCE are tested for outliers using a combination of Z-score, IQR, and multivariate (Mahalanobis Distance) methods. None were detected by any test.

H_{3a}: Increase in TCT vs Gap and Treatment

To begin assessing the effect of instructional method on retention, the relationship between iUCE and gap, visualized in Figure 5.31, is considered. A pair of linear regression models (LMs) were fit, both using iTCT as the response, with gap and treatment as predictors. One specified additive terms and the other included interactions between gap and treatment.

A comparison of model performance statistics are summarized in Table 5.20. Columns for the model name and type are followed by AIC, BIC, and their corresponding weights, standard and adjusted R-squared (R^2), and the Root Mean Squared Error (RMSE). AIC/BIC weights represent the relative likelihood of each model being the most accurate of those considered.

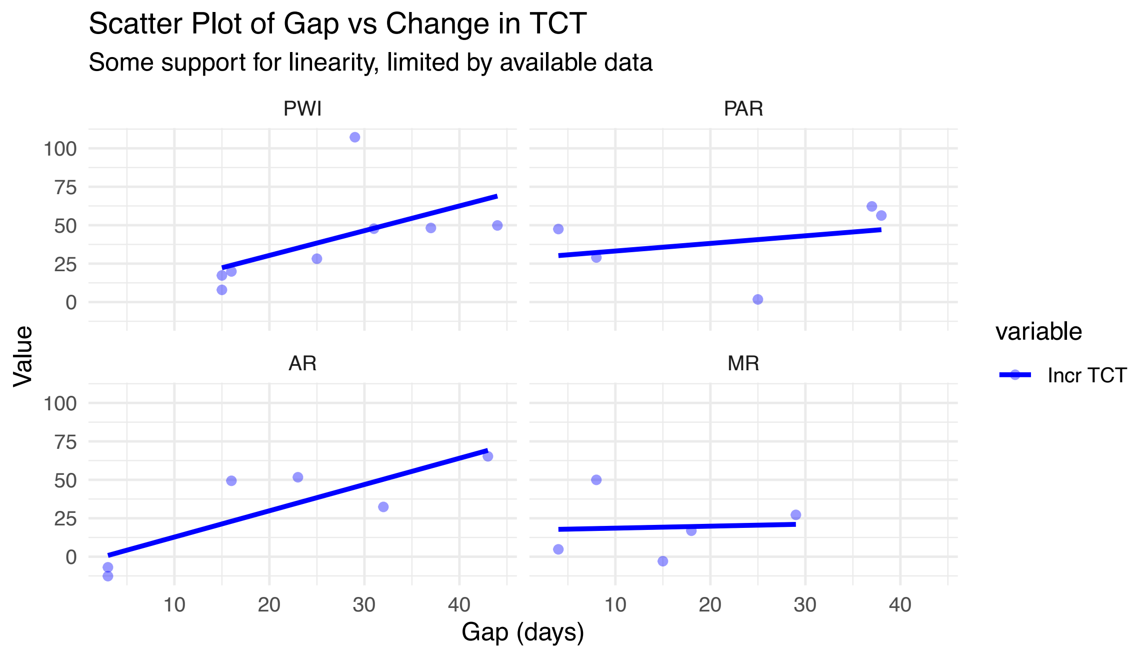


Figure 5.30: Average Gap vs Increase in TCT by Treatment

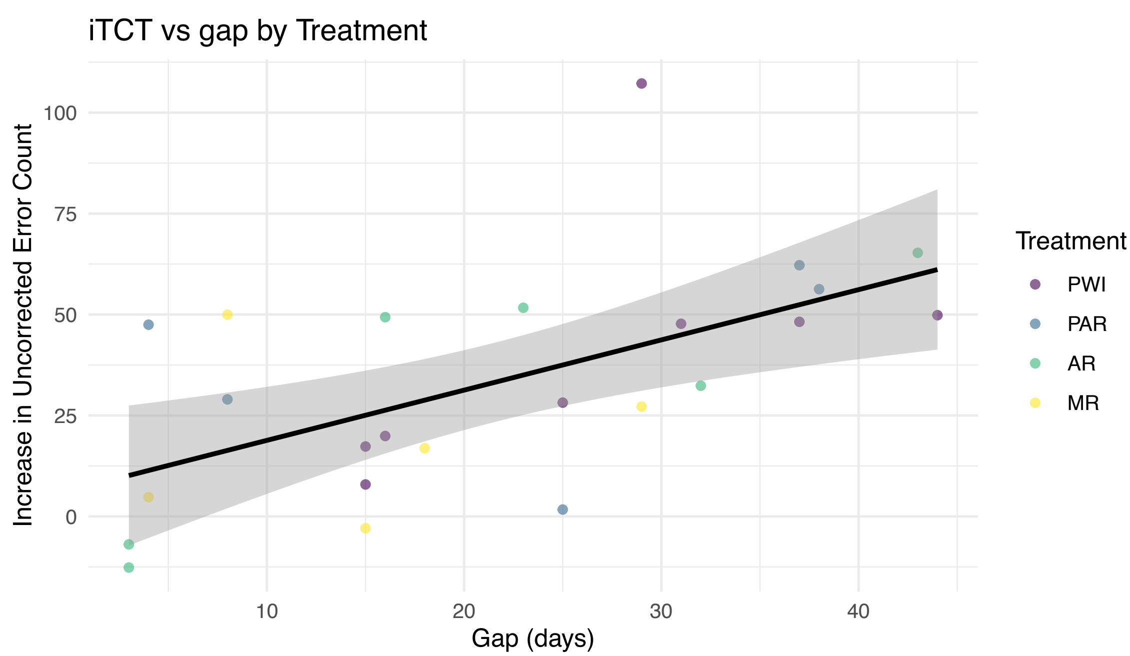


Figure 5.31: Increase in Task Completion Time vs Gap by Treatment

Table 5.20: Model Performance Results

Name	Model	AIC (weights)	BIC (weights)	R2	R2 (adj.)	RMSE
m1_tct_add	lm	228.3 (0.81)	235.3 (0.96)	0.36	0.22	21.90
m2_tct_int	lm	231.2 (0.19)	241.8 (0.04)	0.44	0.19	20.53

In this comparison, both AIC and BIC strongly prefer the additive model. While the R^2 of the model with interactions is higher, the additive model scores a higher R^2_{adj}., where complexity is penalized. Despite a slightly higher RMSE, the additive model outperforms based on the principle of parsimony, which favors the least complex model that effectively explains the data. Details for the additive model are summarized in Table 5.21. The intercept represents the mean iTCT for the reference treatment (AR), with a gap of zero. The treatment coefficients (MR, PAR, PWI) indicate the difference in mean iTCT between each treatment and AR, assuming the same gap value. Since there are no interaction terms in the model, the effect of gap on iTCT is constant across all treatments.

Table 5.21: Increase in TCT LM

Table 5.21: Fixed Effects

Parameter	Est Coef	SE	95% CI	t(19)	p	Sig
(Intercept)	6.587	13.059	(-20.75, 33.92)	0.504	0.6198	
gap	1.163	0.417	(0.29, 2.04)	2.788	0.0117	*
treatmentMR	-4.625	15.059	(-36.14, 26.89)	-0.307	0.7621	
treatmentPAR	6.692	14.935	(-24.57, 37.95)	0.448	0.6592	
treatmentPWI	3.377	13.564	(-25.01, 31.77)	0.249	0.8061	

Only the gap predictor is statistically significant, with an estimated coefficient of 1.163 (p = 0.0117), indicating that iTCT increases with gap at a rate of approximately 1.16 seconds per day, which is practically significant. The coefficients for the treatment levels are not statistically significant, which implies that the observed differences in iTCT between the treatments may not be robust or replicable. This result must be interpreted with care, as the overall model fit is marginal. Gap and treatment together account for only about 36% of the variance in iTCT (R^2 = 0.358 and adjusted $R^2 = 0.222$). The p-value for the F-statistic of 2.645 on 4 and 19 degrees of freedom is 0.0656. At the conventional threshold of 0.05, this fails to reject the null hypothesis that all regression coefficients except the intercept are zero. Together, these results provide only limited evidence that the predictors collectively contribute to explaining the variability in iTCT, with gap making the only significant contribution. Further validation with a larger sample size and additional diagnostics would be prudent to confirm these findings and ensure robustness.

To validate this model a number of checks were performed. A Shapiro-Wilk test indicates residuals are normally distributed (p = 0.08). The Q-Q plot of the residuals in Figure 5.32 shows most points are on or near the reference line, despite some deviance in the right tail. The second plot in the figure shows the residuals are scattered randomly around the axis without clear curvature or other systematic deviations. This provides good support for the LM's linearity assumption within the limits of the available data. The variance inflation values for gap (1.06) and treatment (1.02) are both near 1, indicating very low multicollinearity. These results confirm the model is valid.

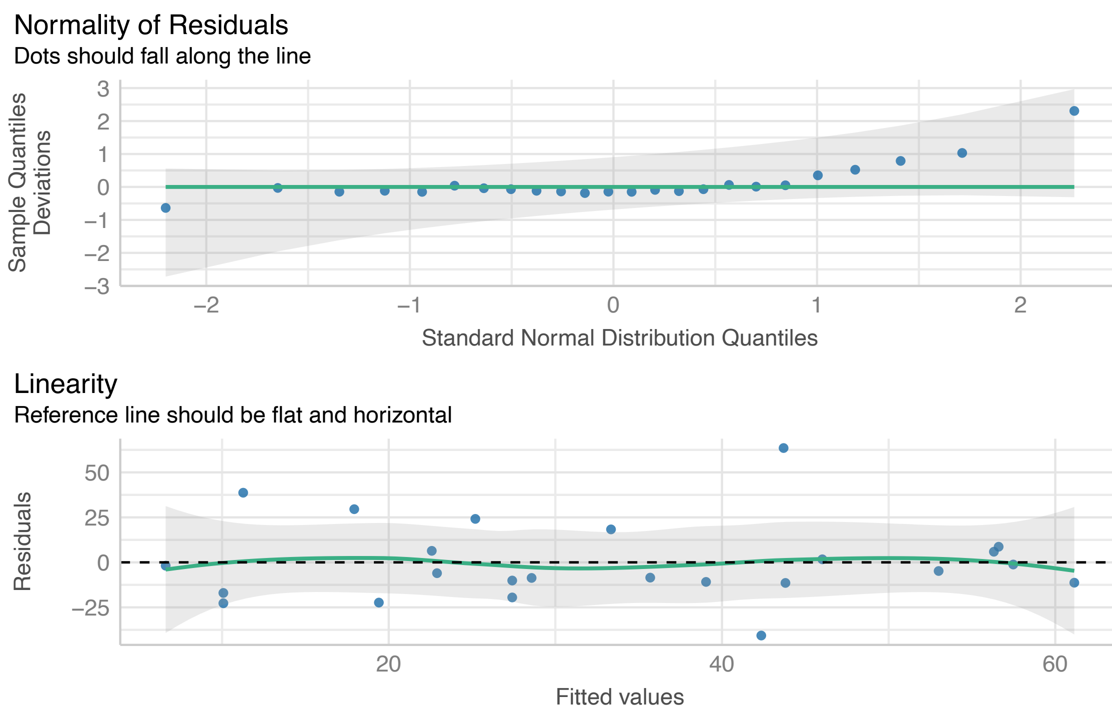


Figure 5.32: Model validation for H3a

H_{3b}: Gap and Treatment vs Incr UCE

Continuing the analysis of H_3, the second measure of learning retention, iUCE, is tested. A pair of generalized linear models were fit using a Poisson distribution. As before, gap and treatment were the predictors, and iUCE the response. One model includes only the main effects of the predictors, while the other also include their interactions. The relationship between iUCE and gap is visualized in Figure 5.33, which includes a linear regression with confidence intervals for the full dataset.

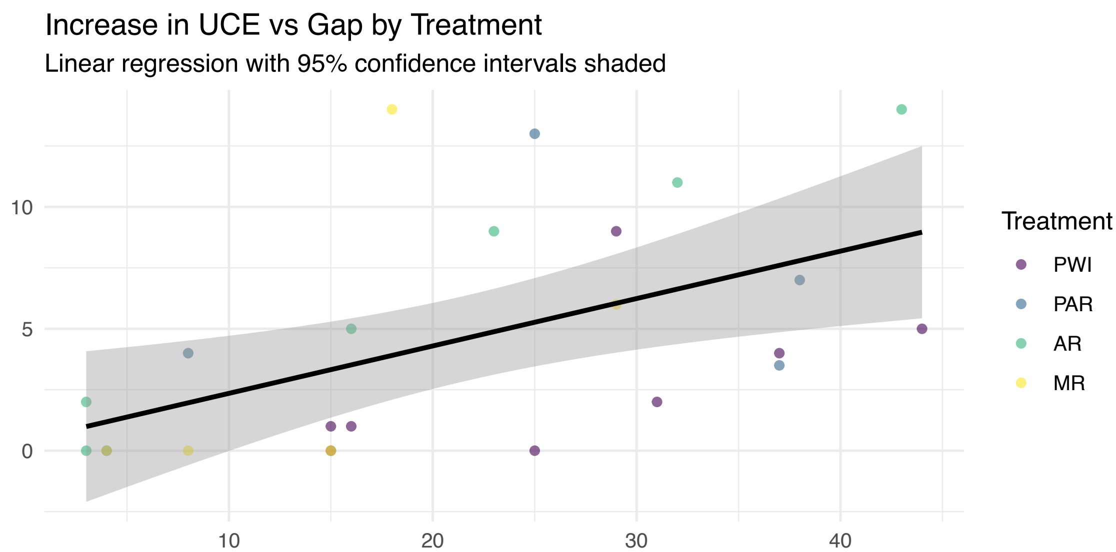


Figure 5.33: Increase in Uncounted Errors vs Gap by Treatment

Because the Poisson distribution is used to model count data, it requires all non-negative integer values for the response variable. Since iUCE was measured from the average of the last two values during recall, some non-integer observations are included. To correct for this without altering the relative differences between observations, iUCE was doubled prior to the model fit. When interpreting the model results, the coefficients and other estimates will be on the scale of the doubled iUCE and should be halved to interpret them on the original scale.

Table 5.22: Model Performance Results

Name	Model	AIC (weights)	BIC (weights)	Nagelkerke's R2	RMSE
m1_uce_poisson	glm	226.6 (0.21)	232.5 (0.61)	0.99	7.05
m2_uce_poisson	glm	223.9 (0.79)	233.3 (0.39)	0.99	6.92

A comparison of model performance statistics are summarized in Table 5.22. The columns for AIC, BIC, and RMSE are as previously described. The standard and adjusted R^2 terms are replaced with Nagelkerke's R^2. This is the equivalent measure for GLMs with count data outcomes, and is interpreted in the same manner. The second model includes interactions of gap and treatment. It performs better on AIC, BIC, and RMSE metrics, and has the same R^2 value, indicating that the additional model complexity provides a substantially better fit.

Despite these results, post-hoc validation steps reveal conflicts with the model assumptions. Dispersion is measured at 9.04 ($p < 0.001$), which is both substantially larger than 1.0 and statistically significant. This indicates overdispersion, a condition where the observed variance is greater than that predicted by the model. A second test found that zero inflation is present ($p = 0.048$), indicating that the count of zero values in the data is greater than the model expects. Both violate the assumptions of the Poisson-based model, and prompt the switch to the Negative Binomial (NB) distribution, which accounts for overdispersion. Based on these findings, Zero-Inflated Negative Binomal (ZINB) equivalents are included for comparison.

Table 5.23: Model Performance Results

Name	Model	AIC (weights)	BIC (weights)	RMSE
m1_uce_nb_add	negbin	155.0 (0.03)	162.1 (0.09)	10.06
m2_uce_nb_int	negbin	157.7 (8.00e-03)	168.3 (4.00e-03)	19.09
m3_uce_zinb_add	zeroinfl	149.6 (0.44)	157.9 (0.75)	7.54
m4_uce_zinb_int	zeroinfl	149.3 (0.52)	161.1 (0.15)	16.80

As seen in Table 5.23, both ZINB options (model 3 and 4) strongly outperform the NB alternatives in terms of AIC and BIC. While model 4 (with predictor interactions) has a slightly lower AIC, the main effects version (model 3) shows substantially better BIC and RMSE. Based on this, the additive ZINB model, which provides the best balance of model complexity, fit, and predictive power, is selected.

Table 5.24: ZINB Model Summary

Table 5.24: Fixed Effects

Parameter	Est Coef	SE	95% CI	z	p	Sig	Adj Coef
(Intercept)	6.709	2.875	(2.90, 15.54)	4.441	< 0.001	***	3.354
gap	1.033	0.014	(1.01, 1.06)	2.414	0.0158	*	0.516
treatmentMR	1.480	0.646	(0.63, 3.48)	0.899	0.3689		0.740
treatmentPAR	0.865	0.309	(0.43, 1.74)	-0.404	0.6859		0.433
treatmentPWI	0.394	0.136	(0.20, 0.78)	-2.693	0.0071	**	0.197

Details for the main effects of the ZINB model are summarized in Table 5.24. The parameters for this model must be interpreted differently than in the previous linear model. In a GLM, the response is log-transformed, resulting in exponentiated coefficients for predictors. As a consequence, the effects are *multiplicative* rather than additive. The intercept represents the expected log count of the response variable when all predictors are at their reference levels. The predictor coefficients then represent the multiplicative effects on the expected count compared to the baseline.

To ease interpretation, the values for estimated coefficients and standard errors are exponentiated in this table, revering the log transform applied by the NB model fit. This adjustment puts the coefficients in the original 2x scale that was used to ensure integer values for the response. Additionally, the "Adj Coef" column reverses the 2 x iUCE transform, allowing interpretation in the observed scale. Importantly, these transformations do not alter the multiplicative nature of GLM results.

Here, the intercept ($p < 0.001$), gap ($p = 0.02$), and PWI treatment effect ($p = 0.007$) are all statistically significant. The adjusted intercept value indicates that baseline performance for the reference treatment level (AR) and a gap of zero is 3.354 errors. The adjusted gap coefficient indicates that the expected error count increases by 0.516 per day for the AR treatment. The effect of other treatments is multiplicative. This implies that the gap coefficient for PWI is changed by a *factor* of 0.197. These relationships can be interpreted as follows:

$$\text{iUCE}_{\text{AR}} = 3.354 + (0.516 \times \text{gap}) \tag{5.7}$$

$$\begin{aligned} \text{iUCE}_{\text{PWI}} &= 3.354 + (0.516 \times \text{gap}) \times 0.197 \\ &= 3.354 + (0.102 \times \text{gap}) \end{aligned} \tag{5.8}$$

PAR and MR are statistically equivalent to AR, despite estimated differences that are meaningful if real. To assess the overall model fit, it is compared with a null (intercept only) version of the model using the likelihood-ratio statistic and a χ^2 distribution. The result ($p = 0.03$) confirms that gap and treatment effects have significant explanatory power. Nagelkerke's R^2 suggests that the model explains about 36.67% of iUCE's observed variance. Additionally, the model output confirms that both dispersion (θ, $p = 0.003$) and zero-inflation ($p = 0.05$) were present and properly accounted for, validating the model selection process. Finally, Figure 5.34 shows that, despite a few notable high points, the residuals lack sequential patterns, clustering, or other systematic trends. This randomness provides further evidence that the model is adequately capturing most of the structure in the data. Together, these findings give confidence in the model results.

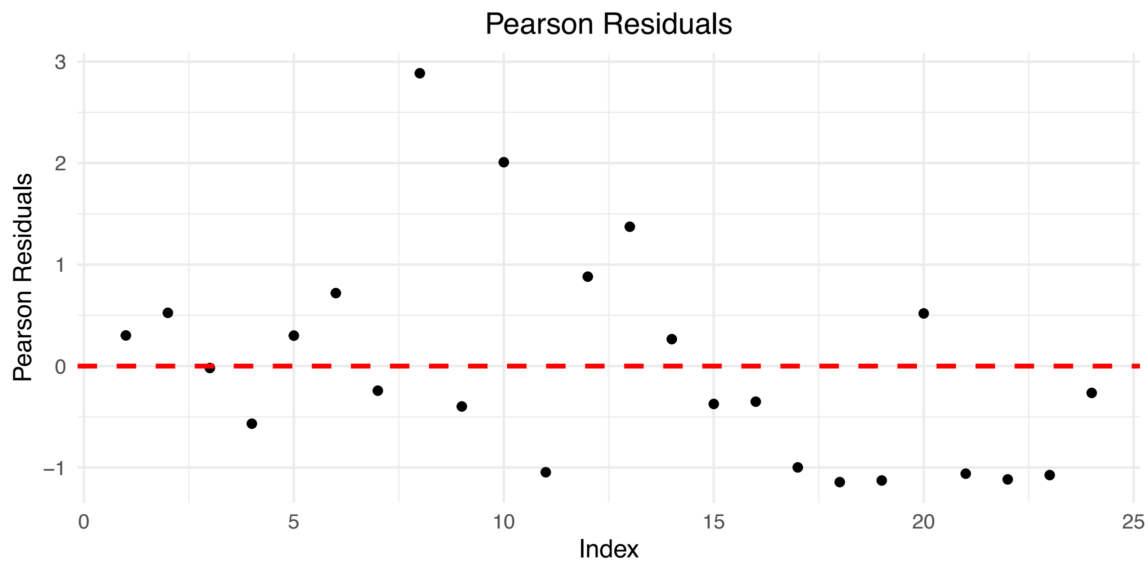


Figure 5.34: Pearson Residuals

5.5.2 H_3 Result

No statistically significant evidence was found for the effect of instructional method on the increase in TCT over the retention interval. Treatment coefficients for the linear model suggested meaningful differences exist, but that could not be substantiated with the limited data available. The overall model fit was marginal and was not improved by including interaction effects for the gap and treatment predictors. Therefore, H_{3a} is rejected for a lack of evidence in support of the claim that learning retention varies with treatment.

A significant treatment effect on the increase in UCE over time was identified using a zero-inflated negative binomial GLM. Specifically, the PWI method showed a significantly lower rate of error increase compared to the reference AR method. As with the TCT analysis, other treatment coefficients suggest that a meaningful effect may exist, but did not reach the traditional threshold of 95% confidence. Despite these limited findings, the model was statistically significant overall. Sufficient evidence does exist to validate H_{3b}'s broad claim that the change in error rate over time varies by treatment.

Overall, the implications for H_3 are mixed, with the UCE model providing stronger evidence for an effect of instructional method on retention than the TCT model. This difference suggests that time-based performance improvements may not be as sensitive to instructional methods as error rates. While it has little impact on this study, identifying gap is a significant predictor for both iTCT and iUCE is logical and expected, as retention naturally declines over time.

Given the limitations of the study (small sample size, single retention test, variable gap), it would be prudent to interpret these findings as preliminary evidence for the hypothesis, rather than a definitive confirmation. To strengthen these conclusions, further research with larger samples and more repeated measures would be valuable to increase the power to detect treatment effects and estimate them more precisely.

5.6 Workload, Usability, and User Experience

This section explores the relationships between perceived workload, system usability ratings, overall user experience, and the instructional technology assigned. It is accomplished by a thorough analysis of data collected from the TLX and SUS instruments, along with a synthesis of qualitative feedback recorded during each session. Together they provide a rich expression of each participant's experience.

This analysis focuses on two important aspects of the TLX and SUS data: the results from the learning phase, where instructional technologies were actively used, and how those measures differ from the those observed during recall. This approach provides insight into the perceived workload and usability of each treatment both during use and, by contrast, in its absence. Changes in workload and usability between phases will be evaluated in the context of the experimental design, which suggests any improvement[3] must be attributable to some combination of the following factors:

1. Shift in the priorities and objectives of the recall phase. Except for participant idiosyncrasies, this effect is assumed consistent for all treatments.
2. The elimination of training support during recall, which requires participants to work from memory with limited access to instructions. This factor may vary if some treatments do more to engender reliance than support learning.
3. Learning outcomes during the first phase. Some treatments may lead to a better combination of proficiency and confidence than others.

[3] Care must be taken in evaluating "improvement" for these measures. Where an increase in usability (SUS) is seen as an improvement, the opposite is true for measures of workload (TLX), where decreases are desirable.

Table 5.25: Summary Statistics for TLX Workload Measures

Phase	Measure	Mean	SD	Min	Median	Max
1 - Learning	Source Weight	2.50	1.64	0.00	3.00	5.00
	Source Rating	48.87	26.05	0.00	50.00	100.00
	Source Workload	144.74	125.86	0.00	120.00	500.00
	Task Workload	57.90	13.73	15.67	57.67	87.67
2 - Recall	Source Weight	2.50	1.68	0.00	2.50	5.00
	Source Rating	35.77	25.17	0.00	30.00	100.00
	Source Workload	104.55	112.25	0.00	67.50	500.00
	Task Workload	41.82	14.69	11.00	38.50	82.00

The second and third factors both relate to the instructional treatment, though both express different dimensions of effectiveness. All other known and unknown confounders are controlled for by the study design.

5.6.1 NASA Task Load Index Analysis

All observed and calculated measures from the NASA TLX are summarized by phase in Table 5.25. As detailed in Section 4.4.1, source weight, rating, and workload refer to the ranking, value, and weighted score for all workload sources, while task workload is the composite measure of all contributing sources. These data reveal a consistent reduction in perceived workload as participants transitioned to the recall phase. This is evidenced by decreases in overall Task Workload (Learning: M = 57.90, SD = 13.73; Recall: M = 41.82, SD = 14.69) and individual Source Ratings and Workloads. Notably, the Source Weight remained constant (M = 2.50) across phases, suggesting that while the perceived importance of workload sources didn't change, their intensity diminished.

Despite this overall trend, the data exhibits considerable individual variability, as shown by high standard deviations and wide ranges, particularly in Source Workload (0 to 500 in both phases). The consistently lower median values compared to means indicate positively skewed distributions, suggesting that while most participants experienced reduced workload during recall, a subset continued to find certain aspects highly demanding.

These findings provide a basis for understanding how different instructional methods may have influenced workload perceptions across the experimental conditions. This section will explore each of the following aspects of that influence workload during learning:

- Q1: Which sources are most important to participants' perception of workload, regardless of treatment?
- Q2: How does the perceived impact of each source differ across treatment groups?
- Q3: For each treatment, which source(s) had the greatest overall influence on perceived workload?
- Q4: Is there a reliable relationship between overall workload and the instructional treatment?

Together, the answers will provide a comprehensive understanding of relationships between how participants perceive workload, how they quantify it for each task / treatment, and factors underlying that association. The results will offer insights into the cognitive, physical, and emotional demands imposed by different instructional technologies and their implications for task performance and learning outcomes.

TLX Q1: Participant Ranking of Source Weights

To understand which of the TLX's six sources are most important to all participants' perception of workload, a Friedman test of repeated measures was conducted. This non-parametric test does not assume normality and accounts for the within-participant variability inherent in the repeated measures of survey data. Using `friedman_test` from the `coin` package confirmed significant (122.1, $p < 0.001$) differences in source rankings by participants. The subsequent post-hoc test calculated pairwise differences using a Wilcoxon signed-rank test and Bonferroni correction for multiple comparisons.

The results of `stats::pairwise.wilcox.test`, as tabulated in Table 5.26, show significant differences for most pairs. In particular, Physical Demand (PD), Performance Factor (PF), and Temporal Demand (TD) are significantly different from all other sources, with the exception of TD vs PF and FF (Frustration Factor). Pairs with significant differences indicate substantial disagreement among participants about the ranking of the corresponding sources. Less polarizing pairs have higher p-values, with those over 0.05 of similar importance to participants. Overall, the number of significant and strongly significant findings shows that participants rank sources very differently. The distribution of rank by source is shown in Figure 5.35.

Kendall's coefficient of concordance, a measure of inter-rater agreement, was used to further assess the rankings across the six sources of workload. Results from the `kendall` function of the `irr` package ($W = 0.01$, $p = 1.00$) indicated negligible agreement among

Table 5.26: Pairwise comparisons for TLX Source of Workload Rankings

group1	EF	FF	MD	PD	PF
FF	0.18	NA	NA	NA	NA
MD	1.00	1.00	NA	NA	NA
PD	< 0.001	< 0.001	< 0.001	NA	NA
PF	< 0.001	0.004	< 0.001	< 0.001	NA
TD	< 0.001	0.02	< 0.001	< 0.001	1.00

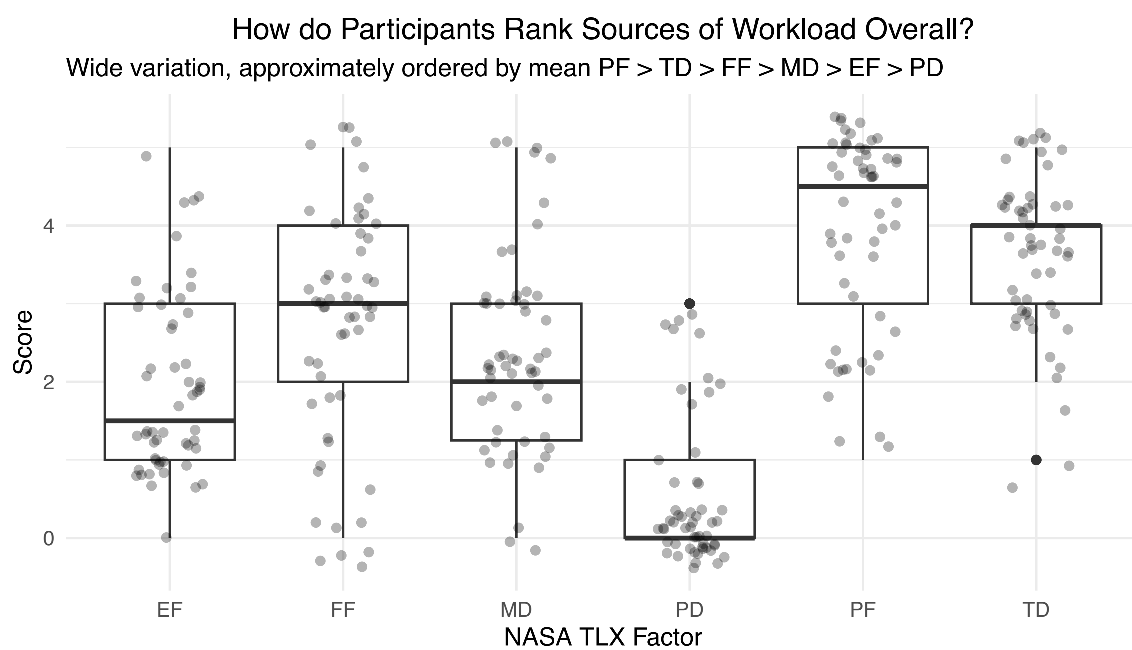


Figure 5.35: Distribution of Factor Scores Across All Treatments

participants, confirming that workload rankings are highly individualized. Low agreement in ranking has lead other researchers to question the utility of the TLX's weighting scheme *(Bustamante and Spain 2008, Byers et al 1989)*. Regardless, it remains the officially recommended methodology *(Heart 2006)*.

TLX Q2: Source Ratings by Treatment

Figure 5.36 was created to understand how the sources of perceived workload vary by treatment group. All combinations of treatment and source of workload are arranged in a heatmap, where each cell is the mean rating for the corresponding source and treatment. Temporal Demand consistently rate as the most important workload source across all treatments, with mean ratings all above 66.1. This suggests that all participants experienced substantial time pressure. Physical Demand is consistently rated as the least important source, with all mean ratings below 25.8. This indicates that the assembly task is not perceived as physically demanding in any group.

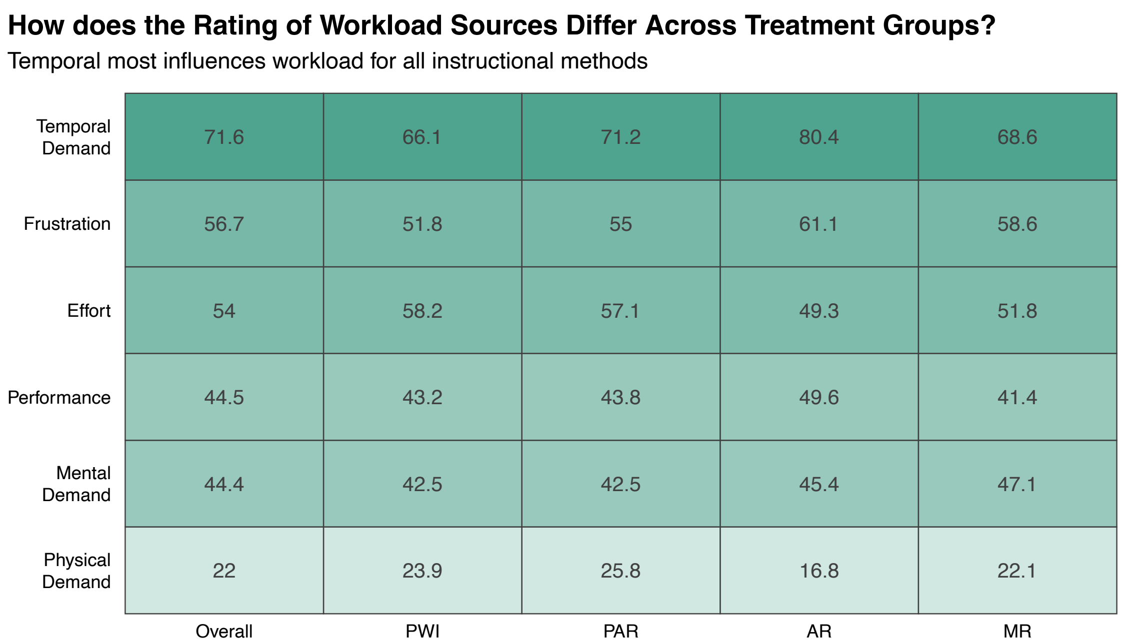


Figure 5.36: Heatmap for Source of Workload Rating by Treatment

There are other subtle but noteworthy differences between treatments. AR has the highest Temporal Demand rating, perhaps due to the groups' lack of familiarity with the system. PAR and PWI have the highest Effort ratings, indicating those participants felt challenged by their work, or less concerned with other factors. Despite these differences, the overall pattern of factor importance remains relatively consistent across treatments, as seen in the uniform vertical gradient of the heatmap.

To better quantify these differences, significance tests were conducted for the difference in rating, by treatment, for all sources. The Kruskal-Wallis rank sum test, which is appropriate for these discrete ordinal values, failed to identify any significant effects, further evidence that the ratings for each source are consistent across treatment groups. There appears to be no meaningful relationship between the task-treatment a participant completes and their perception of the factors most important to their overall workload.

TLX Q3: Treatment Effect on Weighted Source Workloads

Where the previous section explored changes across the treatment axis of fig-tlx-q2-heat for unweighted source ratings, this section tests the differences across sources for weighted source workloads. Figure 5.37 illustrates the relative contributions of each source to the total workload in each treatment. A fairly consistent progression from the highest contributor, Temporal Demand, to the lowest, Physical Demand is obvious. While there is some variation in between, those sources act as the high / low points for all treatments.

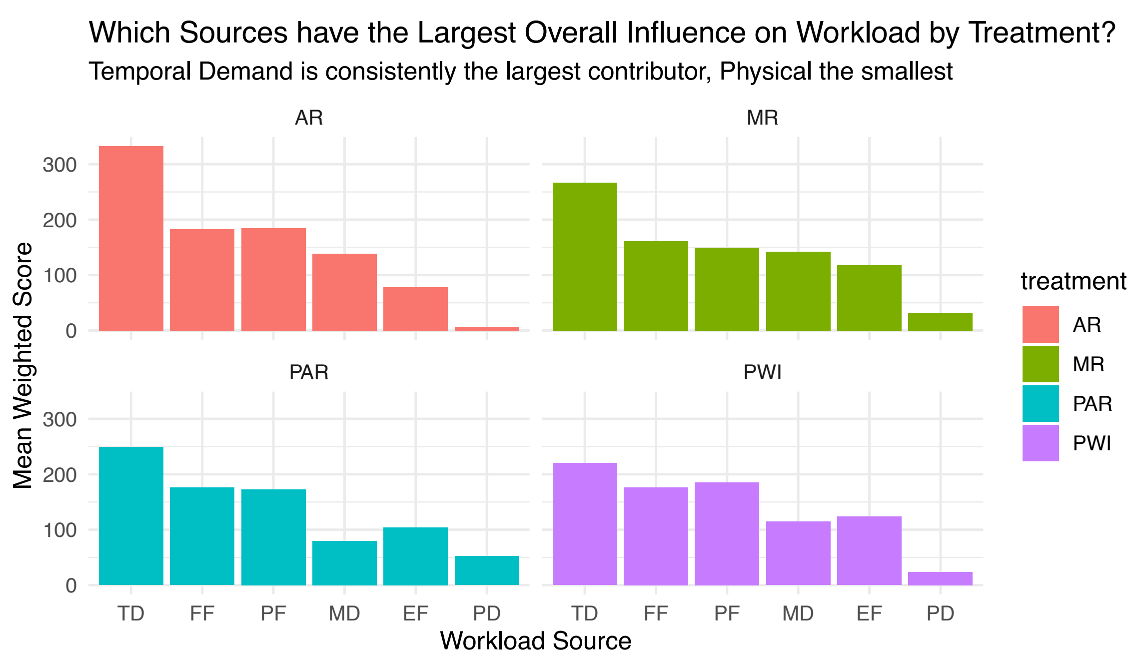


Figure 5.37: Weighted Workload Source Scores by Treatment

Friedman tests show these workload differences are significant for all treatments ($p < 0.001$ in each), confirming that the sources perceived most influential to overall workload vary by treatment. To explore the nature of those differences, Dunn's test for pairwise multiple comparisons of ranked data is applied with Bonferroni correction. Of the 60 pairs tested[4], seventeen significant differences were identified (AR: 6, PWI: 4, PAR: 4, and MR: 3), all involving Temporal Demand, Physical Demand, or both. The four central sources (FF, PF, MD, and EF) are each involved in 3-4 significant pairs.

In summary, while all treatments show significant differences in how source workloads are perceived, only Temporal and Physical Demands emerge as reliably different in all cases. The relative importance of other workload sources varies in magnitude and significance, as Figure 5.37 implies.

TLX Q4: Task Workload by Treatment

Having examined the individual components of workload, the analysis now focuses on overall Task Workload (TWL, the sum of weighted source workloads) and its relationship with the assigned instructional treatment. Of particular interest is whether any treatments have a significant positive or negative influence on perceived workload. Such influences could indicate aspects of user experience that are uncontrolled mediators of task performance. Once again, the Kruskal-Wallis test is used for comparison and the results are shown in Figure 5.38.

No significant differences are identified ($p = 0.53$). Visual inspection shows that median TWL values are quite similar for all treatments and all IQRs overlap. This result is somewhat surprising, given the added complexity and occasional issues associated with the AR and MR interfaces. It may suggest that technical complexity has limited effect on perceived workload, especially in the recruited demographic.

TLX Results

The TLX analysis revealed several important insights into the workload experienced by participants across different instructional technologies. Notably, there was a consistent reduction in perceived workload as participants transitioned from the learning to the recall phase, suggesting that all instructional methods led to some degree of task familiarization. Temporal Demand emerged as the most significant contributor to workload across all treatments, indicating that time pressure was a universal challenge. Conversely, Physical Demand was

[4] Four treatments with six sources each = 4 * (6 * (6 - 1)) / 2 = 4 * 15 = 60

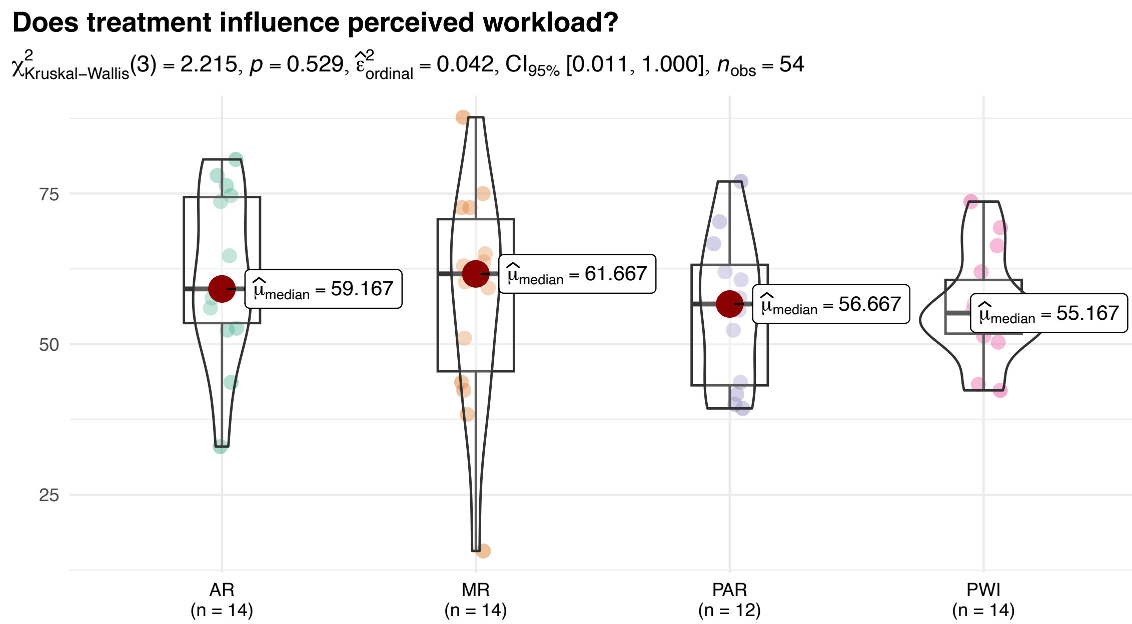


Figure 5.38: Composite Task Workload by Treatment

consistently rated as the least important factor, implying that the assembly task was not physically taxing. Interestingly, despite the varying complexities of the different instructional technologies, no significant differences were found in overall task workload between treatments. This unexpected result warrants further investigation.

Consistent patterns of source ratings and workload across treatments are evident, but accompanied by little statistical significance. This may suggest that the inherent nature of the task, rather than the assigned instructional method, primarily determines which aspects of workload are most important to the participants. The high individual variability across TLX metrics also highlights the importance of considering personal factors in designing and implementing instructional technologies.

5.6.2 System Usability Scale Analysis

The System Usability Score (SUS) provides a complementary view of the participant experience. Where TLX is focused on perceived workload, SUS is designed to provide a simple, one number assessment of overall usability. Despite its simplicity, the SUS is well-respected and often used. Table 5.27 summarizes the raw SUS scores collected during the learning phase of this study, representing the usability of the instructional technology. Five-number summaries are provided for each treatment group for both phases.

Table 5.27: Summary Statistics for Raw System Usability Scores

Treatment		Mean	SD	Min	Median	Max
AR	Score	63.0	15.9	37.5	66.2	90.0
	Rank	42.9	31.3	0.7	44.5	96.1
MR	Score	62.5	20.8	7.5	61.2	87.5
	Rank	43.9	35.5	0.0	29.6	94.1
PAR	Score	68.8	14.1	35.0	72.5	85.0
	Rank	54.8	30.9	0.4	63.8	91.3
PWI	Score	71.8	16.6	37.5	77.5	92.5
	Rank	60.4	35.8	0.7	77.2	97.5

The benchmark distribution for these scores is SUS $\sim \mathcal{N}(\mu = 68, \sigma = 12.5)$. Despite being on a 100-point scale, raw SUS scores should not be interpreted as percentages. For example, a score of 68 is indeed 68% of the maximum, but only represents the 50th percentile. To improve the interpretability of the raw scores, it is considered best practice to convert to percentile rank by normalizing:

$$\text{SUS}_{rank} = \Phi\left(\frac{\text{SUS} - \mu}{\sigma}\right) \times 100 \tag{5.9}$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. Values transformed in this manner can be interpreted in terms of letter grades on the traditional 10-point scale. The benchmark parameters μ and σ were primarily developed during 500 studies of commercial, publicly available products. In research and product development, where SUS percentile ranks and letter grades may seem harsh or unrealistic, they are more commonly used to track improvement over time rather than for absolute comparison.

With that in mind, percentile ranks are also included in Table 5.27. The overall median rank of 56.4% gives these systems a failing grade overall, *relative to commercially available products*. Performance by treatment ranges from 77.2% for PWI to 29.6% for MR, with PAR scoring 63.8% and AR 44.5%. While all of these roughly align with observed interactions, the D letter grade for PAR, a mature commercial product, is noteworthy.

Following best practices, subsequent SUS analysis will utilize the transformed SUS percentile ranks. In addition to improved interpretability, this also accounts for the non-linear nature of the raw scores, enabling more meaningful comparisons. The monotonic nature of this transformation, which preserves the data's original order, will have no affect on the statistical conclusions for the methods employed. Figure 5.39 compares the distribution of SUS across treatments.

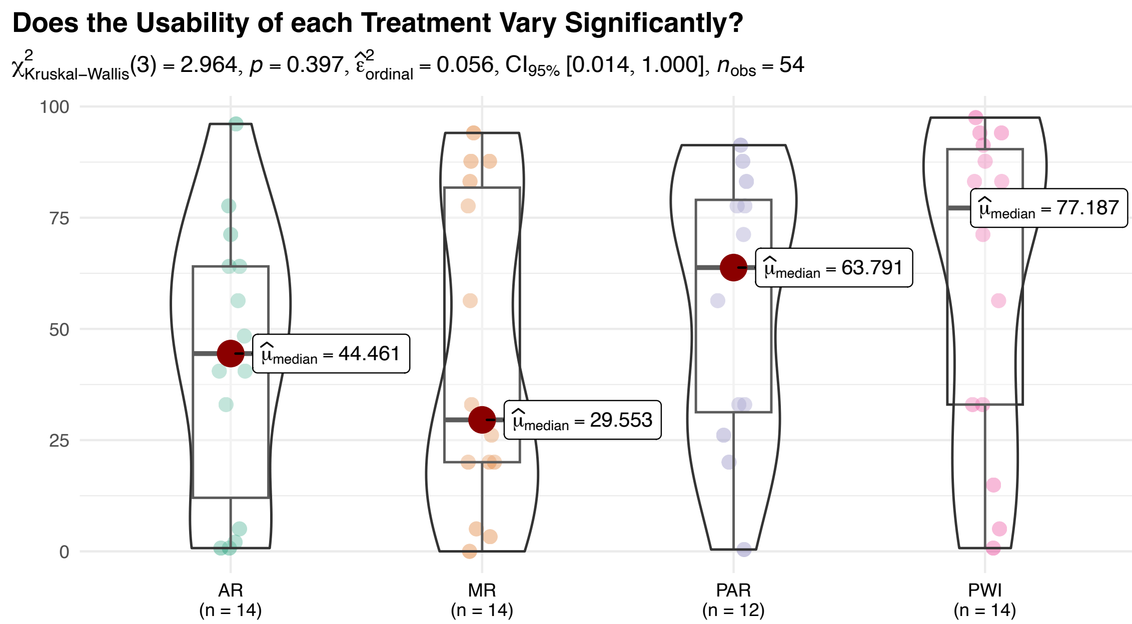


Figure 5.39: System Usability Score for Each Instructional Method

No statistically significant overall difference is identified (p = 0.40), likely due to the very large range of scores found in every treatment. While the effect size is quite small ($\hat{\varepsilon}^2_{rank}$ = 0.056), the progression of median values aligns with expectations based on the complexity associated with each treatment: PWI > PAR > AR > MR. This suggests that, while the technologies differ in complexity, their perceived usability during learning is not drastically different–an unexpected result.

Those learning scores were compared with SUS during recall to test for significance using the Wilcoxon signed rank test for paired samples. As seen in Figure 5.40, the difference is statistically significant (p < 0.001) and practically meaningful. The median value for SUS_R (87.70) is a 31.34 point improvement in usability from the learning phase (SUS_L = 56.36). The large negative effect size ($\hat{r}^2_{biserial}$ = -0.644) with relatively narrow 95% confidence interval [-0.792, -0.424] suggests this finding is robust.

This result implies that usability is increased in recall due to some combination of (1) eliminating the instructional technologies as a source of complexity, and (2) learning effects preparing participants for independent work. To evaluate the relationships between that difference and the treatments, two more tests were conducted. First, the Wilcoxon test is

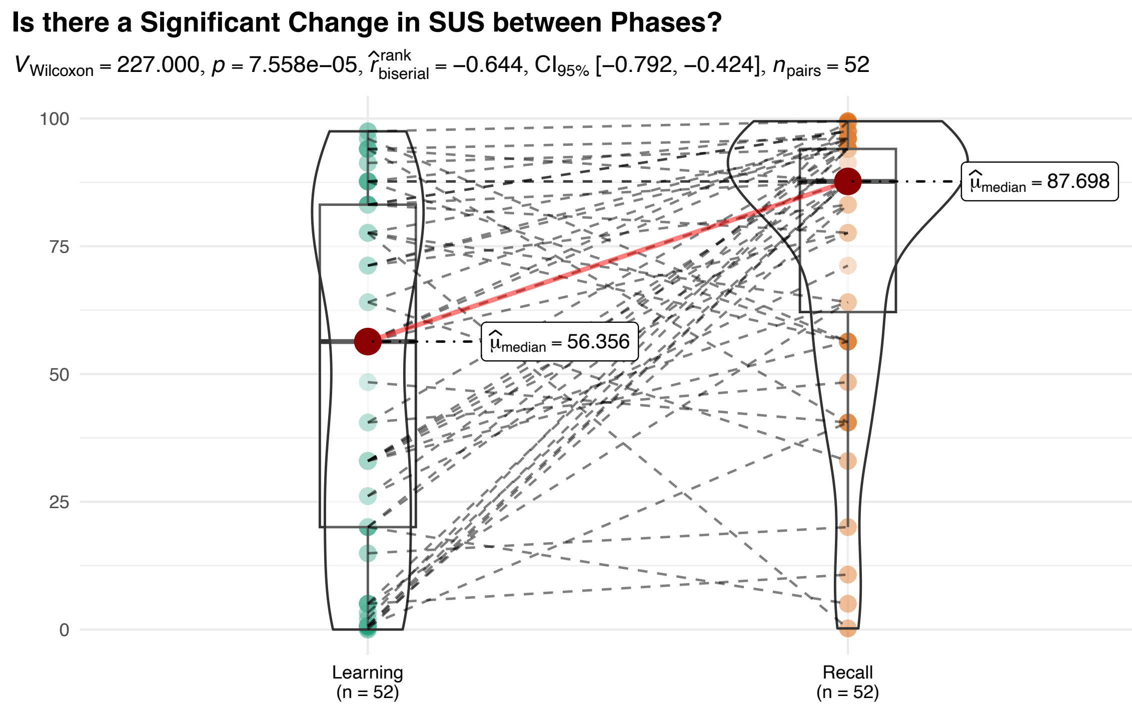


Figure 5.40: System Usability Score for Learning and Recall Phases

again employed to confirm that the change in SUS is significant (i.e., reliably non-zero) within each treatment. The results of these paired tests are summarized in Table 5.28.

These calculations utilize a normal approximation to address ties (AR) and zero values (PWI and MR), and p-values are adjusted for multiple comparisons using the Holm method. The results indicate that, despite large effect sizes ($r > 0.50$ for all treatments), none of the observed differences can be conclusively attributed to the treatments (adjusted p-values > 0.08 in all cases) given the sample size and variability. A final test was conducted to determine if the changes differ by treatment, as shown in Figure 5.41. Once again, no significant differences were identified (p = 0.479).

Table 5.28: Paired Wilcoxon Comparison for SUS Percentiles between Learning and Recall

```
# A tibble: 4 x 8
  treatment     n median_diff wilcox_statistic p_value pv_adj sig   effsize
  <chr>     <int>       <dbl>            <dbl>   <dbl>  <dbl> <chr>   <dbl>
1 PWI          14        5.41               79  0.0211 0.0844 ns      0.655
2 MR           14       44.9                76  0.0360 0.108  ns      0.537
3 AR           14       31.2                86  0.0382 0.108  ns      0.562
4 PAR          10       23.1                46  0.0665 0.108  ns      0.596
```

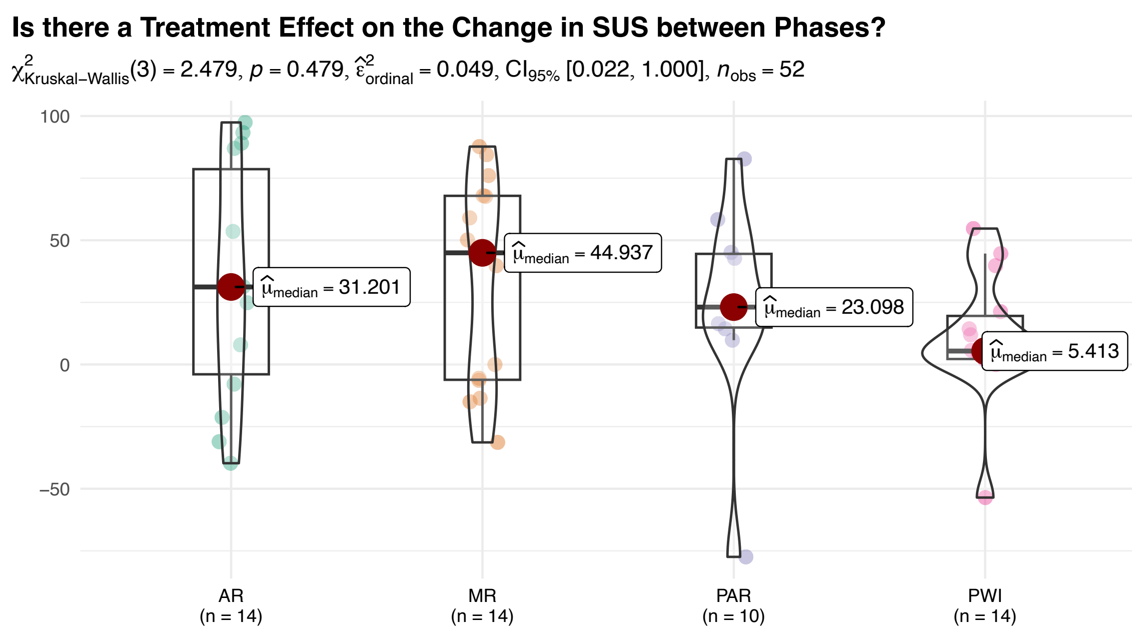


Figure 5.41: System Usability Score for Learning and Recall Phases

SUS Results

The SUS analysis revealed several key findings. During the learning phase, the overall usability reported for all instructional technologies was substantially below commercial benchmarks. This includes both the PAR, a commercial system, and PWI, the experimental control and standard instructions for the lab. While poor by SUS benchmarks, the relative scores for each treatment generally aligned with observed interactions and participant feedback. No statistically significant differences in usability were reported for the treatments used during learning. This unexpected result suggests that, while the technologies differ in complexity, their perceived usability during learning is not drastically different. Individually, participants experienced a significant increase in SUS with phase, but there isn't sufficient evidence to attribute that change to the assigned treatment, despite large observed differences and effect sizes. Furthermore, the magnitude of usability improvement did not differ significantly between treatments.

Based on the design of this study, it is likely that the improvement in SUS primarily originates from some combination of reduced complexity during retention and participant preparation during learning. However, these results provide no evidence regarding the balance of those effects. Overall, these results challenge some common assumptions about the usability of different instructional technologies and highlight the need for a more nuanced understanding of how usability impacts learning and performance in manufacturing assembly tasks.

5.6.3 Qualitative Feedback Analysis

As the final lens through which to view the overall participant experience, all provided feedback was analyzed using thematic, sentiment, and comparative analysis techniques. Several recurring themes emerged, including device usability, learning experience, physical comfort, instruction clarity, participant strategy, cognitive aspects, and overall experience. This analysis was greatly aided by the use of a large language model (Claude 3.5 Sonnet) to find patterns and synthesize hundreds of lines of feedback. The resulting insights were grouped into by- and cross-treatment observations, as summarized in the following sections.

PWI Treatment Group

Generally mixed to slightly negative sentiment.

Key Findings:

- Commonly cited difficulties in distinguishing parts, especially black ones.
- Considered simpler and more flexible than technological alternatives
- Hampered by poor overall quality of the instructions.
- Quick memorization of the process, typically by the 3rd or 4th car.
- Development of personal strategies for part retrieval and assembly order.
- Reliance on spatial memory and mental imagery of the finished car.
- Dissatisfied by the fixture requirement, wanted to be able to rotate the car.

The PWI method, while familiar and flexible, presents challenges in instruction clarity, particularly for complex or visually similar parts. Its effectiveness appears to be significantly influenced by individual characteristics like color perception and prior LEGO experience. The method promotes the development of personal memory strategies but may benefit from incorporating more detailed visual cues or 3D representations.

PAR Treatment Group

Initially positive sentiment, shifting towards negative as users became familiar with the task.

Key Findings:

- Considered helpful for initial learning but often too constraining after task familiarity.
- Inconsistent auto-advance feature was a major source of frustration, affecting agency and motivation.
- Tendency to work ahead of instructions once familiar with the process.
- Use of bin arrangement as an unexpected memory aid.
- Compared favorably to PWI for learning but considered less suitable for continued use.

PAR shows promise for initial training but faces challenges in adapting to user progress. The system's pace and control mechanisms significantly impact user experience, sometimes negatively affecting motivation and sense of agency. Future iterations might benefit from more adaptive pacing and reliable user control features.

AR Treatment Group

Overall positive sentiment, with enthusiasm for the technology tempered by usability issues.

Key Findings:

- Generally seen as better for training and confidence-building compared to other methods.
- Frequently described as "fun" or "cool," indicating high user engagement.
- Common issues with unreliable button interaction, limited field of view, and neck strain.
- Holograms sometimes obscured real-world view or lacked precision for small parts.
- Unforeseen ergonomic challenges and perceptual impacts (e.g., dimming effect).

AR demonstrates strong potential for engaging and effective training, particularly for complex tasks or diverse learning styles. However, current hardware limitations and ergonomic issues present significant challenges. Future developments should focus on improving user interaction, expanding field of view, and addressing physical comfort concerns.

MR Treatment Group

Positive overall sentiment, with appreciation for learning benefits despite technical limitations.

Key Findings:

- Valued for combining digital instructions with real-world overlay, particularly for complex steps.
- Consistent issues with tracking, limited field of view, and occasional misalignment of virtual objects.
- Development of strategies to work around system limitations.
- Seen as having a steeper learning curve than simpler methods.
- Unanticipated use of digital PWI within the MR system for better part differentiation.

Most feedback echoed that from the AR treatment group, which is expected given the similarities. MR users more commonly cited technical issues related to misalignment and tracking issues. Very few took advantage of, and none specifically cited, the added flexibility offered by MR, allowing users to freely manipulate the workpiece.

Cross-treatment Observations

Many participants noted a change in their preference and performance as they became more familiar with the task. While users of augmented methods (AR, MR, PAR) often reported these methods as helpful for initial learning, several mentioned they became limiting once the task had been mastered. Some participants shared how this and related issues affected their motivation and sense of control. Collectively, those comments suggested that good training methods don't simply direct the learner's passive actions. Instead, they promote active engagement in the learning process.

Participants also noted how their individual characteristics affected the experience. Feedback about their height, vision, or learning style influencing method effectiveness was provided for all treatments. To address these issues, several participants reported developing their own strategies to work with or around the systems. For example, many used the correspondence between assembly order and left to right bin arrangement to help learn the task, regardless of the instructional method provided.

Environmental factors were more common than expected in participant feedback. Comments about background noise, workspace setup, and lighting conditions were scattered across different treatment groups. While uncommon, this feedback emphasizes the importance of considering the suitability of these systems for the environents they are intended for deployment in.

5.7 Challenges and Unexpected Observations

A variety of challenges and unexpected observations were encountered during the conduct of the protocol. While some echo participant feedback, the following observations were recorded by the research team. All are included to provide a complementary viewpoint of participant experience. Beyond mundane challenges associated with participant scheduling and attendance, and the expected challenges of consistency in data collection (both documented elsewhere), most of what is described below can be grouped into instructional design, technical issues, and participant idiosyncracies.

Instructions in the PWI treatment were unexpectedly problematic. An assumption of the experimental design was that the PWI would provide a well-tested and validated procedure for baseline performance. The research team's prior experience with the instructions and untested confidence in their effectiveness led us to overlook these issues, which significantly impacted the PWI results. Their limitations quickly became clear, particularly when it came to discerning the type, placement, and orientation of black parts. Where the edge outline of other parts are relatively obvious, the edges of black parts are much more difficult to discern, especially in the top and three-quarter views. To compound this problem, various arrangements of parts 75, 59, 75, and 67 could achieve the results pictured in steps one and two. As a result, many participants incorrectly placed these parts, either due to misreading the instructions or simply deciding to improvise. It should be noted that the proper placement of these pieces was much more easily interpreted from the second and third step images, where earlier pieces are shown in grey, with obvious outlines. This seemingly obvious technique was only identified by one participant, at which time it was a revelation even to the research team.

Another limitation of the PWI was an over reliance on top-down imagery. Problems related to this were less prevalent, but some participants did struggle with the placement of parts 1 and 3 in the final step due to their obscured placement. This limitation was acknowledged, in part, by the inclusion of a three-quarter view of the completed assembly, but it provided

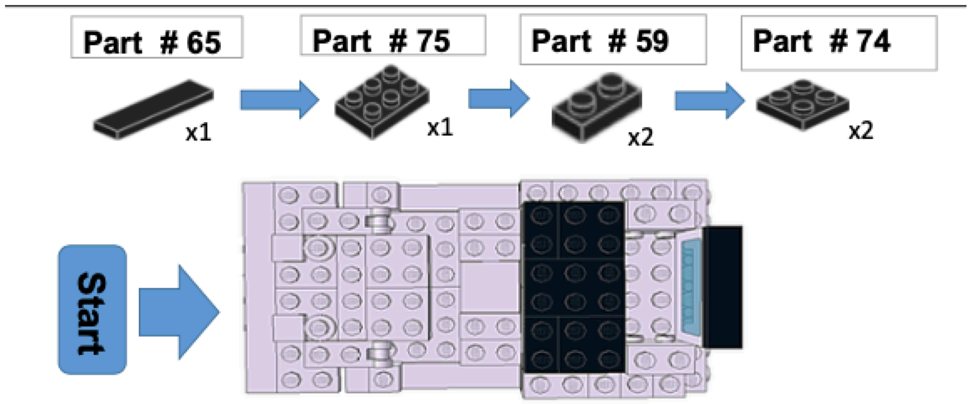


Figure 5.42: Low contrast edges for black parts on PWI instuctions.

little assistance for the problematic parts, which are also obscured in that view. This issue was less prominent it did negatively influence all measures of PWI performance.

The augmented systems were not without issue. Because gesture based systems are not 100% reliable, some participants from all augmented treatments encountered control issues that ranged in frequency and severity. In the PAR treatment, inconsistencies in the auto-advance detection systems led to false positive (unintended) and false negative (not recognized) triggers. When either occurred, participants had to recognize the error and manually correct for it, finding the correct place in sequence with forward / back buttons. A mix of confusion, frustration, and interrupted flow resulted, though its impact varied. Similar problems plagued the manual trigger methods used by PAR, AR, and MR. Some participants using HoloLens2 based systems also encountered tracking discontinuities, during which the interface momentarily goes blank (dropouts). The time lost to these events was accounted for in the TCT calculations but their indirect effects on participant performance are not easily quantified. This was especially common in the MR treatment where sophisticated model-based tracking methods were employed, and was exacerbated by some physical characteristics of individual participants.

Participants exhibited a range of tendencies in the way they directed their view. While all used some a combination of head and eye motion, some were much more reliant on one than the other. The HL2's limited field of view exaggerated these differences by requiring AR and MR users to adjust move more to keep their view centered in the display. Taller participants that were more reliant on head motion typically resulted in very top-down view of the work surface, with too few contrasting features to support robust tracking. This resulted in more tracking dropouts that occasionally led to other system instabilities. In contrast, shorter participants experienced more robust tracking overall in AR and MR, but may have benefited less from its overlays due to the less informative viewing angle. Participants of average or greater height that relied mostly on eye motion were prone to have a line of sight centered *below* the HL2's display when looking at the work surface. This obviously limited the utility of AR and MR instructions, but that effect was not directly quantifiable. Participants that exhibited this particular behavior were labeled "down-lookers" by the research team, a sign that the problem was unexpected but not uncommon.

Formally, these forms of head-eye coordination are described as "gaze patterns," and our so-called down-lookers would be classified as "eye-dominant gazers." The ramifications of this were largely overlooked by the research team prior to witnessing it during the trials. Presumably, we naturally accounted for any issues encountered by adjusting our viewpoint, HL2 fitment, and/or behavior, without acknowledging the corrections. Unfortunately, the experimental design minimized ways that participants could compensate. In particular, the fixed height work surface required for the PAR system prevented us from accommodating for these individual differences.

The tracking accuracy and display brightness of the HL2 sometimes made it difficult for participants to discern the proper placement of parts. The small size of Lego parts and precise layout of their studs made the tasks sensitive to tracking accuracy. Poor registration or drift in tracking sometimes made it difficult to interpret the instruction and place bricks properly. At times this seemed complicated by the presence of the brick hologram itself, which partially obscured the real-world objects.

Surprisingly few participants assigned to the MR treatment took advantage of its primary affordance. Despite the absence of a fixture, many completed each task with the workpiece in its initial orientation. Some rotated it, but very few picked it up to take full advantage of the manipulation capabilities of this system. Worse yet, the lack of a fixture led to issues with workpiece stability, causing unwanted translation and/or rotation that sometimes resulted in breakages that might not have otherwise occurred.

AR and MR users encountered a variety of other minor issues related to the system's tracking, battery life, performance, and brightness. Participants with long hair that draped into view, bulky long sleeves, or unusually long fingernails led to hand tracking issues that were mostly easily corrected for. The limited battery life of the HL2 required careful maintenance when several trials were conducted back-to-back. The limited capabilities of the the HL2's CPU/GPU sometimes led to framerate and/or latency issues that complicated the choreography of MR demonstrations, along with monitoring, recording, and editing videos, but rarely had any noticeable effect on participant performance.

Except as/if noted elsewhere, no modifications were made to the protocol to account for these issues. Most patterns emerged late enough that such changes would have invalidated previous results and compromised any findings. Given the practical constraints of this study, it was decided to carry on with an understanding of the potential limitations and how they might affect the results. Mixed-effect modeling methods were used to help isolate the random effect of some of these issues, but that was not always the most suitable analytical approach. Resulting implications to interpretation and generalization will be detailed in the Conclusions chapter to follow, along with recommendations to improve future work based on these findings.

6 Conclusions

This chapter synthesizes the key findings, insights, and implications of this study on augmented technologies in manufacturing assembly training. It begins with a comprehensive summary of results, addressing the initial research questions and hypotheses. That is followed by a discussion of the study's main conclusions, its contributions to the field, and the implications for both research methods and theory. An examination of the the limitations of the work and its generalizability to broader contexts follows. Finally, directions for future research are outlined, aiming to build upon this study's foundations and address its identified limitations.

6.1 Summary of Results

This study examined the effects of different instructional methods (PWI, PAR, AR, MR) on learning, recall, and retention of a manufacturing assembly task. The results provide insights into the effectiveness of these methods across multiple phases and dimensions of performance.

6.1.1 Hypothesis Testing

This section will briefly summarize the quantitative results in the context of the three questions originally posed as hypotheses in Section 4.3.4.

H_1: How does each treatment affect performance during the learning phase?

Table 6.1 summarizes the outcomes related to H_1. These results suggest that while traditional paper instructions and projected AR allowed for faster initial performance, the more immersive AR and MR technologies led to higher quality outcomes, albeit with slower execution.

Table 6.1: Summary of Quantitative Results for Learning Claims

Claim	Result	Key Findings
H_{1a}: Average task completion time varies with treatment	Accepted	PWI and PAR are both significantly faster than AR and MR. No significant difference between PWI and PAR or between AR and MR.
H_{1b}: Learning rates vary with treatment	Accepted	PWI demonstrates the steepest learning curve when controlling for iTCT, followed closely by PAR. AR and MR show slower rates of improvement. Relationship: PWI ≈ PAR > (AR ≈ MR)
H_{1c}: Average error count per car varies with treatment	Accepted	All augmented instruction methods (PAR, AR, MR) result in significantly lower error rates compared to PWI.

The observed performance differences observed during the learning phase stem from a balance between the familiarity and simplicity of traditional methods versus the potentially deeper, more spatially-integrated learning offered by PAR, AR, and MR technologies. Initial performance in the PWI condition was likely influenced by a combination of participants' prior experience with paper-based instructions and the lack of built-in error checking to slow them down. But when combined with limitations in the instructional design, PWI resulted in the lowest quality by far. In contrast, the relatively high TCTs and low UCEs observed for augmented methods are the result of unfamiliar interfaces and integrated error-checking through step-by-step guidance and real-time feedback. As predicted by cognitive learning theory and embodied cognition, the additional time and engagement likely contributed to the low error counts observed for augmented instruction.

H_2: How does each treatment affect performance during the recall phase?

The Recall-related results summarized in Table 6.2 suggest that augmented technologies, particularly AR and MR, may lead to better retention and more independent task performance after initial training.

Table 6.2: Summary of Quantitative Results for Recall Claims

Claim	Result	Key Findings
H_{2a}: OEE varies with treatment	Accepted	PAR, AR, and MR all result in statistically significant and practically meaningful improvements in OEE compared to PWI.
H_{2b}: PWI reliance varies with treatment	Accepted	Statistically significant differences in reliance exist between treatments, but specific pair-wise differences are not clear enough to declare. AR and MR tend to reduce reliance more than PAR and PWI.

The observed differences in recall phase performance likely stem from the nature of learning facilitated by each instructional method. The significantly lower OEE for PWI can be attributed to errors that negated its effectiveness. This confirms that the initial speed advantage of PWI came at the cost of proper skill acquisition and retention. In contrast, higher OEE scores for AR methods, particularly AR and MR, indicate better retention and more independent task performance. This aligns with theories of embodied cognition and active learning, where the spatial integration and hands-on interaction provided by these technologies likely led to deeper processing and more robust mental models of the assembly process. The reduced reliance on instructions for AR and MR users further supports this interpretation. While PAR showed improvements over PWI, its intermediate performance in both OEE and reliance hints at a balance between the benefits of augmentation and the limitations of 2D presentation.

H_3: How does each treatment affect performance during the retention phase?

Table 6.3 summarizes the findings that relate iTCT and iUCE with treatment. Overall, the implications for retention are mixed, with the UCE model providing stronger evidence for an effect of instructional method on retention than the TCT model. This difference suggests that error-related performance improvements may be more sensitive to instructional methods than completion times.

Table 6.3: Summary of Quantitative Results for Retention Claims

Claim	Result	Key Findings
H_{3a}: Increase in TCT over retention interval varies with treatment	Rejected	No statistically significant evidence found for the effect of instructional method on the increase in TCT over the retention interval.
H_{3b}: Increase in UCE over retention interval varies with treatment	Accepted	PWI method showed a significantly lower rate of error increase compared to the reference AR method. Other treatments suggest meaningful effects but did not reach statistical significance.

The lack of significant differences in iTCT across treatments can likely be taken at face value–the observed decreases in temporal performance are dominated by the natural process of skill decay. While the evidence that PWI users experienced a lower rate of error increase seems to challenge expectations, it can likely be attributed to a ceiling effect on UCE. Simply put, it was hard for many PWI results to get worse. Another plausible contributor, that augmented users were too dependent on the provided guidance, seems discredited by the reliance findings during recall.

6.1.2 Survey Results

The surveys of workload (TLX) and usability (SUS) bring another dimension to the analysis by quantifying critical symptoms of user experience and possibly suggesting uncontrolled moderating factors. Table 6.4 summarizes those findings:

Table 6.4: Summary of Quantitative Results for Workload and Usability Claims

Claim	Result	Key Findings
Workload Analysis (NASA TLX)	No Significant Differences	No significant differences in overall task workload between treatments. Temporal Demand was consistently the most significant contributor to workload across all treatments.

Claim	Result	Key Findings
Usability Analysis (SUS)	No Significant Differences	No statistically significant differences in usability were reported for the treatments used during learning. Significant increase in SUS from learning to recall phase for all treatments, but not attributable to specific treatments.

The lack of significant differences in TLX scores across treatments suggests that participants adapted to the challenges of each method, resulting in similar overall workload experiences. The consistency of Temporal Demand as the primary contributor to workload indicates that the timed nature of the task, rather than the instructional method, dominated workload perception. The unintended emphasis on time pressure likely overshadowed other potential differences between treatments.

User adaptation likely accounts, in part, for the absence of significant differences in SUS scores during learning, despite varying technological complexities and observed performance differences. This also suggests that the benefits of interactive features in AR/MR systems were likely balanced by technical issues and comfort concerns, resulting in similar overall usability perceptions across treatments. Those perceptions may also reflect factors that are disconnected from performance metrics. For example, the novelty of AR/MR technologies might have positively influenced usability ratings despite initial performance differences. In short, SUS may not be a valid measure of the efficacy of these training methods.

The significant increase in SUS scores from learning to recall across all treatments likely reflects both reduced interface complexity and increased user confidence. However, the lack of treatment-specific differences in this improvement indicates that all methods similarly prepared users for independent task performance, regardless of their initial technological sophistication.

6.1.3 Qualitative Results

Participant feedback revealed nuanced experiences across treatments. PWI was perceived as simple and flexible, but hampered by poor instruction quality, particularly between similar parts. PAR was initially helpful but became constraining as users gained familiarity. The inconsistent auto-advance feature significantly impacted user experience, highlighting the importance of reliable automation and/or direct user control in training systems. AR and MR were seen as engaging and effective for training, generating enthusiasm despite hardware limitations. Users appreciated the combination of digital instructions with real-world overlay, particularly for complex steps. However, technical issues like unreliable interactions, limited field of view, and tracking problems presented significant challenges.

Across all treatments, participant preferences and performance evolved with task familiarity. Many developed personalized strategies to overcome various system limitations, and individual characteristics like height and gaze patterns notably influenced method effectiveness. Environmental factors such as background noise, workspace setup, and lighting conditions were also noted by some participants. All of this highlights the importance of user-centered instructional design, adaptive training methods, flexible process designs, and environmental considerations in the design and implementation of these systems.

6.1.4 Researcher Observations

Several common themes emerged when observing participant trends and behavior, some of which are supported by participant feedback from Section 5.6.3 and briefly summarized above. Most obviously, nearly all participants showed a clear improvement in speed and confidence as their trial progressed, regardless of treatment. The rate with which they did so varied, and some seemed more confident than others with the nature and design of the experiments. While each treatment had its unique challenges, participants generally adapted and improved over time. The AR and MR treatments offered helpful visualization but came with technical challenges, while the PWI treatment allowed for more flexibility but sometimes led to confusion with part identification and orientation.

Among AR and MR participants, the impact of tracking and FOV limitations described in Section 5.7 varied with individual height and preferred gaze patterns, an outcome that was neither expected nor accounted for in the study design. For all augmented treatments, the inconsistency of gesture detection systems was a common cause of frustration. The same could be said of the shortcomings of PWI instructions, all of which led to adaptation.

Though specifically discouraged, it was not uncommon for participants to develop individual assembly strategies. Efforts to improve their performance and/or address perceived limits or shortcomings of the instructional method typically involved skipping, combining, or reordering steps. This was most common in the PWI treatment, where no step-wise controls existed. It was also prevalent in the PAR condition, where users that struggled with the gesture recognition system would sometimes resort to batching steps: working from memory for a few steps before advancing through all the relevant instructions.

6.1.5 Collective Insights

The quantitative results align with much of the qualitative feedback. For instance, PWI's faster TCT but higher error rates corresponds with participant comments about its simplicity and poor instruction quality. Similarly, the improved OEE and reduced reliance on instructions for AR and MR users during recall aligned with feedback about these methods being effective for training and confidence-building.

However, several unexpected findings and diverging quantitative / qualitative results also emerged. While PWI showed poorer performance in terms of error rates during learning, it unexpectedly demonstrated potential for improving learning durability. The lack of significant differences in TLX scores between treatments contrasted with the clear preferences and challenges expressed in qualitative feedback. AR and MR showed slower TCT but lower error rates quantitatively, yet participants generally provided positive feedback about these technologies' effectiveness. Finally, although SUS scores improved across all treatments from learning to recall, qualitative feedback indicated ongoing usability challenges with augmented technologies.

These divergences emphasizes the importance of considering both quantitative and qualitative data in evaluating instructional methods, as each provides unique insights into user experience and performance. They may also help to explain the mixed results obtained during this study.

6.1.6 Key Findings

To complete the summary of results, the ten most significant findings are enumerated below, listed roughly in order of importance:

1. Error Reduction with Augmented Methods: AR and MR technologies led to significantly lower error rates compared to traditional methods (PWI) during the learning phase. While this reduction is notable, it may be partially attributable to limitations

in PWI instructions and potential floor effects for augmented methods. Further research is needed to fully understand the extent and implications of this error reduction across various task complexities and over longer periods.

2. Speed-Accuracy Trade-off: While augmented methods reduced errors, they also led to slower initial execution compared to traditional methods. This trade-off is crucial for organizations to consider when implementing training systems and warrants further investigation, particularly in more complex tasks and real-world manufacturing environments.
3. Evolution of Learning Effectiveness: The effectiveness of instructional methods varied across learning phases, highlighting the need for longitudinal studies. AR and MR showed benefits in initial error reduction and recall performance, while PWI unexpectedly demonstrated potential for improving error-rate durability in retention. Research exploring longer intervals between tasks could provide valuable insights into the long-term effectiveness of different methods.
4. Adaptive User Behavior: Participants across all treatments developed personal strategies and preferences as they gained task familiarity. This emphasizes the importance of flexible training systems that accommodate user adaptation and suggests a need for research into how instructional methods can best support this natural learning process.
5. Impact of Individual Differences: Characteristics such as height, gaze patterns, and learning styles significantly influenced the effectiveness of each method, especially for AR and MR systems. This underscores the need for adaptable designs in training technologies and further research into how these individual differences interact with various instructional methods.
6. Complex Relationship between Perceived and Actual Performance: The discrepancy between quantitative workload measures (TLX) and qualitative feedback reveals that user adaptation may result in similar perceived workload despite varying technological complexities. This highlights the need for comprehensive evaluation methods in future research.
7. Affordances and Limitations Balance: While AR/MR affordances like spatial registration and user-centric displays may contribute to deeper learning, current technical limitations partially offset these benefits. Further research is needed to quantify the impact of specific affordances and overcome existing limitations.
8. Dynamic User Experience: User preferences and satisfaction evolved as participants became more familiar with the task, indicating that the effectiveness of training meth-

ods may vary depending on the stage of learning. This suggests a need for research into adaptive instructional methods that evolve with user proficiency.

9. Environmental Factors: The unexpected influence of environmental conditions on method effectiveness emphasizes the importance of considering deployment contexts in training system design and future research.
10. Long-term Usability Perceptions: SUS scores improved from learning to recall across all treatments, suggesting that familiarity with the task may improve perceived usability regardless of the initial technological sophistication. This highlights the need for longitudinal studies in usability research for training systems.

6.2 Conclusions

Together, these findings reveal complex relationships between instructional methods, various performance metrics, and individual traits. The central research question of this work, per Section 3.2 is:

> *How do different AR/MR instructional methods, designed to leverage specific affordances, impact operator learning, recall, and retention in a real-world manufacturing assembly training context?*

Realizing this question is difficult to answer directly in a concise manner, it was broken down into five supporting questions, each of which are revisited below.

> *What are the relative effects of various AR/MR technologies on immediate learning outcomes, such as task completion time and error rates, compared to traditional paper-based instructions?*

These findings reveal a clear trade-off between speed and accuracy in immediate learning outcomes. While traditional methods facilitated faster initial performance, AR/MR technologies promoted higher quality outcomes. This suggests that the choice of instructional method should be guided by whether initial speed or accuracy is prioritized in a given training context.

> *How do these AR/MR technologies influence long-term recall and retention of assembly skills, as measured by performance on the same task after a designated period without further training?*

The impact of AR/MR on long-term performance appears more nuanced than initially expected, with differing effects on speed versus accuracy retention. While augmented technologies generally led to better recall performance, unexpected results in the retention phase highlight the complexity of long-term skill retention and the need for further investigation in this area.

> *To what extent do the specific affordances of each AR/MR technology, such as hands-free interaction, spatial registration, and user-centric displays, contribute to the observed learning, recall, and retention outcomes?*

Affordances such as spatial registration and user-centric displays appear to contribute significantly to learning outcomes, albeit with some trade-offs. These features may initially slow task completion but potentially lead to deeper learning and better long-term retention. However, current technical limitations partially offset these benefits, underscoring the need for continued refinement of AR/MR technologies.

> *How do operator characteristics, such as related experience or demographics, influence the effectiveness of each instructional method?*

This study reveals that individual differences significantly impact the effectiveness of each instructional method. Factors such as height, gaze patterns, and learning styles interact with the features of each method, highlighting the need for adaptable designs that can accommodate a range of user characteristics and preferences.

> *What are the perceived workload, usability, and user satisfaction associated with each AR/MR technology, and how do these factors relate to learning, recall, and retention outcomes?*

The relationship between perceived workload, usability, and actual performance is complex. While quantitative measures showed little difference between methods, qualitative feedback revealed evolving user preferences and adaptation strategies. This suggests that effective implementation of AR/MR in training contexts requires consideration of both objective performance metrics and subjective user experiences.

In conclusion, this study reveals that AR/MR instructional methods in manufacturing assembly training offer a nuanced trade-off between immediate performance and long-term learning outcomes. While these technologies can enhance accuracy and potentially deepen learning, their effectiveness is moderated by individual user characteristics and evolving preferences. The impact of AR/MR methods on learning, recall, and retention is not uniform across all aspects of performance, highlighting the need for context-specific implementation strategies. As these technologies continue to develop, their unique affordances promise to reshape training approaches, but realizing their full potential requires careful consideration of both immediate task demands and long-term skill development goals.

6.3 Contributions

This work makes several significant contributions to the field of AR/MR technologies in manufacturing training, directly addressing the central research question: "How do different AR/MR instructional methods, designed to leverage specific affordances, impact operator learning, recall, and retention in a real-world manufacturing assembly training context?" The primary contributions relate to theoretical, methodological, empirical, practical, and human factors considerations. Each correspond with one of the supporting research questions, as detailed below.

6.3.1 Theoretical: Affordance-Based Evaluation Framework

This study develops and applies an affordance-based framework for evaluating AR/MR technologies in manufacturing training contexts, addressing the research question: "To what extent do the specific affordances of each AR/MR technology contribute to the observed learning, recall, and retention outcomes?" The framework provides a new theoretical lens through which to understand and assess the effectiveness of these technologies. By focusing on specific affordances such as spatial registration, user-centric displays, and hands-free interaction, this framework offers a more structured and theoretically grounded method of assessing AR/MR technologies in manufacturing training, enabling more precise comparisons and insights across different systems and contexts. Evidence from the study suggests that affordances like spatial registration and user-centric displays contribute significantly to learning outcomes, albeit with some trade-offs in initial task completion time.

6.3.2 Methodological: Comprehensive Multi-Phase Study Design

A comprehensive study that captures learning, recall, and retention outcomes was designed and implemented to address the question: "How do these AR/MR technologies influence long-term recall and retention of assembly skills?" This multi-phase approach allows for the evaluation of both immediate learning outcomes and longer-term skill retention, offering insights into how the effectiveness of different instructional methods evolves over time. The study revealed that while AR and MR showed benefits in initial error reduction and recall performance, traditional methods unexpectedly demonstrated potential for improving error-rate durability in retention.

6.3.3 Empirical: Quantification of Speed-Accuracy Trade-offs

By providing empirical evidence to quantify the speed-accuracy trade-offs between traditional and AR/MR-based instructional methods in manufacturing assembly tasks, this study addresses the question: "What are the relative effects of various AR/MR technologies on immediate learning outcomes?" It reveals that while AR/MR technologies often lead to slower initial performance, they result in higher quality outcomes with fewer errors. Specifically, PWI and PAR demonstrated faster initial performance but higher error rates, while AR and MR showed slower initial execution but significantly lower error rates during the learning phase. By quantifying this trade-off across different phases of learning and retention, this study provides valuable insight into important implementation considerations.

6.3.4 Practical: Insights on the Interplay of Environmental and User Characteristics

The question: "How do operator characteristics and environmental conditions influence the effectiveness of each instructional method?" is addressed through relevant practical insights. The research found that factors such as user height, workspace setup, and lighting conditions had some unexpected effects on the effectiveness of AR and MR systems. For example, taller participants and those with specific gaze behaviors (e.g., "down-lookers") experienced more tracking issues with AR and MR systems. Additionally, the study revealed how environmental factors like background noise and lighting affected the user experience across different instructional methods. These findings provide valuable guidance for the real-world implementation of AR/MR training systems, informing the design of more flexible and effective solutions that can accommodate a diverse range of users and deployment contexts in manufacturing settings.

6.3.5 Human Factors: User Adaptation and Perceived Workload / Usability

This study contributes valuable insights into the cognitive and experiential aspects of AR/MR training systems, addressing the question: "What are the perceived workload, usability, and user satisfaction associated with each AR/MR technology, and how do these factors relate to learning outcomes?" By examining the complex relationships between perceived workload, usability, and actual performance across different phases of learning, this research reveals important considerations for the design and implementation of AR/MR training systems. The study's significant, counter-intuitive findings concerning the relationships (or lack thereof) between complexity, workload, and usability suggest that the participants are less sensitive to technical complexity and more influenced by other factors (e.g., time pressures) than expected. This contribution underscores the importance of integrating both objective performance metrics and subjective user experiences in the evaluation and development of AR/MR training systems for manufacturing contexts.

These contributions collectively advance our understanding of AR/MR technologies in manufacturing training. Collectively, they shed significant light on the central research question and offer a variety of insights that can guide future research and implementation in this rapidly evolving field.

6.4 Implications to Method and Theory

The contributions outlined above have important implications for both research and practice in augmented-assisted training.

6.4.1 Methodical Implications

1. Research Design: The multi-phase, mixed-methods approach employed in this study demonstrates the value of comprehensive assessment in capturing the effects of

AR/MR technologies on learning, recall, and retention. Future studies may consider adopting similar approaches to fully understand the impact of these technologies.

2. Training System Design: The observed trade-offs between speed and accuracy, as well as the impact of individual differences, underscore the need for adaptable AR/MR training systems. Designers should consider incorporating features that can adjust to user characteristics and learning progress.

3. Implementation Strategies: The varying effectiveness of different methods across learning phases suggests that organizations should consider using a combination of instructional approaches, potentially transitioning between methods as learners progress from novice to expert.

4. Evaluation Metrics: The discrepancies observed between performance and user experiences highlight the importance of supporting comprehensive quantitative performance measures with diverse qualitative metrics when assessing the effectiveness of AR/MR training systems in the context of participant idiosyncrasies.

5. Technology Development: The identified limitations of current AR/MR systems, particularly in terms of ergonomics and adaptability, provide clear directions for technology developers to improve these systems for manufacturing training applications.

6.4.2 Theoretical Implications

1. Cognitive Load Theory: The findings on speed-accuracy trade-offs and learning effectiveness across phases enhance our understanding of how different instructional methods affect cognitive load over time.

2. Embodied Cognition: The benefits observed from spatial registration and user-centric displays in AR/MR systems provide empirical support for embodied cognition theories in learning processes.

3. Active Learning Theory: The observation of adaptive user behavior across all treatments supports the importance of active engagement in the learning process, as posited by active learning theories.

4. Technology Acceptance Model: The complex relationship found between perceived workload, usability, and actual performance suggests a need to refine how we apply technology acceptance models to AR/MR training systems.

5. Affordance Theory: The study's findings on the balance between AR/MR affordances and limitations contribute to our understanding of how technological affordances translate into learning outcomes in practical settings.

Together, these implications highlight the complexity of evaluating AR/MR training systems and suggest several directions for refining both research methodologies and theoretical frameworks in this field. They emphasize the need for more nuanced, multi-faceted approaches to studying the effectiveness of these technologies in real-world learning contexts.

6.5 Limitations and Generalizability

While this study provides valuable insights into the effectiveness of different instructional methods for manufacturing assembly tasks, several limitations should be considered when interpreting the results. All are discussed below, grouped into the categories ecological generalizability, data measurement and analysis, and technology limitations.

A number of considerations may limit the ecological generalizability of these findings for real-world industrial tasks, workers, and environments. The convenience sample was relatively small (n=54) and comprised primarily of university students without significant manufacturing experience. The assembly task, while designed to simulate real-world manufacturing processes, was relatively simple and short in duration, leading to floor effects in the observed error rates. Only a single assembly task was simulated, and the controlled laboratory setting may not fully reflect the conditions of a real manufacturing environment. Results may differ for experienced participants or different task types, including those that are more complex or of longer duration, in real world settings. Additionally, the fixed ergonomics of the workstation couldn't accommodate diverse user needs, especially in AR/MR conditions, potentially impacting comfort, system performance, and AR/MR effectiveness for some participants.

Limitations in data measurement and analysis should also be considered. While comprehensive, the measures used (e.g., TCT, UCE, OEE) may not capture all relevant aspects of performance and learning in manufacturing contexts. The qualitative feedback, while valuable, may be subject to recall bias or influenced by participants' preconceptions about different technologies. Furthermore, the novelty of PAR, AR, and MR technologies to all participants may have influenced engagement and performance in ways that might not persist with long-term use. The variable retention period with a single measure (up to 44 days) may not provide a complete understanding of training durability. The imbalanced repeated measures due to varying task iterations during the learning phase led to compromises in analysis, challenging model fitting and reducing the accuracy of extrapolations. Additionally, the analysis did not employ cross-validation or other methods to prevent bias and limit overfitting for model-based approaches, potentially leading to underestimated error rates and overly optimistic model performance estimates.

Finally, the PAR, AR, and MR systems used in the study had various known limitations in terms of field of view, comfort, reliability, and interaction methods. More advanced systems or implementations might yield different results, highlighting the need for caution when comparing these findings with results from other technological setups.

These limitations do not invalidate the findings but provide important context for their interpretation and application. They also suggest avenues for future research to address these constraints and further validate the results in diverse settings and with larger, more representative samples.

6.6 Future Work

Building on the findings and limitations of this study, future work will focus on furthering the understanding of AR/MR technologies in manufacturing training while increasing the ecological validity of the research. Several publications are planned from this work, each focusing on specific contributions of the work. As summarized in Table 6.5, these publications will extend the current analysis by exploring interactions between affordances, refining performance metrics (e.g., a more nuanced OEE calculation), conducting error type analysis, and investigating the relationships between workload, usability, and performance measures.

Table 6.5: Proposed Publications

Working Title	Focus
"An Affordance-Based Framework for Evaluating AR/MR Training Systems in Manufacturing"	Affordance-based framework, Theoretical underpinnings, Application to AR/MR training systems
"Designing and Implementing AR/MR Technologies for Manufacturing Assembly Training: A Comparative Study of Learning Outcomes"	Literature review and theoretical background, Experimental design and methodology, Learning phase results, Immediate impacts on task completion time and error rates
"Long-term Effectiveness of AR/MR Technologies in Manufacturing Training: An Analysis of Recall and Retention"	Recap of experimental design, Recall phase results, Retention phase findings, Implications for long-term skill acquisition
"Beyond Performance Metrics: Assessing Workload, Usability, and Individual Factors in AR/MR-Assisted Manufacturing Training"	TLX and SUS results analysis, Demographic and individual factors, Interplay between workload, usability, and performance, User-centered design implications
"Modeling Learning Curves in AR/MR-Assisted Manufacturing Training: Challenges and Methodological Approaches"	Methodological aspects, Challenges in modeling learning curves, Solutions for imbalanced repeated measures data
"Qualitative Analysis of Error Types in AR/MR vs. Traditional Manufacturing Training: Implications for Training Design and Quality Control"	Error type analysis, Qualitative insights into error nature, Impact of instructional methods on errors

These topics are suitable for publication in a range of reputable journals, including IEEE Transactions on Visualization and Computer Graphics, International Journal of Human-Computer Studies, Journal of Manufacturing Systems, and Computers in Industry; a list that reflects the interdisciplinary nature of the research and the value of its contributions.

Subsequent studies with increasing complexity and ecological validity are intended. Each will integrate improvements to both the experimental and system design intended to address the limitations and implications previously identified. Anticipated system upgrades include improved application design, variable-height workstation with integrated multi-camera video capture, and modern VR / VST hardware. This would provide enhanced user experience, streamlined data capture, more robust tracking, and support user adaptation.

Enhancements to the experimental design will prioritize ecological validity, the robustness of results, and greater insight into underlying learning mechanisms. This can be accomplished by refining participant selection, task design, and study duration, implementing more comprehensive assessment methods with additional qualitative and quantitative measures, and exploring the impact of instructional design methods on skill acquisition and retention.

This research trajectory is best suited for a series of industrial experiments, progressing from controlled lab settings to longitudinal on-site studies with industry partners. Ideally, these studies will feature real-world tasks that vary in type and complexity to help validate and refine our understanding of AR/MR effectiveness in authentic manufacturing contexts.

Through this future work, we aim to bridge the gap between theoretical insights and practical applications, ultimately contributing to the development of more effective and adaptable AR/MR training systems for the manufacturing industry.

6.7 Closing

In closing, this study provides valuable insights into the effectiveness of AR/MR technologies in manufacturing assembly training, highlighting both their potential benefits and current limitations. The research reveals a nuanced interplay between instructional methods, performance metrics, and individual user characteristics, emphasizing the need for adaptive and context-sensitive training solutions. While AR/MR technologies demonstrate promise in reducing errors and potentially fostering deeper learning, their implementation requires careful consideration of speed-accuracy trade-offs, user adaptation processes, and environmental factors.

As the field of AR/MR-assisted training continues to evolve, future research should focus on addressing the limitations identified in this study and expanding our understanding of these technologies in diverse manufacturing contexts. By building on the affordance-based framework and methodological approaches developed here, researchers and practitioners can work towards creating more effective, user-centered AR/MR training systems that balance immediate performance needs with long-term skill development goals. Ultimately, this line of inquiry has the potential to significantly enhance workforce training in manufacturing, contributing to improved productivity, quality, and adaptability in an increasingly complex industrial landscape.

References

Albert, R., Patney, A., Luebke, D., & Kim, J. (2017). Latency Requirements for Foveated Rendering in Virtual Reality. *ACM Transactions on Applied Perception*, *14*(4), 1–13. https://doi.org/10.1145/3127589 (cit. on p. 32).

Alexander, S. A., Rozo, J. S., Donadio, B. T., Tenhundfeld, N. L., de Visser, E. J., & Tossell, C. C. (2019). Transforming the Air Force Mission Planning Process with Virtual and Augmented Reality. *SIEDS, 2019 Systems and Information Engineering Design Symposium (SIEDS)*, 1–4. https://doi.org/10.1109/SIEDS.2019.8735617 (cit. on p. 44).

Ali, R. M., & Deif, A. M. (2014). Dynamic Lean Assessment for Takt Time Implementation. *Procedia CIRP*, *17*, 577–581. https://doi.org/10.1016/j.procir.2014.01.128 (cit. on p. 88).

Alves, J. B., Marques, B., Ferreira, C., Dias, P., & Santos, B. S. (2022). Comparing augmented reality visualization methods for assembly procedures. *Virtual Reality*, *26*(1), 235–248. https://doi.org/10.1007/s10055-021-00557-8 (cit. on pp. 58–62).

Ante, L. (2021). Digital Twin Technology for Smart Manufacturing and Industry 4.0: A Bibliometric Analysis of the Intellectual Structure of the Research Discourse. *Manufacturing Letters*, *27*, 96–102. https://doi.org/10.1016/j.mfglet.2021.01.003 (cit. on pp. 11, 14).

Aqlan, F., & Hui, Y. (2020). Detecting mental disorders with sensor-based virtual reality: Gaming technology creates environment that simulates lab assessment tests. *ISE: Industrial & Systems Engineering at Work*, *52*(10), 32–35 (cit. on p. 43).

ASTM. (2019). Standard Specification for 3D Imaging Data Exchange, Version 1.0. (Cit. on p. 120).

Atkinson, R. C., & Shiffrin, R. M. (1968, January). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (pp. 89–195, Vol. 2). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3 (cit. on p. 96).

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355 (cit. on pp. 1, 20–22).

Azuma, R. T. (2016). The Most Important Challenge Facing Augmented Reality. *Presence*, *25*(3), 234–238. https://doi.org/10.1162/PRES_a_00264 (cit. on p. 63).

Azuma, R. T. (2017). Making Augmented Reality a Reality. *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, JTu1F.1 (cit. on pp. 25, 64).

Azuma, R. T. (2019). The road to ubiquitous consumer augmented reality systems. *Human Behavior and Emerging Technologies*, *1*, 26–32. https://doi.org/10.1002/hbe2.113 (cit. on pp. 1, 21, 42, 52).

Azuma, R. T., & Bishop, G. (1994). Improving Static and Dynamic Registration in an Optical See-through HMD Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques. *SIGGRAPH '94*, 197–204. https://doi.org/10.1145/192161.192199 (cit. on pp. 1, 18, 33, 42).

Bailenson, J. (2018, January). *Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do*. W. W. Norton & Company. (Cit. on pp. 20, 21, 45).

Ballard, V., O'Leary, D. J., & Sesek, R. F. (2024). The Role of Technology in Enhancing Workplace Ergonomics for Neurodiverse Employees: A Performance and Safety Perspective. *Applied Ergonomics Conference 2024* (cit. on p. 106).

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776 (cit. on p. 73).

Barfield, W. (2015). Wearable Computers and Augmented Reality: Musings and Future Directions. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 22, 44, 65).

Bass, L., Mann, S., Siewiorek, D., & Thompson, C. (1997). Issues in wearable computing. *ACM SIGCHI Bulletin*, *29*(4), 34–39. https://doi.org/10.1145/270950.270975 (cit. on p. 38).

Behzadan, A. H., Dong, S., & Kamat, V. R. (2015). Recent Advances in Augmented Reality for Architecture, Engineering, and Construction Applications. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on p. 36).

Bilal, M. (2023, February). Here are 8 AI-powered apps that will make your academic research easier and faster... (Cit. on p. 4).

Billinghurst, M., Reichherzer, C., & Nassani, A. (2015). Collaboration with Wearable Computers. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on p. 22).

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals* (Vol. Handbook I: Cognitive Domain). David McKay Company. (Cit. on pp. 48, 96).

Borsci, S., Lawson, G., & Broome, S. (2015). Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance. *Computers in Industry*, *67*, 17–26. https://doi.org/10.1016/j.compind.2014.12.002 (cit. on pp. 69, 73, 74).

Bottani, E., Assunta, C., Murino, T., & Vespoli, S. (2017, September). *From the Cyber-Physical System to the Digital Twin: The process development for behaviour modelling of a Cyber Guided Vehicle in M2M logic*. (Cit. on pp. 9, 13).

Braun, H. I., Jackson, D. N., & Wiley, D. E. (Eds.). (2001, August). *The Role of Constructs in Psychological and Educational Measurement*. Routledge. https://doi.org/10.4324/9781410607454 (cit. on p. 96).

Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press. (Cit. on p. 73).

Brooke, J. (2013). SUS: A Retrospective. *Journal of usability studies*, *8*(2), 29–40 (cit. on p. 73).

Bruner, J. S. (1960). *The Process of Education*. Vintage Books. (Cit. on p. 47).

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company. (Cit. on pp. 6, 15).

Büttner, S., Prilla, M., & Röcker, C. (2020). Augmented Reality Training for Industrial Assembly Work - Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376720 (cit. on pp. 2, 55, 59, 60, 62, 63).

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2 (cit. on p. 49).

Chen, C., Zhang, L., Luczak, T., Smith, E., & Burch, R. F. (2019). Using Microsoft HoloLens to improve memory recall in anatomy and physiology: A pilot study to examine the efficacy of using augmented reality in education. *Journal of Educational Technology Development and Exchange*, *12*(1). https://doi.org/10.18785/jetde.1201.02 (cit. on pp. 1, 46).

Chiarello, F., Trivelli, L., Bonaccorsi, A., & Fantoni, G. (2018). Extracting and mapping industry 4.0 technologies using wikipedia. *Computers in Industry*, *100*, 244–257. https://doi.org/10.1016/j.compind.2018.04.006 (cit. on p. 8).

Chu, C.-H., Liao, C.-J., & Lin, S.-C. (2020). Comparing Augmented Reality-Assisted Assembly Functions—A Case Study on Dougong Structure. *Applied Sciences*, *10*(10), 3383. https://doi.org/10.3390/app10103383 (cit. on pp. 54, 55, 59, 60, 62, 63).

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together*. MIT Press. (Cit. on p. 51).

Clayton, A. S. (2017). *Multiplayer educational role playing games (MPERPGs) and the application of leadership* [Doctoral dissertation, Grand Canyon University]. (Cit. on p. 47).

Clayton, A. S., & Straub, C. (2020). Virtual Leadership Simulator: The Missing Gap in Soft Skills Training. *I/ITSEC, Interservice/Industry Training, Simulation, and Education Conference 2020 Proceedings*, 13 (cit. on p. 44).

Cook, M., Lischer-Katz, Z., Hall, N., Hardesty, J., Johnson, J., McDonald, R., & Carlisle, T. (2019). Challenges and Strategies for Educational Virtual Reality. *Information Technology and Libraries*, *38*(4), 25–48. https://doi.org/10.6017/ital.v38i4.11075 (cit. on pp. 38, 46, 65).

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, *35*(6), 64–72. https://doi.org/10.1145/129888.129892 (cit. on p. 24).

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row. (Cit. on p. 48).

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2014). Flow. In *Flow and the Foundations of Positive Psychology* (pp. 227–238). Springer Netherlands. (Cit. on p. 48).

Culot, G., Nassimbeni, G., Orzes, G., & Sartor, M. (2020). Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics*, *226*, 107617. https://doi.org/10.1016/j.ijpe.2020.107617 (cit. on pp. 7, 8, 10).

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740 (cit. on p. 20).

Danielsson, O., Holm, M., & Syberfeldt, A. (2020). Augmented reality smart glasses in industrial assembly: Current status and future challenges. *Journal of Industrial Information Integration*, *20*, 100175. https://doi.org/10.1016/j.jii.2020.100175 (cit. on pp. 1, 72–74).

Dewey, J. (1938). *Experience and Education*. Macmillan. (Cit. on p. 47).

Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., & Makedon, F. (2020). A Review of Extended Reality (XR) Technologies for Manufacturing Training. *Technologies*, *8*(4), 77. https://doi.org/10.3390/technologies8040077 (cit. on p. 52).

Drath, R., & Horch, A. (2014). Industrie 4.0: Hit or Hype? [Industry Forum]. *IEEE Industrial Electronics Magazine*, *8*(2), 56–58. https://doi.org/10.1109/MIE.2014.2312079 (cit. on pp. 6, 9).

Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: Part I. *IEEE Robotics & Automation Magazine*, *13*(2), 99–110. https://doi.org/10.1109/mra.2006.1638022 (cit. on p. 36).

Ebbinghaus, H. (2013). Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, *20*(4), 155–156. https://doi.org/10.5214/ans.0972.7531.200408 (cit. on p. 96).

ETSI Augmented Reality Framework. (n.d.). (Cit. on p. 65).

Evans, P., & Annunziata, M. (2013). Industrial Internet: A European perspective. *GE Paper* (cit. on p. 7).

Ferrone, H. (2021, October). *Learning C# by Developing Games with Unity 2021: Kickstart your C# programming and Unity journey by building 3D games from scratch* (6E). Packt Publishing Ltd. (Cit. on p. 118).

Fischer, J. (2015). Visual Consistency in Augmented Reality Compositing. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 25, 32, 33, 64).

Gay-Bellile, V., Bourgeois, S., Larnaut, D., & Tamaazousti, M. (2015). Applications of Augmented Reality for the Automotive Industry. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 25, 36, 42, 63, 64).

Geng, J., Song, X., Pan, Y., Jianjun, T., Yu, L., Dongping, Z., & Yongsheng, M. (2020). A systematic design method of adaptive augmented reality work instruction for complex industrial operations. *Computers in Industry*. https://doi.org/10.1016/j.compind.2020.103229 (cit. on p. 70).

Gonzalez-Franco, M., Pizarro, R., Cermeron, J., Li, K., Thorn, J., Hutabarat, W., Tiwari, A., & Bermell-Garcia, P. (2017). Immersive Mixed Reality for Manufacturing Training. *Frontiers in Robotics and AI*, *4*. https://doi.org/10.3389/frobt.2017.00003 (cit. on pp. 53, 54, 59, 62).

Green, J. G., DeYoung, G., Wogan, M. E., Wolf, E. J., Lane, K. L., & Adler, L. A. (2019). Evidence for the Reliability and Preliminary Validity of the Adult ADHD Self-Report Scale V1.1 (ASRS V1.1) Screener in an Adolescent Community Sample. *International Journal of Methods in Psychiatric Research*, *28*(1), e1751. https://doi.org/10.1002/mpr.1751 (cit. on p. 106).

Greenwold, S. (2003). *Spatial computing* [Doctoral dissertation, Massachusetts Institute of Technology]. (Cit. on pp. 21, 50).

Grieves, M. (2011, November). *Virtually Perfect: Driving Innovative and Lean Products through Product Lifecycle Management*. Space Coast Press. (Cit. on p. 11).

Grieves, M. (2015). Digital Twin: Manufacturing Excellence through Virtual Factory Replication (cit. on pp. 15, 42).

Grieves, M., & Vickers, J. (2017, August). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (pp. 85–113). Springer International Publishing. (Cit. on pp. 11, 13–15).

Guttag, K. (2021). Magic Leap 2 (Pt. 3): Soft Edge Occlusion, a Solution for Investors and Not Users. *KGOnTech* (cit. on p. 37).

Hardy, Q. (2014). Consortium Wants Standards for Internet of Things. *The New York Times*, *27* (cit. on p. 7).

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909 (cit. on pp. 72, 104).

Havard, V., Baudry, D., Jeanne, B., Louis, A., & Savatier, X. (2021). A use case study comparing augmented reality (AR) and electronic document-based maintenance instructions considering tasks complexity and operator competency level. *Virtual Reality*, *25*(4), 999–1014. https://doi.org/10.1007/s10055-020-00493-z (cit. on pp. 56, 59, 60, 62, 63).

Heilig, M. L., & States, U. (1960). *Stereoscopic-television apparatus for individual use*. (Cit. on pp. 16, 17).

Hoover, M., Miller, J., Gilbert, S., & Winer, E. (2020). Measuring the Performance Impact of Using the Microsoft HoloLens 1 to Provide Guided Assembly Work Instructions. *Journal of Computing and Information Science in Engineering*, *20*(6). https://doi.org/10.1115/1.4046006 (cit. on pp. 56, 59, 60, 62, 63).

Hyre, A., Harris, G., Osho, J., Pantelidakis, M., Mykoniatis, K., & Liu, J. (2022). Digital twins: Representation, Replication, Reality, and Relational (4Rs). *Manufacturing Letters*, *31*, 20–23. https://doi.org/10.1016/j.mfglet.2021.12.004 (cit. on p. 14).

IEEE. (2020). IEEE 1589-2020 - Standard for Augmented Reality Learning Experience Model. *IEEE Std 1589-2020*, 1–48. https://doi.org/10.1109/IEEESTD.2020.9069498 (cit. on p. 65).

INCOSE. (2015, July). *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities* (D. D. Walden, G. J. Roedler, K. J. Forsberg, R. D. Hamelin, & T. M. Shortell, Eds.). John Wiley & Sons. (Cit. on p. 15).

International Organization for Standardization. (2018). Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. (Cit. on p. 73).

ISO. (2015). Information technology - Automatic identification and data capture techniques - QR Code bar code symbology specification. *ISO/IEC 18004:2015* (cit. on p. 35).

Jackson, S. A., & Marsh, H. W. (1996). Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, *18*(1), 17–35. https://doi.org/10.1123/jsep.18.1.17 (cit. on p. 72).

Jerald, J. (2016). *The VR Book: Human-centered Design for Virtual Reality*. Association for Computing Machinery. (Cit. on pp. 16, 18, 24, 26, 32, 39).

Jetter, J., Eimecke, J., & Rese, A. (2018). Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits. *Computers in Human Behavior*, *87*, 18–33. https://doi.org/10.1016/j.chb.2018.04.054 (cit. on pp. 71, 73, 74).

Jiang, Z., & Benbasat, I. (2007). Investigating the Influence of the Functional Mechanisms of Online Product Presentations. *Information Systems Research*, *18*(4), 454–470. https://doi.org/10.1287/isre.1070.0124 (cit. on p. 20).

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, *29*, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002 (cit. on pp. 11–13).

Kagermann, H. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0 (cit. on p. 7).

Kagermann, H., Lukas, W.-D., & Wahlster, W. (2011). Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution. *VDI nachrichten*, *13*(1), 2–3 (cit. on pp. 7, 10).

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*(4), 706–726. https://doi.org/10.1177/0018720820904229 (cit. on pp. 2, 66, 67, 97).

Kaplanyan, A. S., Sochenov, A., Leimkühler, T., Okunev, M., Goodall, T., & Rufo, G. (2019). DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics*, *38*(6), 1–13. https://doi.org/10.1145/3355089.3356557 (cit. on p. 31).

Kappes, H. B., & Morewedge, C. K. (2016). Mental Simulation as Substitute for Experience. *Social and Personality Psychology Compass*, *10*(7), 405–420. https://doi.org/10.1111/spc3.12257 (cit. on pp. 45, 50).

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3 (cit. on p. 72).

Kersten-Oertel, M., Jannin, P., & Collins, D. L. (2015). Augmented Reality for Image-Guided Surgery. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 43, 65, 70).

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele, UK, Keele University*, *33*(2004), 1–26 (cit. on p. 4).

Kiyokawa, K. (2015). Head-Mounted Display Technologies for Augmented Reality. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 17, 24, 26, 30, 32, 37, 39).

Knowles, M. S. (1970). *The Modern Practice of Adult Education: Andragogy Versus Pedagogy*. Association Press. (Cit. on p. 47).

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, *27*(2), 97–111. https://doi.org/10.1093/comjnl/27.2.97 (cit. on p. 159).

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall. (Cit. on p. 47).

Kolla, S. S. V. K., Sanchez, A., & Plapper, P. (2021). Comparing Effectiveness of Paper Based and Augmented Reality Instructions for Manual Assembly and Training Tasks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3859970 (cit. on pp. 57, 59, 60, 62, 63).

Kress, B. C. (2015). Optics for Smart Glasses, Smart Eyewear, Augmented Reality, and Virtual Reality Headsets. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 31, 41).

Kress, B. C. (2020). *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets*. SPIE. (Cit. on pp. 1, 2, 16, 18, 22, 23, 25, 26, 28, 31, 32, 37–42, 44, 64).

Kress, B. C., & Cummings, W. J. (2017). Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. *SID Symposium Digest of Technical Papers*, *48*(1), 127–131. https://doi.org/10.1002/sdtp.11586 (cit. on pp. 24, 38).

Kress, B. C., Pace, M., & Chatterjee, I. (2020). Artificial intelligence (AI) as a key enabling technology for next generation mixed reality (MR) experiences leading to mass adoption in enterprise and consumer spaces Emerging Topics in Artificial Intelligence 2020. *11469*, 1–8. https://doi.org/10.1117/12.2571473 (cit. on pp. 24, 25).

Kress, B. C., Saeedi, E., & Brac-de-la-Perriere, V. (2014). The segmentation of the HMD market: Optics for smart glasses, smart eyewear, AR and VR headsets Photonics Applications for Aviation, Aerospace, Commercial, and Harsh Environments V. *9202*, 107–120. https://doi.org/10.1117/12.2064351 (cit. on pp. 26, 40).

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018*, *51*(11), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474 (cit. on pp. 11, 14).

Kwon, C. (2018). Verification of the possibility and effectiveness of experiential learning using HMD-based immersive VR technologies. *Virtual Reality*, *23*(1), 101–118. https://doi.org/10.1007/s10055-018-0364-1 (cit. on pp. 45, 46, 48).

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*, 239–242. https://doi.org/10.1007/s12599-014-0334-4 (cit. on pp. 7, 10).

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. (Cit. on p. 47).

Laviola, E., Gattullo, M., Manghisi, V. M., Fiorentino, M., & Uva, A. E. (2021). Minimal AR: Visual asset optimization for the authoring of augmented reality work instructions in manufacturing. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-021-08449-6 (cit. on pp. 70, 73).

Lee, E. A. (2006). Cyber-Physical Systems-Are Computing Foundations Adequate ? *Position paper for NSF workshop on cyber-physical systems: research motivation, techniques and roadmap* (cit. on p. 9).

Lee, E. A. (2015). The Past, Present and Future of Cyber-Physical Systems: A Focus on Models. *Sensors*, *15*(3), 4837–4869. https://doi.org/10.3390/s150304837 (cit. on p. 9).

Lee, J., Bagheri, B., & Kao, H.-A. (2015). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, *3*, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001 (cit. on pp. 9, 10).

Leonard, S. N., & Fitzgerald, R. N. (2018). Holographic learning: A mixed reality trial of Microsoft HoloLens in an Australian secondary school. *Research in Learning Technology*, *26*(0). https://doi.org/10.25304/rlt.v26.2160 (cit. on pp. 1, 45, 46, 48, 63, 65).

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *22 140*, 55–55 (cit. on p. 105).

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x (cit. on p. 20).

Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & Industrial Engineering*, *113*, 144–159. https://doi.org/10.1016/j.cie.2017.09.016 (cit. on pp. 69, 73).

Marchand, E., Uchiyama, H., & Spindler, F. (2016). Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, *22*(12), 2633–2651. https://doi.org/10.1109/TVCG.2015.2513408 (cit. on p. 34).

Masood, T., & Egger, J. (2019). Augmented reality in support of Industry 4.0—Implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, *58*, 181–195. https://doi.org/10.1016/j.rcim.2019.02.003 (cit. on pp. 2, 71, 73, 74).

Masood, T., & Egger, J. (2020). Adopting augmented reality in the age of industrial digitalisation. *Computers in Industry*, *115*, 103112. https://doi.org/10.1016/j.compind.2019.07.002 (cit. on pp. 2, 71, 73, 74).

Microsoft Mixed Reality Toolkit. (n.d.). *MRTK2* (cit. on p. 65).

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329 (cit. on p. 16).

Millican, T. (2017). *Virtual Reality in Higher Education: A Case Study at the Air University's Squadron Officer College* [Doctoral dissertation, Auburn University]. (Cit. on pp. 44, 45, 48).

Moro, C., Birt, J., Stromberga, Z., Phelps, C., Clark, J., Glasziou, P., & Scott, A. M. (2021). Virtual and Augmented Reality Enhancements to Medical and Science Student Physiology and Anatomy Test Performance: A Systematic Review and Meta-Analysis. *Anatomical Sciences Education*, *14*(3), 368–376. https://doi.org/10.1002/ase.2049 (cit. on p. 67).

Moyo, Y. (2013, December). *Developing a Multidisciplinary Best Practice Manufacturing Education and Research Laboratory for 21st Century Competitiveness* [Doctoral dissertation, Auburn University]. (Cit. on p. 84). Accepted: 2013-12-10T15:59:41Z.

Nakajima, S. (1988). *Introduction to TPM: Total Productive Maintenance*. Productivity Press. (Cit. on p. 97).

Negri, E., Fumagalli, L., & Macchi, M. (2017). A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy*, *11*, 939–948. https://doi.org/10.1016/j.promfg.2017.07.198 (cit. on pp. 7, 11).

Ng Corrales, L. d. C., Lambán, M. P., Hernandez Korner, M. E., & Royo, J. (2020). Overall Equipment Effectiveness: Systematic Literature Review and Overview of Different Approaches. *Applied Sciences*, *10*(18), 6469. https://doi.org/10.3390/app10186469 (cit. on p. 97).

Norman, D. (2013, November). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books. (Cit. on p. 21).

Ohno, T. (1988, March). *Toyota Production System: Beyond Large-Scale Production*. CRC Press. (Cit. on p. 88).

Olson, J. L., Krum, D. M., Suma, E. A., & Bolas, M. (2011-03-19/2011-03-23). A design for a smartphone-based head mounted display. *2011 IEEE Virtual Reality Conference*, 233–234. https://doi.org/10.1109/VR.2011.5759484 (cit. on pp. 18, 19).

OpenCV Perspective-n-Point (PnP) pose computation. (n.d.). (Cit. on p. 35).

Oztemel, E., & Gursev, S. (2020). Literature Review of Industry 4.0 and Related Technologies. *Journal of Intelligent Manufacturing*, *31*, 127–182. https://doi.org/10.1007/S10845-018-1433-8 (cit. on pp. 1, 7, 42).

Palmarini, R., Erkoyuncu, J. A., & Roy, R. (2017). An Innovative Process to Select Augmented Reality (AR) Technology for Maintenance. *Procedia CIRP Proceedings of the 5th International Conference in Through-life Engineering Services Cranfield University, 1st and 2nd November 2016*, *59*, 23–28. https://doi.org/10.1016/j.procir.2016.10.001 (cit. on pp. 68, 73).

Parsons, D., & MacCallum, K. (2021). Current Perspectives on Augmented Reality in Medical Education: Applications, Affordances and Limitations. *Advances in Medical Education and Practice*, *Volume 12*, 77–91. https://doi.org/10.2147/amep.s249891 (cit. on p. 75).

Pascual, D. G., Daponte, P., & Kumar, U. (2019). *Handbook of Industry 4.0 and SMART Systems*. CRC Press. (Cit. on pp. 9, 10).

PEO Soldier PM IVAS. (n.d.). (Cit. on p. 44).

Perrow, C. (1999). *Normal accidents: Living with high risk technologies*. Princeton university press. (Cit. on p. 15).

Piaget, J. (1928). *Judgment and Reasoning in the Child*. Littlefield, Adams. (Cit. on p. 47).

Radziwon, A., Bilberg, A., Bogers, M., & Madsen, E. S. (2014). The Smart Factory: Exploring Adaptive and Flexible Manufacturing Solutions. *Procedia Engineering 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013*, *69*, 1184–1190. https://doi.org/10.1016/j.proeng.2014.03.108 (cit. on p. 8).

Roberts, M., & Paczan, N. (2021). Hypersim: A Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 10892–10902 (cit. on p. 64).

Roscoe, S. N. (1971). Incremental Transfer Effectiveness. *Human Factors*, *13*(6), 561–567. https://doi.org/10.1177/001872087101300607 (cit. on p. 97).

Saltzer, J. H. (2020). The Origin of the “MIT License”. *IEEE Annals of the History of Computing*, *42*(4), 94–98. https://doi.org/10.1109/MAHC.2020.3020234 (cit. on p. 115).

Sauro, J. (2011). *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Measuring Usability LLC. (Cit. on p. 73).

Schonberger, J. L., & Frahm, J.-M. (2016). Structure-from-Motion Revisited. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.445 (cit. on p. 36).

Schwab, K. (2015). The Fourth Industrial Revolution: What it Means and How to Respond. *Foreign Affairs* (cit. on p. 6).

Shabanov, I. (2024, March). The Effortless Blog. (Cit. on p. 4).

Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., & Wang, L. (2012). Modeling, simulation, information technology & processing roadmap. *National Aeronautics and Space Administration*, *32*(2012), 1–38 (cit. on p. 11).

Shelton, B. E., & Hedley, N. R. (2003). Exploring a Cognitive Basis for Learning Spatial Relationships with Augmented Reality. *Technology, Instruction, Cognition and Learning*, *1*(4) (cit. on pp. 1, 46, 50).

Simpson, E. J. (1966). *The Classification of Educational Objectives, Psychomotor Domain*. University of Illinois. (Cit. on p. 48).

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603 (cit. on p. 20).

Starner, T. (2015). Wearable Computing. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 22, 42).

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x (cit. on p. 20).

Sutherland, I. (1965). The Ultimate Display. *Proceedings of IFIP Congress*, *506* (cit. on p. 17).

Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5 (cit. on p. 49).

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205 (cit. on p. 49).

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. *Proceedings of the Conference on Human Factors in Computing Systems - CHI '03*. https://doi.org/10.1145/642611.642626 (cit. on pp. 53, 59, 60, 62, 63, 67).

Tao, F., & Zhang, M. (2017). Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing. *IEEE Access*, *5*, 20418–20427. https://doi.org/10.1109/ACCESS.2017.2756069 (cit. on pp. 7, 14).

Taylor, N. E. (2021). Form from Function: Applying Flow Driven Experiential Learning to the Integration of Immersive Technology in Formal Military Training Programs. *I/ITSEC* (cit. on pp. 69, 73).

UL. (2022). UL 8400 - Virtual Reality, Augmented Reality and Mixed Reality Technology Equipment. *UL CSDS Proposal 8400, Edition 1*, 1–57 (cit. on pp. 16, 26, 32, 39, 65, 132).

Vanneste, P., Huang, Y., Park, J. Y., Cornillie, F., Decloedt, B., & Van den Noortgate, W. (2020). Cognitive support for assembly operations by means of augmented reality: An exploratory study. *International Journal of Human-Computer Studies*, *143*, 102480. https://doi.org/10.1016/j.ijhcs.2020.102480 (cit. on pp. 56, 59, 60, 62, 63).

Ventura, J., & Höllerer, T. (2015). Urban Visual Modeling and Tracking. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on p. 36).

Vygotskiĭ, L. S., & Kozulin, A. (1986). *Thought and Language*. MIT Press. (Cit. on p. 47).

Wang, Z., Bai, X., Zhang, S., Billinghurst, M., He, W., Wang, Y., Han, D., Chen, G., & Li, J. (2021). The role of user-centered AR instruction in improving novice spatial cognition in a high-precision procedural task. *Advanced Engineering Informatics*, *47*, 101250. https://doi.org/10.1016/j.aei.2021.101250 (cit. on pp. 53, 57–59, 62, 63).

Weiser, M. (2002). The computer for the 21st Century. *IEEE Pervasive Computing*, *1*(1), 19–25. https://doi.org/10.1109/MPRV.2002.993141 (cit. on p. 21).

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322 (cit. on p. 49).

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686 (cit. on p. 72).

Woetzel, J., Chen, Y., Manyika, J., Roth, E., Seong, J., & Lee, J. (2015). The China effect on global innovation. *McKinsey Global Institute*, 41–102 (cit. on p. 7).

World Wide Web Consortium (W3C). (2008, November). *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. Manual. (Cit. on p. 151).

Wright, L., & Davidson, S. (2020). How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences*, *7*(1), 13. https://doi.org/10.1186/s40323-020-00147-4 (cit. on pp. 12, 13).

Yang, X., & Cheng, K. T. T. (2015). Scalable Augmented Reality on Mobile Devices: Applications, Challenges, Methods, and Software. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 33–36).

Yaniv, Z., & Linte, C. A. (2015). Applications of Augmented Reality in the Operating Room. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on p. 43).

Yim, M. Y.-C., Chu, S.-C., & Sauer, P. L. (2017). Is Augmented Reality Technology an Effective Tool for E-commerce? An Interactivity and Vividness Perspective. *Journal of Interactive Marketing*, *39*(1), 89–103. https://doi.org/10.1016/j.intmar.2017.04.001 (cit. on pp. 20, 32).

You, S., & Neumann, U. (2015). Visual Tracking for Augmented Reality in Natural Environments. In W. Barfield (Ed.), *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press. (Cit. on pp. 33, 36).

A Team Contribution Matrix

Table A.1 is meant to properly account for the various contributions made to this work. Single check-marks indicate assistance and double-checks indicate primary contributors for each category / sub-category. I1 and I2 refer to the two investigations that were included in the IRB, as described in Section 4.7.1.

Table A.1: Contributions to this work.

Category	Subcategory	DJO	RS	GH	JE	KM	DO	VB	MH	AB	DT	CT	DC	KT
IRB	Admin	✓✓	✓											
	Strategy	✓✓	✓	✓	✓			✓✓	✓					
	Write I1	✓✓						✓						
	Write I2							✓✓	✓✓					
	Review		✓	✓				✓✓	✓					
	Edit	✓✓						✓						
Experiments	Design	✓✓	✓					✓		✓	✓	✓		
	Administration	✓✓						✓✓						
	Conduct	✓✓						✓	✓	✓	✓	✓	✓	✓
	Recruitment	✓✓						✓✓						
Data	Collection	✓✓						✓	✓	✓	✓	✓		✓
	Analysis	✓✓						✓						
Software	Direct / Manage	✓✓								✓				
	Development	✓								✓✓	✓✓	✓✓		
Dissertation	Plan	✓✓	✓	✓	✓									
	Write	✓✓												
	Review		✓	✓	✓	✓	✓	✓						
	Edit	✓✓												

Legend:

- Author
  - DJO: Danny J. O'Leary
- Dissertation Committee
  - RS: Dr. Richard Sesek
  - GH: Dr. Gregory Harris

- JE: Dr. John Evans
  - KM: Dr. Konstantinos Mykoniatis
  - DO: Darren Olsen, JD

- Research Collaborators
  - VB: Victoria Ballard
  - MH: Md Monir Hossain

- Research Assistants
  - AB: Alex Barras
  - DT: David "Brown" Teague
  - CT: Carson Tillery

- Protocol Associates
  - DC: Diego Roberto Caputo Rodriguez
  - KT: Kralyn Thomas

A portion of this research was funded by the Industrial Base Analysis & Sustainment Program, Dual-Purpose Robotic 3d Scanning System for Dimensional & Welding Inspection of Large Ground Vehicles, agreement number CS-20-1601.

B Video Tool Selection

These informal notes were taken during the search for tool(s) to process and annotate video efficiently and export machine-readable metadata for those annotations.

Looking for tools to build a workflow to: - trim and combine videos (split-screen) - export a small, high quality version - mark up the combined video with observed events using markers and sub-clips - export annotations with timecode

Based on the following notes I chose a combination of Filmora, Handbrake, and Kyno.

B.1 Specialty Tools

I identified several candidates but ultimately tried only one from this list:

- Observer XT https://www.noldus.com/observer-xt - seems spendy and probably more full featured than required
- BORIS https://www.boris.unito.it/pages/features.html - looks pretty good, free and open-source, actively maintained
- Transana https://www.transana.com - affordable and full featured, may lean more towards transcripts than we need?
- ChronoViz https://chronoviz.com - free, Mac only
- ELAN https://archive.mpi.nl/tla/elan - free, open source, mult-platform; seems focused on linguistics
- Anvil http://www.anvil-software.de - not updated since 2017
- VideoTagger http://www.anvil-software.de - not updated since 2017

B.1.1 BORIS

Designed for this kind of thing but clunky and overkill for the needs of this project. Windows only (for now). Supports mult-cam playback but difficult to sync. Workspace not saved.

B.2 Traditional Video Editors

B.2.1 Adobe Premiere Pro

Steep learning curve with many unnecessary features.

B.2.2 Camtasia

Easy to set up and synchronize videos. Poor playback controls. Big / slow exports. No support for sub-clips and limited metadata. No export of markup? Project files include all video - loads of storage required.

Aside: lots of cool EdTech features - clearly their market.

B.2.3 iMovie

Too limited.

B.2.4 Filmora

Good balance of complexity and capability. Playback controls could be better but ticks the other boxes:

- easy to set up split screen and sync tracks
- reasonable export size
- fair price (but shady subscription up-charges and bad purchase UX)
- bonus: effective background noise reduction

B.3 Video Preproduction Tools

B.3.1 Adobe Prelude

No longer available.

B.3.2 Kyno

Designed for prepping clips for production. Streamlined single-track toolset for tagging and export. A sort of Lightroom for video. Great playback controls and metadata support, including a variety of export options (XML flavors, XLS). Metadata is kept in SQLite database, which complicates sharing, but the program can import XML. Does not touch source files. Lightweight. More capabilities than I need but very well suited pro tool that is reasonably priced.

B.3.3 Handbrake

Much more efficient compression than the built-in tools above.

C Video Processing

Step-by-step instructions for processing video. Starts with individual clips (left side and HL2) and ends with 3-view synced video and annotation data.

C.1 Combine Videos in Filmora

1. Rename their raw video files using the format *part number - phase - camera*.
    1. Usually, there will be only four files, named as follows.
        1. `PART#-learn-hl`
        2. `PART#-learn-side`
        3. `PART#-recall-hl`
        4. `PART#-recall-side`
    2. If the trial had to be restarted due to a crash, etc., additional files may exist. In that case, append the affected files with `-p1`, `-p2`, etc.
2. Move the raw video files to `Research Master\raw_videos\PART#` on TBM.
3. Use Filmora to build and export a combined video for the learn and recall phases. For each learn / recall phase repeat the following steps:
    1. Project settings: 16:9 widescreen, 1920 x 1080 (Full HD), 30fps, SDR 709
    2. Add the relevant raw video files to the project media folder.
    3. Drag the clips onto the timeline **one at a time** to create the following video tracks:
        1. Side camera footage, rot -90deg, ≈140% scale, positioned so that the right edge of the video snaps to the centerline of the video window, and the fixture is centered in the lower left quadrant. Crop right edge as necessary. Audio un-muted with denoise activated (mid).
        2. Side camera footage, rot -90deg, scale 90%, snap right edge to center. Crop as necessary and recheck snap to center. Mute the audio.
        3. HoloLens footage, scale 50%, snapped to the top left corner. Audio muted.
        4. Check that all panels are properly aligned with expected overlap.
    4. Save the Filmora project as *part number - Phase*, e.g. `1001-Learn`, in the participant's `videos` folder.

5. Match the action between track 2 and 3 as follows:
    1. Find an event in track 3 (HL) that can be used to match action. Something that will be in view on both recordings, e.g. the last car placed or stopping the timer.
    2. With the HL clip in track 3 selected, add a marker at that moment (M).
    3. Find the matching moment in track 2 (side) and mark it in the same fashion.
    4. Zoom in and slide the HL clip so the markers align.
6. Set the export region as follows:
    1. Deselect all clips / tracks.
    2. Move to a good starting point for the video. Click the scissors on the play head to cut all clips at that point.
    3. Repeat this process at the end point.
    4. Select a continuous track in the middle and Select Clip Range (X).
7. Preview the result before exporting.
8. Save the Filmora project.
9. With the clip range still selected, export the video to working folder (`desktop-export`) with the following settings:
    1. Preset = Match to Project Settings
    2. Format = MP4
    3. Quality = Recommend (this will change the Preset to Custom)
    4. Frame Rate = 30 fps
    5. Enable HW acceleration ON
    6. Upload to Cloud and Add to Project Media OFF
    7. Check *Use last export settings for local* option after initial setup and save these settings as a preset for other projects.
10. Scrub through the resulting video to quickly check the result.

C.2 Recompress Using Handbrake

1. Load the video in Handbrake.
2. Check that the size is 1920 x 1080.
3. Choose the preset Fast 720p30.
4. Append `-hb` to the output filename. (or change settings for output naming)
5. Compress the file.

6. Check the result.
7. Rename it and move it to the participant's `videos` folder.

C.3 Use Kyno to Annotate Events and Export Data

1. Add a title and general notes about the trial in the Metadata panel.
2. Use Markers (M) to annotate instant events. Press M once to place a marker and again to edit its name and description.
3. Use In (I) and Out (O) points to annotate events with a time duration. Press I to set the in point and O to set the out point. Then press S to define a sub-clip based on those points. Press S again to edit its name and description.
4. when annotating video, the two streams may not be perfectly synced and/or may drift apart due to dropped frames on the HL recording - use side cam for car related logging and HL cam for PWI / HL UI related logging to limit effect of lag on times recorded
5. Annotations should include:
    1. Markers
        1. Breakage - car breaks during assembly
        2. Defect - something wrong with the pre-built car
        3. Bin - wrong part found in bin
        4. Correction - participant notices an error and corrects it
        5. Removal - participant uses removal tool
        6. Slip - car slips on surface (MR - no fixture)
        7. System - HL or PAR system issue, e.g. inadvertent HL menu
        8. Crash - system not responsive, reset required
        9. Stop - stop timer to pause trial
        10. Restart - restart timer
        11. Drop - UI dropped - area tracking?
        12. Tracking - loss of model tracking
        13. PWI - for glances at paper work instructions during Recall
        14. Lean - participant leans in to get a closer look at PWI
        15. Instruction - participant asks for clarification
        16. Intervention - observer intervenes on trial
    2. Sub-Clips

1. Car build start (releases car on work surface) and stop (releases car in green or red tray), with names formatted as `Car 1`. Include anything of note for each car in its description.
2. For cars that are not completed, use the following names instead:
    1. Breakage - car was placed in the red tray for breakage
    2. Defect - car was placed on the red tray for defect in prebuilt
    3. Incomplete - ran out of time
3. For the Recall phase, PWI reference start and stop, with names formatted as `PWI 1` and notes as appropriate.
    1. **Note: do not mark PWIs during defective cars?** see 1014
4. For any interruption in video, use `Video 1` with details.
5. Mark any extended repair in the format `Repair 1` with details
6. Mark time lost to UI drops and Tracking reset as `Drop` (area) or `Tracking` (model)
7. Mark time lost to system issues as `System` (obstruction) or `Crash` (reset required)
8. Use `Reset` to indicate the segment during which the video documents the system stop and restart. For clarity only as the timer should be stopped.

6. After annotation is completed, use File > Export > Kyno XML to export it into the participant's `videos` folder.

D Data Organization

TODO: clean this up and eliminate any sensitive folder names / directory structures, etc.

D.1 File Locations

Data and analysis kept in:

- *RAID array - extracted data for each participant and original raw videos
- Local diss dir - manuscript files
- Local dev dir - local analysis files
  - *Local data dir - ingested and processed files, support files, raw notes
- *BOX folder - local sync source for BOX, duplicated from elsewhere, for team member access (ensures changes they make don't affect my own files)

*Detailed below

D.1.1 RAID

On `/Volumes/ThunderBay mini...`:

```
.../Research Master
├── data
│   ├── 1001-PWI-2023-02-10-1200
│   │   ├── 1001-Forms.pdf
│   │   ├── 1001-Script.pdf
│   │   ├── images
│   │   │   ├── 2023-02-10T1500-IMG_4829.jpg
│   │   │   └── ...
│   │   └── videos
│   │       ├── 1001-Learn.mp4
│   │       ├── 1001-Learn.wfp
│   │       ├── 1001-Learn.xml
│   │       ├── 1001-Recall.mp4
│   │       ├── 1001-Recall.wfp
│   │       └── 1001-Recall.xml
│   │
```

```
│   ├── 1003-PAR-2023-02-17-1330
│   └── ...
└── raw_videos
    ├── 1001
    │   ├── 1001-learn-hl.mp4
    │   ├── 1001-learn-side.mov
    │   ├── 1001-recall-hl.mp4
    │   └── 1001-recall-side.mov
    ├── 1003
    └── ...
```

- `Research Master\data` on the TBM RAID has all the data for each participant. When appropriate, this folder is synced to the local BOX folder using a ChronoSync action.
  - `videos` subdirectory will hold the **finished videos**, along with Kyno project and XML files
- `Research Master\raw_videos`, also on the TBM RAID has the **raw** video files, which are moved here as each participant video directory is processed. In order to reclaim space on the internal drive, these files are removed from the local BOX folder during the sync described above.

D.2 Local Development Directory

On `/Users/djo/dev/au/dissertation...`:

```
.../data
├── DataDictionary.docx
├── combined_results.xlsx
├── csv
│   ├── i1_times_v1.csv
│   └── i1_times_v2.csv
├── reports
│   ├── 1001-combined.md
│   └── ...
└── source
    ├── adjusted_drop_events.xlsx
```

```
    ├── i1_raw_data.xlsx
    └── notes
        ├── 1001.md
        └── ...
```

TODO: consolidate diss folders and rename root above (careful with chronosync jobs)

D.2.1 Box Sync Folder

On `/Users/djo/Box Sync/Tiger Motors Research Team Collaboration Files...`:

```
.../Investigation 1 Data Files
├── analysis (sync of Local dev)
│   └── ...
└── trial-data (sync of RAID / data)
    └── ...
```

- The `BOX Sync\...\trial-data` tree holds local copies of the data that get synced from each participant's `videos` folder and then mirrored to the cloud, where other members of the research team can access it.

TODO: GitHub?

D.2.2 Local Data Backup

Incremental hourly external backups with local snapshots. Nightly full external backups with local snapshots. Carbon Copy Cloner to NVMe

D.3 Synchronization / Backup

Related ChronoSync Jobs

- **trial-data** mirrors all the participant data to BOX:
  - `Research Master\data` ☐ `BOX Sync\...\trial-data`
- **analysis** mirrors all the analysis to BOX

– ...\dissertation\data □ BOX Sync\...\analysis

E Institutional Review Board Approval

The final approved version of the IRB begins on the following page.

O'Leary Protocol Review – AU IRB Protocol #22-538

The Effects of Augmented Instruction on Manufacturing Assembly Training

Version History

Version	Description	Comp Date	Sub Date	Resp Date	Status	Outcome
1.0	Original submission	12/4/22	12/7/22	12/27/22	Rejected	Add elements related to COVID protocol and risks
1.1	COVID changes	1/4/23	1/6/23	1/30/23	Approved	Initial protocol approved
2.0	Expanded protocol, 2 investigations	2/2/23	2/2/23	n/a	None	Skipped due to v1.1a, which backport changes; not sub'ed as modification
1.1a	Backport v2 changes for first invest.	2/5/23	2/6/23	2/6/23	Rejected	Not submitted as a modification, adjust format
1.1a1	Resub 1.1a as modification	2/6/23	2/7/23	2/13/23	Approved	v1.1a1 is approved, added team members and part. intake sheet form
2.1	Resub v2 with v1.1a mods inforporated	2/20/23	2/20/23	2/23/23	Approved	v2.1 is approved, expanded team again and started I2
3.0	Expanded recruitment, compensation	4/3/23	4/3/23		Submitted	in review

Thank you for the recent approval of our IRB modification v2.1, dated 2/20/23.

Enclosed please find another modification which incorporates a variety of minor changes along with expanded recruitment plan with provisions for compensation, per Dr. Sesek's conversation with Sally Headley on Friday, March 31st.

Best Regards,

The Research Team
Dan, Monir, and Victoria

Summary of Changes:

- Expanded Recruitment and Compensation, 4/3/2023
  - Protocol Review Form (PRF):
    - Dates and versions updated
    - 6B Participants will be compensated
    - 8B Added description of open house event with retention experiment
    - 9A Updated purpose to reflect intention of expanded recruitment and retention experiment
    - 12A Updated recruitment details and added description of the manufacturing industry recruitment for the first investigation
    - 12B Noted that a self-service system is available to automate the onboarding process
    - 12C Updated participant numbers, added table for clarity
    - 12D Described the plan for compensation
  - Informed Consent (IC):
    - Added description of compensation for both investigations
  - Appendices:
    - B – Recruiting Materials
      - Added a script for the manufacturing industry recruitment
      - Added a script for the open house invitation
      - Added a list describing the posters and flyers incorporated
      - Updated all flyers and slides to address compensation and add a link to the sign up page

            - Added a new flyer for the manufacturing industry recruitment
        - C – Data Collection Instruments
            - Added 6 questions to the participation intake sheet
            - Updated the Task Loading Index
    - Additional Changes
        - None.

Attached:

1. Modification form
2. Updated protocol form, consent documents, and appendix, all with changes highlighted
3. Clean versions of all updated forms that require new IRB stamps.
4. All current IRB stamped docs

AUBURN UNIVERSITY HUMAN RESEARCH PROTECTION PROGRAM (HRPP)

REQUEST for MODIFICATION

For Information or help completing this form, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu**

- **Federal regulations require IRB approval before implementing proposed changes.**
- **Change means any change, in content or form, to the protocol, consent form, or any supportive materials (such as the investigator's Brochure, questionnaires, surveys, advertisements, etc.). See Item 4 for more examples.**

1. Today's Date	4/3/2023

2. Principal Investigator (PI) Name: Dan O'Leary			
PI's Title:	Instructor / PhD Candidate	**Faculty PI (if PI is a student):**	Dr. Richard Sesek
Department:	Industrial & Systems Eng	**Department:**	Industrial & Systems Eng
Phone:	407-399-3189	**Phone:**	334-728-1438
AU-E-Mail:	djo0008@auburn.edu	**AU E-Mail:**	rfs0006@auburn.edu
Contact person who should receive copies of IRB correspondence (Optional):	Click or tap here to enter text.	**Department Head Name:**	Dr. Gregory Harris
Phone:	Click or tap here to enter text.	**Phone:**	334-844-1407
AU E-Mail:	Click or tap here to enter text.	**AU E-Mail:**	gah0015@auburn.edu

3. AU IRB Protocol Identification		
3.a. Protocol Number: 22-538		
3.b. Protocol Title: The Effects of Augmented Instruction on Manufacturing Assembly Training		
3. c. Current Status of Protocol – For active studies, check ONE box at left; provide numbers and dates where applicable		
☐	**Study has not yet begun; no data has been entered or collected**	
☒ ☐	**In progress If YES, number of data/participants entered:** 22 trials in the first investigation and 30 on the second, as of 3/27. **Is this modification request being made in conjunction with/as a result of protocol renewal?** ☐ **YES** ☒ **NO**	**Current Approval Dates** **From:** 1/30/2023
☐	**Adverse events since last review If YES, describe:** Click or tap here to enter text.	**To:** Click or tap to enter a date.
☐	**Data analysis only**	
☐	**Funding Agency and Grant Number:** Click or tap here to enter text.	**AU Funding Information:** Click or tap here to enter text.
☐	**List any other institutions and/ or AU approved studies associated with this project:** Click or tap here to enter text.	

The Auburn University Institutional Review Board has approved this Document for use from
04/07/2023 to -------------
Protocol # 22-538 EP 2301

4. Types of Change **Mark all that apply, and describe the changes in item 5**	
☐	**Change in Key Personnel** List the name(s) of personnel being added to or removed from the study and attach a copy of the CITI documentation for personnel being added to the study.
☐	**Additional Sites or Change in Sites, including AU classrooms, etc.** Attach permission forms for new sites.
☐	**Change in methods for data storage/ protection or location of data/ consent documents**
☐	**Change in project purpose or project questions**
☒	**Change in population or recruitment** Attach new or revised recruitment materials as needed; both highlighted version & clean copy for IRB approval stamp Expanded recruitment for first investigation to include volunteers from local manufacturing companies. Increased maximum number of participants for second investigation. Totals for both have increased based on progress and appetite to date.
☒	**Change in study procedure(s)** Attach new or revised consent documents as needed; both highlighted revised copy & clean copy for IRB approval stamp Added compensation as described in the attached protocol review form. Updated informed consent documents accordingly.
☒	**Change in data collection instruments/forms (surveys, data collection forms)** Attach new forms as needed; both highlighted version & clean copy for IRB approval stamp Added questions to the participant intake form and updated the NASA TLX formatting.
☐	**Other** (BUAs, DUAs, etc.) Indicate the type of change in the space below, and provide details in the Item 5.c. or 5.d. as applicable. Include a copy of all affected documents, with revisions highlighted as applicable. Click or tap here to enter text.

5. Description and Rationale
5.a. For each item marked in Question #4 describe the requested change(s) to your research protocol, and the rationale for each.
Boosted participant counts to improve the statistical power of both investigations based on available time, resources, progress, and appetite. Expanded recruitment for investigation 1 to compare student and industry results. Added compensation to incentivize performance and thank industry volunteers.
5.b. Briefly list (numbered or bulleted) **the activities that have occurred up to this point, particularly those that involved participants.**
Over 50 total trials have been completed as of 3/27. Additional trials are scheduled. All those will continue to utilize the methods and forms previously approved. This modification creates no material change in the either investigation, except for compensation, which past participants will also qualify for.
5.c. Does the requested change affect participants, such as procedures, risks, costs, benefits, etc.
No.
5.d. Attach a copy of all “IRB stamped” documents currently used. (Information letters, consent forms, flyers, etc.)
Attached.
5.e. List all revised documents and attach two copies of the revised documents – one copy which highlights the revisions and one clean copy of the revised documents for the IRB approval stamp.

Attached.

6. Signatures
Principal Investigator: ______
Faculty Advisor PI, if applicable: ______

Version Date: **4/3/2023**

Modified Forms

AUBURN UNIVERSITY INSTITUTIONAL REVIEW BOARD for RESEARCH INVOLVING HUMAN SUBJECTS

PROTOCOL REVIEW FORM
FULL BOARD or EXPEDITED REVIEW

For assistance, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu** Web Address: **http://www.auburn.edu/research/vpr/ohs**
Submit completed form and supporting materials as one PDF through the IRB Submission Page
Handwritten forms are not accepted. Where links are found hold down the control button (Ctrl) then click the link.

1. Proposed Start Date of Study:1/11/2023 Today's Date: April 3, 2023
Submission Status (Check One): ☐ New ☒ Revisions (to address IRB Review Comments)
Proposed Review Category (Check One): ☐ Full Board (greater than minimal risk) ☒ Expedited
If Expedited, Indicate Category(ies) ((Link to Expedited Category Review Sheet) 6

2. Project Title: The Effects of Augmented Instruction on Manufacturing Assembly Training

3. Principal Investigator (PI): Dan O'Leary Degree(s): BS Mech Eng, MS Eng Mgmt
Rank/Title: Graduate Student Department/School: Industrial & Systems Engineering
Role/responsibilities in this project: Organize and conduct research, perform data collection and analysis
Preferred Phone Number: 407-399-3189 AU Email: djo0008@auburn.edu

Faculty Advisor Principal Investigator (if applicable): Richard Sesek
Rank/Title: Associate Professor Department/School: Industrial & Systems Engineering
Role/responsibilities in this project: Supervise and advise the design and execution of the experiment
Preferred Phone Number: 334-728-1438 AU Email: rfs0006@auburn.edu

Department Head: Gregory Harris Department/School: Industrial & Systems Engineering
Preferred Phone Number: 334-844-1407 AU Email: gah0015@auburn.edu
Role/responsibilities in this project: Dissertation co-chair and primary project advisor

4. Funding Support: ☒ N/A ☐ Internal External Agency: n/a Pending ☐ Received ☐
For federal funding, list funding agency and grant number (if available): n/a

5. a) List any contractors, sub-contractors, and other entities associated with this project: n/a

b) List any other AU IRB approved protocols associated with this study and describe the association: n/a

c) List any other institutions associated with this study and submit a copy of their IRB approval(s): n/a

Protocol Packet Checklist

Check all applicable boxes. A completed checklist is required.

☒ **Protocol Review Form** (All required signatures included and all sections completed)
(Examples of appended documents are found on the website: https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)

☒ **CITI Training Certificates** for key personnel

☒ **Consent Form or Information Letter** and any releases (audio, video or photo) that participants will review and/or sign

☒ **Appendix A** "Reference List"

☒ **Appendix B** if e-mails, flyers, advertisements, social media posts, generalized announcements or scripts, etc., will be used to recruit participants.

☒ **Appendix C** if data collection sheets, surveys, tests, other recording instruments, interview scripts, etc. will be used for data collection. Attach documents in the order they are listed in item 13c.

Continued on Page 2

☒ **Appendix D** if they study will use a debriefing form or will include emergency plans/ procedures and medical referral lists. (A referral list may be attached to the consent document.)

☐ **Appendix E** if research is being conducted at sites other than Auburn University or in cooperation with other entities. A **permission letter** from the site/ program director must be included indicating their cooperation or involvement in the project. NOTE: If the proposed research is a multi-site project, involving investigators or participants at other academic institutions, hospitals or private research organizations, a letter of **IRB approval** from each entity is required prior to initiating the project.

☐ **Appendix F** Written evidence of approval by the host country, local IRB or institutions if research is conducted outside the United States

6. General Research Project Characteristics

6A. Research Methodology

Check all descriptions that best apply to the research methodology.

Data Source(s): ☒ New Data ☐ Existing Data

Will recorded data directly or indirectly identify participants?
☒ Yes ☐ No

Data collection will involve the use of:

☒ Educational Tests (cognitive diagnostic, aptitude, etc.)
☒ Interview
☒ Observation
☐ Locations or Tracking Measures
☐ Physical / Physiological Measures or Specimens
☒ Surveys / Questionnaires
☐ Other: Click or tap here to enter text.

☒ Internet / Electronic
☒ Audio
☒ Video
☒ Photos
☐ Digital Images
☐ Private records or files

6B. Participant Information

Check all descriptors that apply to the TARGET population.
(link to definition of target population)
☐ Males ☐ Females ☐ AU students

Vulnerable Populations
☐ Pregnant Women/Fetuses ☐ Prisoners ☐ Institutionalized
☐ Children and / or Adolescents (under age 18 in AL; if minor participants, at least 2 adults must be present during all research procedures that include the minors)

Persons with:
☐ Economic Disadvantages ☐ Physical Disabilities
☐ Educational Disadvantages ☐ Intellectual Disabilities

Will participants be compensated? ☒ Yes ☐ No

6C. Risks to Participants

Identify all risks participants might encounter in this research.

☒ Breach of Confidentiality* ☐ Coercion
☐ Deception ☐ Physical
☐ Psychological ☐ Social
☐ None
☒ Other (COVID-19, other medical):
COVID-19 Exposure; Discomfort, including possibility of mild nausea, see section 14

*Note that if the investigator is using or accessing confidential or identifiable data, reach of confidentiality is always a risk.

6D. Corresponding Approval/ Oversight

- **Does the study include participant exposure to radiation?** ☐ **Yes** ☒ **No**
  If yes indicate: ☐ **DEXA** ☐ **PQCT** ☐ **Other**

- **Is IBC Approval required for this study?**
  ☐ **Yes** ☒ **No**

  If yes, BUA # Click or tap here to enter text. **Expiration Date** Click or tap to enter a date.

- **Is IACUC Approval required for this study?**
  ☐ **Yes** ☒ **No**

  If yes, PRN # Click or tap here to enter text. **Expiration Date** Click or tap to enter a date.

- **Does this study involve the Auburn University MRI Center?**
  ☐ **Yes** ☒ **No**

Which MRI(s) will be used for this project? (Check all that apply)
☐ **3T** ☐ **7T**

Continued on Page 3

Does any portion of this project require review by the MRI Safety Advisory Council?
☐ **Yes** ☒ **No**

Signature of one MRI Center Representative:____
Required for all projects involving the AU MRI Center
Appropriate MRI Center Representatives:
Dr. Thomas S. Denney, Director AU MRI Center
Dr. Ron Beyers, MR Safety Officer

7. Project Assurances

7A. Principal Investigator's Assurances

1. I certify that all information provided in this application is complete and correct.
2. I understand that, as Principal Investigator, I have ultimate responsibility for the conduct of this study, the ethical performance this project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the Auburn University IRB.
3. I certify that all individuals involved with the conduct of this project are qualified to carry out their specified roles and responsibilities and are in compliance with Auburn University policies regarding the collection and analysis of the research data.
4. I agree to comply with all Auburn policies and procedures, as well as with all applicable federal, state, and local laws regarding the protection of human subjects, including, but not limited to the following:
   a. Conducting the project by qualified personnel according to the approved protocol
   b. Implementing no changes in the approved protocol or consent form without prior approval from the Office of Research Compliance
   c. Obtaining the legally effective informed consent from each participant or their legally responsible representative prior to their participation in this project using only the currently approved, stamped consent form
   d. Promptly reporting significant adverse events and / or effects to the Office of Research Compliance in writing within 5 working days of the occurrence.
5. If I will be unavailable to direct this research personally, I will arrange for a co-investigator to assume direct responsibility in my absence. This person has not been named as co-investigator in this application, or I will advise ORC, by letter, in advance of such arrangements.
6. I agree to conduct this study only during the period approved by the Auburn University IRB.
7. I will prepare and submit a renewal request and supply all supporting documents to the Office of Research Compliance before the approval period has expired if it is necessary to continue the research project beyond the time period approved by the Auburn University IRB.
8. I will prepare and submit a final report upon completion of this research project.

My signature indicates I have read, understand and agree to conduct this research project in accordance with the assurances listed above.

____Dan O'Leary_________________	_______________________________	_2/20/2023_
Principal Investigator Name	**Principal Investigator Signature**	**Date**

7B. Faculty Advisor / Sponsor's Assurances

1. I have read the protocol submitted for this project for content, clarity, and methodology.
2. By my signature as faculty advisor / sponsor on this research application, I certify that the student or guest investigator is knowledgeable about the regulations and policies governing research with human subjects and has sufficient training and experience to conduct this particular study in accord with the approved protocol.
3. I agree to meet with the investigator on a regular basis to monitor study progress. Should problems arise during the course of the study, I agree to be available, personally, to supervise the investigator in solving them.
4. I assure that the investigator will promptly report significant incidents and / or adverse events and / or effects to the ORC in writing within 5 working days of the occurrence.
5. If I will be unavailable, I will arrange for an alternate faculty sponsor to assume responsibility during my absence, and I will advise the ORC by letter of such arrangements. <u>If the investigator is unable to fulfill requirements for submission of renewals, modifications or the final report, I will assume that responsibility.</u>

Revised 07/12/2022

____Richard Sesek________________ ________________________ 4/3/23

Faculty Advisor / Sponsor Name **Faculty Advisor Signature** **Date**

Continued on Page 4

7C. Department Head's Assurance

By my signature as department head, I certify that I will cooperate with the administration in the application and enforcement of all Auburn University policies and procedures, as well as all applicable federal, state, and local laws regarding the protection and ethical treatment of human participants by researchers in my department

____Gregory Harris_______________ ________________________ ______________

Department Head Name **Department Head Signature** **Date**

8. Project Overview:

8A. A summary of relevant research findings leading to this research proposal:
(Cite source; include a "Reference List" as ***Appendix A****.)*

This experiment incorporates two separate but related investigations that employ similar methods. Each is described separately in sections of the proposal, as required. Where no distinction is made, the protocol is identical.

First Investigation

Augmented Reality (AR) systems "combine real and virtual, are interactive in real time, and are registered in 3-D" [1]. By realistically integrating informative and/or interactive virtual objects in our view of the world, AR aims to enhance the users' interaction with and perception of it. Its essential affordance is the direct and natural manipulation of virtual objects in everyday surroundings. Relative to metaphorical digital interfaces, this is thought to improve the uptake of knowledge by reducing the overall cognitive load and better distributing it across multiple sensory pathways [2]. AR-assisted learners demonstrate improved perception, performance, and understanding of spatial concepts, with outcomes correlated to the amount of physical engagement involved [3]. As a result, AR is thought to be well-suited for task-related learning. Using untethered, hands-free devices with optical see-through head-mounted displays, AR can continuously enhance the user's actions in the real world [4]. These benefits have broad industrial applications.

In manufacturing, operator support has been a common application of AR research and development since the early 1990s [5]. It is also seen as a source of innovative operator training methods required to meet rapidly increasing demand for skilled labor due to high retirement rates, global expansion, and increasing specialization [6]. Manufacturing support, training, and related applications have been identified in the areas of assembly, maintenance, operations, quality control, safety, design, visualization, logistics, and marketing [7].

Despite great potential, the adoption of AR is slowed by technical, market, and other important social and legal obstacles [8]. To successfully transition from research projects and proof of concepts and gain widespread adoption in manufacturing, AR must demonstrate a worthwhile return on investment [9; 10]. But AR remains a highly fragmented market, including a diverse selection of screen-based, projected, and head-mounted technologies [6]. Studies show that the efficacy of these systems varies with the task type, technology used, application design, and other factors [11]. Thus, the success rate of AR adoption in industry would be improved by frameworks for strategic decision making based on quantified benefits in various scenarios [12–14]. Research in this area is young but accelerating. Most of it focuses on efficiency (task time) and accuracy (error count). These are relevant but incomplete measures for assessing training outcomes, where the learning rate and transfer effectiveness must also be considered [15]. This investigation extends prior work [16] to explore the relationship between a variety of AR technologies and their underlying affordances [17] and learning outcomes for manufacturing assembly operations. By controlling for the task type and application design we hope to better understand the relative value of these systems, filling in important gaps that can lead to a cohesive framework for successful adoption.

Second Investigation

Cognitive load is challenging to measure but is essential in designing systems for worker safety, reliability, quality, and health [1]. Typically, cognitive load assessment is estimated by directly querying subjects using survey instruments such as the NASA Task Load Index (TLX) [1]. The NASA-TLX is perhaps the most widely used such instrument, having been adapted for use in many fields during its almost 40 years in application [2], [3].

In the National Occupational Research Agenda for Healthy Work Design and Well-Being (January 2020), the first objective is: "Identify and examine the impact of worker demographics on employer or organizational practices and worker safety, health, and well-being." [4] This is because worker characteristics can adversely affect these areas. I propose to include information on the participant's status of disability, particularly ADHD, to gather data on how the presence of this disability affects the worker's performance in a manufacturing setting. The goal is to gather data to develop best practices for workers with particular needs to increase worker safety, health, and overall well-being. ADHD is an optimal place to start with this type of investigation because of its high prevalence in society, an estimated 11% adults have some level of ADHD [5]. A study of days lost and safety incidents have also shown that adults with ADHD are twice as likely to have a safety mishap in the workplace [6]. This statistic highlights the importance of gathering more information on the mental load of persons with ADHD in the manufacturing environment and investigating ways to design the workplace to accommodate their specific needs. Reliable assessment of the presence of ADHD symptoms is effectively obtained through use of the World Health Organization Adult ADHD Self-Report Scale, which we have incorporated into our exit survey in both investigations [7].

Since the early 1990s, Lean Production (LP) has been widely used in manufacturing because of its effectiveness and efficiency in waste reduction, lead time shortening, and productivity improvement [8]. LP primarily focuses on standardizing work, reducing the non-value-added activities, shifting the production systems from capacity to demand-oriented, and installing a distributed production improvement system with closed loops between workstations [8]–[10]. Meanwhile, Industry 4.0 (I-4.0) technologies have also been diffused in manufacturing in the last decade, enabling cyber-physical systems, the Internet of Things (IoT), Augmented Reality (AR), Sensor technology, and others [{Citation}].

As both Lean and I-4.0 paradigms are being used in the manufacturing world simultaneously, a question raised by manufacturers: Is there any complementary effect of I-4.0 on the performance improvement of LP systems? We conducted a literature review to find the answer to the abovementioned question. In several studies, authors revealed a significant co-relationship between Lean & I-4 [5-8]. In a study, the authors mentioned that LP could be integrated with I-4.0 technologies to meet customers' changing demand [11].

In most cases, the complementary effects of Lean and I-4.0 are conceptual. For instance, the authors suggested integrating I-4.0 technology in the LP system to overcome some limitations of Lean [12], while they did not specify strategies [13]. It is stated that LP can be considered a pre-requisite for the I-4.0 application [14], but it is not demonstrated how Lean and I-4.0 co-exist together. Several authors also acknowledge that direction on how Lean and I-4.0 work together are immature [8], [15], [16].

The literature review revealed a gap in the current body of knowledge. The gap is a lack of empirical studies of the interaction between Lean and I-4.0. To answer this need, we plan to conduct an experimental investigation of the interaction between Lean and I-4.0.

8B. A brief summary/abstract of the study methodology, including design, population, and variables of interest. (350 word maximum, in language understandable to someone who is not familiar with your area of study. Note this summary/abstract can be used to prepare the concise summary in the consent document.):

Figure 1- LEGO Speedster Assembly

This experiment will be conducted in the Tiger Motors Lean Education Center, which simulates automotive manufacturing best practices using LEGO® cars. Participants will act as operators assembling the SUV (Model T) car at stations 8 and 10. This process has been used thousands of times in INSY 5/6800 without significant incident.

Figure 2 - Work Station 8

Figure 3- Work Station 10

First Investigation

A population of 40-60 adults will be recruited from Auburn University. Candidates with experience using head-mounted or projected AR or building cars in the Lean Lab will be excluded. Participants in this between-groups design will experience a single level of the Instructional Media Type (IMT) treatment, with increasingly augmented work instructions:

1. Paper Work Instructions (PWI): traditional printed instructions (control)
2. Projector Augmented Reality (PAR): interactive instructions projected on the work surface via the LightGuide system with a stationary model
3. Head-Mounted Display AR (HMDAR): interactive instructions presented in the user's field of view using the HoloLens2 (HL2) HMD with a stationary model
4. HMD Mixed Reality (HMDMR): extends the third treatment by leveraging advanced capabilities of the HL2, allowing for more natural interactions and movement of the model

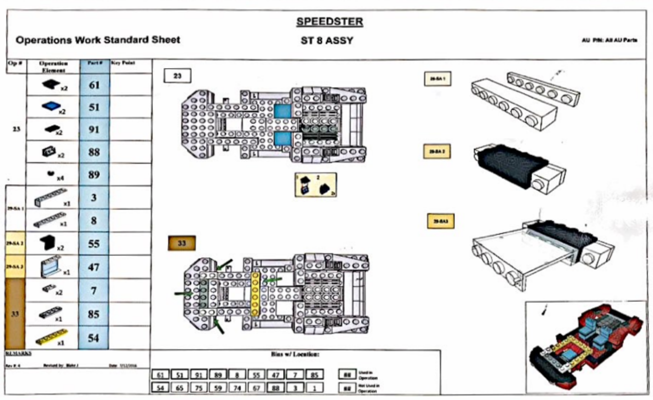


Figure 4 – Paper Work Instructions for Station 8

Figure 5- LightGuide Work Instructions

Participant groups will be set randomly. We hypothesize that HDMR will outperform other treatments in accuracy-based performance measures, as well as learning rate and transfer. In contrast, we expect participants assigned the PWI treatment to have the best times.

Figure 6 – HoloLens2 Wireless, See-Through Design

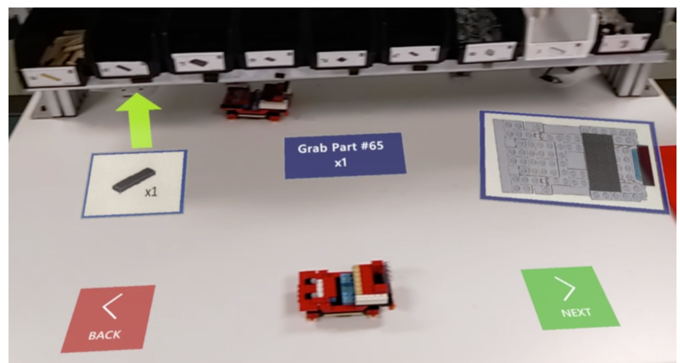


Figure 7- HoloLens2 Work Instructions, 1st Person View

First, participants will be shown how to interpret paper work instructions and use them to construct a sample LEGO assembly. Next, those assigned to an AR treatment level are given a brief introduction to its operation. Questions are allowed throughout this process.

The hypotheses are then tested in two phases. The first compares the effects of instructional media on the speed (task completion time) and accuracy (number and type of corrected and uncorrected errors) with which participants perform each repetition of the task. These measures are tracked for each assembly completed in the 10-minute session, allowing us to assess learning rates.

During the second phase, participants repeat the task four times in the control condition while the same measures are observed. Their results in each phase will be analyzed to compare transfer effectiveness between treatments.

Second Investigation

A population of 30-40 adults will be recruited from Auburn University. Participant treatment order will be set randomly. Participants in this within-subjects design will experience one of four scenarios in a random order:

1. Control: Paper Work Instructions (PWI): traditional printed instructions.
2. Lean Tool: Pre-made finished car provided for quality checks.
3. I-4.0 Tool: Inspection camera for quality check.
4. Lean + I-4.0 Tools: Pre-made finished car and inspection camera.

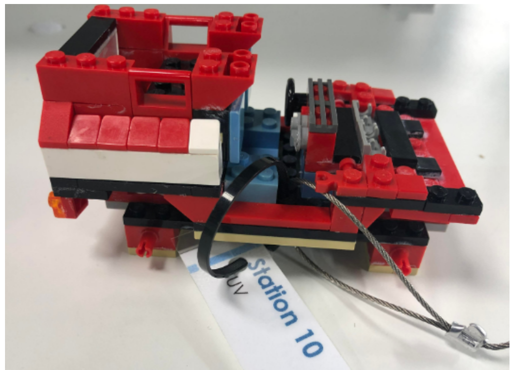


Figure 8- Pre-Made Finished Car Provided for Quality Checks

Before beginning any of the treatments, participants will be shown how to interpret the paper work instructions and use them to construct a sample LEGO assembly. Participants will practice the assembly five times. Questions are allowed throughout this process.

We hypothesize that treatment four will outperform other treatments in accuracy-based performance measures. In contrast, we expect treatment two to have the best times.

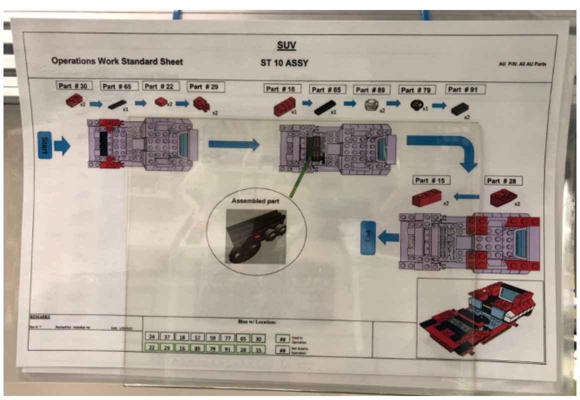


Figure 9 – Paper Work Instructions for Station 10

Figure 10 – Working Station 10

Retention Experiment

All participants recruited from the Auburn University community by either investigation will be invited to an open house event at the end of the Spring 2023 semester. The purpose of this event is thank our participants, give them the opportunity to experience other treatments and demos of the hardware, and to get additional data about how well they retained what was learned in the prior trials. Attendees that agree to the retention test will repeat the control treatment from one or both prior investigations. The same methods and metrics will apply.

9. Purpose

9A. State the purpose of the study and all research questions or aims. (Include a sentence that begins, "The purpose of this study is…")

First Investigation

The purpose of this study is to measure the effect of instructional media type (IMT) on learning rates and skills transfer for industrial assembly tasks. The first phase will help us understand how each IMT affects the operator's learning rate (time or cycles to learn the process) and ultimate measures of performance (speed and accuracy). The second will help assess how learning transfer varies with each treatment. Finally, the exit surveys will help us understand the relationship between those results and perceived workload, system usability, and the participant's self-reported behavioral control. Additionally, the industry recruitment will allow us to compare the results between our initial convenience sample and a more representative population to see how well the results generalize.

Second Investigation

The purpose of this study is to measure the effect of Lean and I-4.0 Tools on process performance and quality.

Research questions:

- Does the interaction between Lean & I-4.0 tools significantly impact the operator Performance?
- Are there significant differences in the performance, cognitive load, and usability scales for participants with few self-reported behavioral control symptoms or many?

Additionally, as with the first investigation, the exit surveys will help us understand the relationship between these results and perceived workload, system usability, and the participant's self-reported behavioral control.

Retention Experiment

The purpose of this phase of the study is to see how much of their prior experience is retained, and how that relates to the time elapsed and original methods used.

9B. Describe how results of this study will be used? (e.g., presentation? publication? thesis? dissertation?)
The data collected during this project will be used for thesis and dissertations, scholarly publications and presentations, and grant proposals.

10. Key Personnel. Describe responsibilities as specifically as possible. Include information on research training or certifications related to this project. **To determine key personnel see decision tree at https://cws.auburn.edu/OVPR/pm/compliance/irb/training. Submit a copy of CITI training documentation for all key personnel.** (For additional personnel, add lines as needed).

To determine Auburn University HIPAA – covered entities click link to HIPAA Policy.

If any key personnel have a formal association with institutions/entities involved in the study (for example is an employee or supervisor at the site research will occur), describe that affiliation. For all non-AU affiliated key personnel, submit a copy of their IRB approval.

Principal Investigator: Dan O'Leary **Rank/Title:** Graduate Student
Email Address: djo0008@auburn.edu **Degree(s):** BS ME, MS Eng Mgmt
Dept / Affiliation: Industrial & Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Overall responsibility for the project, including design and administration of experiments, coordinating recruitment, obtaining consent, and data collection and analysis.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 8/26/2025

Individual: Richard Sesek **Rank/Title:** Associate Professor
Email Address: rfs0006@auburn.edu **Degree(s):** BS, MS, MPH, PhD
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Advise, oversee, and assist with experiment design, IRB review process, obtaining consent, conducting trials, data collection and analysis.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a

Revised 07/12/2022

- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 4/25/2023

Choose a course Expiration Date

Individual: Gregory Harris **Rank/Title:** Associate Professor
Email Address: gah0015@auburn.edu **Degree(s):** PhD
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: Dissertation co-chair and primary advisor

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 5/12/2024

Choose a course Expiration Date

Individual: Gregory Purdy **Rank/Title:** Assistant Professor
Email Address: greg.purdy@auburn.edu **Degree(s):** PhD
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: primary advisor for second investigation

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 2/1/2026

Choose a course Expiration Date

Individual: Victoria Ballard **Rank/Title:** Graduate Student
Email Address: vzb0024@auburn.edu **Degree(s):** BS CHE, MS CivE
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: Lab manager, design and conduct research

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 2/9/2025

Choose a course Expiration Date

Revised 07/12/2022

Individual: Md Monir Hossain **Rank/Title:** Graduate Student
Email Address: mzh0116@auburn.edu **Degree(s):**BS BE, MS TM, MS ISE
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Lab assistant, design and conduct research
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 8/29/2025

Choose a course Expiration Date

Individual: Diego Roberto Caputo Rodriguez **Rank/Title:** Graduate Student
Email Address: drc0040@auburn.edu **Degree(s):**BS IE, MEM
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Lab assistant, assists conducting research
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 10/6/2025

Choose a course Expiration Date

Individual: Yuqing "Lucie" Wang **Rank/Title:** Graduate Student
Email Address: yzw0155@auburn.edu **Degree(s):**BS Geo, MS IE
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Lab assistant, assists conducting research
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 11/30/2025

Individual: Yen-Ting Guo **Rank/Title:** Graduate Student
Email Address: yzg0069@auburn.edu **Degree(s):**BS IE, MS IE
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Lab assistant, assists conducting research
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No

Revised 07/12/2022

- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 9/10/2025

Individual: Alex Barras **Rank/Title:** Other (Undergraduate RA)
Email Address: jab0217@auburn.edu **Degree(s):** BS CS/SWE, Spr23
Dept. / Affiliation: Computer Science & Software Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Assist with administration of protocol

- **-** AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 1/13/2026

Choose a course Expiration Date

Individual: David "Brown" Teague **Rank/Title:** Other (Undergraduate RA)
Email Address: dbt0013@auburn.edu **Degree(s):** BS CS/SWE, Spr23
Dept. / Affiliation: Computer Science & Software Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Assist with administration of protocol

- **-** AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 1/16/2026

Choose a course Expiration Date

Individual: Carson Tillery **Rank/Title:** Other (Undergraduate RA)
Email Address: cwt0013@auburn.edu **Degree(s):** BS CS/SWE, Spr23
Dept. / Affiliation: Computer Science & Software Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Assist with administration of protocol

- **-** AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 1/15/2026

Choose a course Expiration Date

Individual: Kralyn Thomas **Rank/Title:** Other (Undergraduate RA)
Email Address: kzt0044@auburn.edu **Degree(s):**n/a
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Assist with administration of protocol

Revised 07/12/2022
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: AU Basic RCR Training 2/13/2026

11. Location of research.

11A. List all locations where data collection will occur. If applicable, attach permission letters as Appendix E. (School systems,
organizations, businesses, buildings and room numbers, servers for web surveys, etc.) **Be as specific as possible.** (See sample letters at https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)
Data collection will take place at the Lean Lab in the basement of the Shelby Center for Engineering Technology, room 0317, located at 345 W Magnolia Ave, Auburn, AL 36849

11B. Will study data be stored within a HIPAA covered facility? Yes ☐ No ☒
If yes, which facility(ies) (To determine AU HIPPA covered entities, go to VII of the HIPPA Hybrid Entity Policy)**:**
n/a

12. Participants (If minor participants, at least 2 adults must be present during all research procedures that include the minors.)

12A. Describe the targeted/ intended participant population for the study. Include the anticipated number of participants and inclusion and exclusion criteria and the procedures to ensure more than 1 adult is present during all research procedures which include the minor.

☐ **Check here if existing data will be used and describe the population from whom data was collected including the number of data files.**
☐ **Check here if permission to access existing data is required and submit a copy of the agreement to access.**

For both investigations a total of between 90 and 150 subjects will be recruited from the Auburn University community. Between 40 and 70 of those will participate in the first investigation, and 50 to 80 in the second. Potential participants in the first investigation will be screened for exclusion based on the following: 1. Under 18 years of age 2. Prone to motion sickness 3. Prior experience with head-mounted or projected AR systems 4. Prior experience building cars in the Lean Lab as part of INSY 5800/6800 or otherwise. Note that third item does not exclude those having experience with Virtual Reality headsets like the Occulus Rift, which are much more commonly available than AR devices. For the second investigation, any volunteer 18 or older will qualify. A shared screening form will be used for both investigations, and candidates will be assigned to one or both investigation(s) accordingly. Active recruiting efforts for I1 will focus on freshman and sophomore engineering students in Industrial & Systems Engineering (ISE), as they are accessible and are likely to meet all requirements. I2 will actively recruit more broadly because more adults are eligible.

In addition to the participants described above, Investigation 1 will recruit 20-30 participants from local manufacturing companies. The same requirements, recruiting, consent, and onboarding procedures will apply, and the experimental protocol is unchanged, except that those participants will be compensated.

12B. Describe, step-by-step in lay language all procedures to recruit participants. Include in **Appendix B** a copy of all e-mails, flyers, advertisements, recruiting scripts, invitations, etc., that will be used to invite people to participate. (See sample documents at https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)

Students and Faculty will be recruited using flyers distributed around the Auburn University campus. Additionally, ISE students will be recruited via in-class announcements and the distribution of emails. Copies of each are included in Appendix B. Interested participants will be instructed to contact the PI for more information. In the call that follows, the PI will: 1. Briefly explain the investigation, recapping and elaborating on the recruiting materials 2. Explain the exclusion criteria and identify relevant issues for the candidate 3. Set expectations for participant involvement, including time commitment and tasks 4. Answer any questions the candidate has regarding participation in the investigation If the candidate is ready and willing to proceed, their information will be collected using the Subject Recruitment Data Sheet provided in Appendix C. They will be assigned a unique participant ID, the investigation(s) most appropriate for their exclusions, and a date and time for data collection. If interest in either investigation exceeds capacity, additional participants will be thanked for their interest and informed that enrollment is limited. They will be given the option to remain "waitlisted" if additional participants or follow-up studies are required.

As an alternative to the manual method described above, participants are able to use a self-service web-based sign-up system which automates the process.

12C. Minimum number of participants required to validate the study? see table, below

Number of participants expected to enroll? see table, below

Provide the rationale for the number of participants. Appropriate for the desired power given the number of treatments and expected differences in outcomes.

Is there a limit to the number of participants that will be included in the study?
☐ **No** ☒ **Yes, the number is** 180 in total

Participant Summary

- I1 is increased based on interest and progress to date, plus the added industry recruitment.
- I2 is increased based on interest and progress to date, and the addition of possible incentives.

		I1 - AR	I2 - Lean / I4	Total*
Original (v2.1)	Minimum	30-40	30-40	60-80
	Expected	35-50	35-50	70-100
	Limit	50	50	100
Expanded (v3.0)	Minimum	60-70	50-60	110-130
	Expected	70-90	55-80	125-170
	Limit	100	80	180

*ignoring duplication - subjects can participate in both studies, so this is total trials not total unique participants

12D. Describe the process to compensate, amount and method of compensation and/or incentives for participants. **AU Procurement and Business Services (PBS) policies**
(benefits to participants are NOT compensation)

If participants will not be compensated, check here: ☐
Indicate the amount of compensation per procedure and in total: see below
Indicate the type of compensation: ☒ Monetary ☒ Incentives
☒ Raffle or Drawing incentive (Include the chances of winning.)
☒ Extra Credit (State the value)
☐ Other

Describe how compensation will be distributed (USPS, email, etc.): in person or via email, whichever is easier considering the necessary procedures and source of funds, while ensuring confidentiality

First Investigation

All subjects recruited from the local manufacturing industry will be offered gift cards in exchange for their participation. This is considered important to the success of that recruitment. All industry volunteers will be given $40 in cash / gift card at the completion of their participation.

All subjects recruited from the Auburn University community will be considered for two different prizes. Anyone that participated is entered into the first drawing. Three winners will be randomly selected to receive $25 in cash / gift card. The chances of winning one of these prizes is approximately 3/60 (5%). In a second drawing, one winner will be randomly selected from all participants that met or exceeded a predefined quality level. That winner will receive $100 in cash / gift card. The chances of winning this prize will depend on the number of qualifying participants. We estimate 1/10 to 1/20 (5-10%). Volunteers that participated before the addition of compensation are also eligible.

Second Investigation

All subjects will be considered for two different prizes. Anyone that participated is entered into the first drawing. Three winners will be randomly selected to receive $25 in cash / gift card. The chances of winning one of these prizes is approximately 3/67.5 (4.4%). In a second drawing, one winner will be randomly selected from all participants that met or exceeded a predefined productivity level. That winner will receive $100 in cash / gift card. The chances of winning this prize will depend on the number of qualifying participants. We estimate 1/10 to 1/20 (5-10%). Volunteers that participated before the addition of compensation are also eligible.

Open House Event

In order to incentivize participation during the open house, two additional awards will be offered.

1. All attendees that participate in the retention experiment for either investigation are entered into a drawing. One entry is granted for each experiment they complete (max 2). Three winners will be randomly selected for each experiment to receive a $25 gift card (total of 6 winners). The chances of winning will depend on the number of entries, but we anticipate odds of approximately 3/25 (12%).
2. The attendee(s) with the best performance in each retention experiment will receive a $50 gift card. Random selection will be used to break any ties. Otherwise, this is a performance based prize, so odds of winning are not applicable.

Additionally, all attendees will be offered the opportunity to experience the other treatments and various demonstrations of the hardware from both investigations. Food and drink will also be offered. Manufacturing Volunteers from the first investigation will not be invited to this event.

General

Winnings are capped at $100 for any participant, which simplifies the relevant procedures. All drawings will be held at the conclusion of the open house event. Winners do not have to be present at the time of the drawing to receive their prize.

No members of the research team are eligible for any of the financial compensation described.

A total of $1800 is budgeted for all described compensation, as summarized in the table below:

	Manuf Volunteers			Qual / Prod Prize			Participation Drawing			Retention Prize			Retention Drawing			
	Qty	Amt	Cost	Qty	Amt	Cost	Qty	Amt	Cost	Qty	Amt	Cost	Qty	Amt	Cost	Total
Invest 1	30	$ 40	$1,200	1	$ 100	$ 100	3	$ 25	$ 75	1	$ 50	$ 50	3	$ 25	$ 75	$1,500
Invest 2	0	$ -	$ -	1	$ 100	$ 100	3	$ 25	$ 75	1	$ 50	$ 50	3	$ 25	$ 75	$ 300
			$1,200			$ 200			$ 150			$ 100			$ 150	**$1,800**

Extra Credit

Any instructor promoting these studies to their students is free to provide extra credit for participation. This is entirely at the discretion of each individual instructor. We encourage no more than 1% on the final class average per investigation. Alternative bonuses should be provided for those unable to participate.

13. Project Design & Methods

13A. Describe, <u>step-by-step</u>, all procedures and methods that will be used to <u>consent</u> participants. If a waiver is being requested, indicate the waiver, and describe how the study meets the criteria for the waiver. If minors will be enrolled describe the process to obtain parental/ legally authorized guardian permission.

☐ **Waiver of Consent (including using existing data)**
☐ **Waiver of Documentation of Consent (use of Information Letter)**
☐ **Waiver of Parental Permission (for college students 18 years or younger)**

As each participant arrives, they will be welcomed and given brief introductions to members of the team administering the study. We will then ask them to review the consent document, encouraging them to ask any questions they have. After a verbal confirmation that the participant has read and is satisfied with the terms of this document, we will ask that they sign and date it.

13B. In lay language, understandable by someone not familiar with the area of study, describe the complete research design and methods that will be used to address the purpose. Include a clear description of who, when, where and how data will be collected. Include specific information about participants' time and effort.

First Investigation

Following the recruitment, eligibility screening, and consent processes described above, participants are asked to provide basic demographic information, read the NASA TLX instruction sheet, and complete a Behavior Control Survey based on the Adult ADHD Self-Report Scale (ASRSv1.1). Finally, emergency procedures are described, and the participant is given the opportunity to use the restroom. Once the intake process is complete, the participant is ushered to work station 8 where a short orientation process acclimates them to the work area. A research associate will point out the key features of a work cell (work surface, part bins, etc.), describe how to interpret the paperwork instructions, demonstrate typical assembly steps, and answer any relevant questions. (5-10 mins)

Next, participants assigned to any AR IMT (PAR, HMDAR, or HMDMR) will receive a brief demonstration of its basic operation. In all cases, the participant will be shown how to use the appropriate forward and back triggers, and how the system signals instructions and feedback related to part bin and placement. PAR and HMDAR users will be instructed that the model must remain in the fixture. HMDMR users will understand that the model can be freely manipulated during assembly. (5-10 mins)

Once orientation and training are complete, the experiment is conducted in two phases. Regardless of IMT assigned, all participants will wear the HL2 during both phases to control for its effects and allow us to record each session from their POV.

In the first phase, participants will be asked to complete the assembly process for as many cars as they can, while learning the steps and limiting the number of errors produced. This phase will be conducted with the support of the assigned IMT and will last 10 minutes. Observations will be recorded on Data Collection Sheet #1. During that time, we expect that each participant will produce between 3 and 6 cars, based on prior performance data and the 60-second takt time for which the instructions were designed. (10 mins)

During a short break to reset the workstation, the participant will complete the NASA TLX and System Usability surveys for the assigned treatment. In the second phase each participant will build 4 more cars using only paper work instructions. Their stated goal will be to deliver error-free results quickly, while referencing the instructions only when necessary. Observations will be recorded on Data Collection Sheet #2. (5-10 mins)

Participant performance in both phases will be recorded on two cameras, one first-person view from onboard the HL2, and one third-person view from a camera mounted nearby. Experimental data will be derived from subsequent analysis of these videos. Participants will not be allowed to ask questions during either data collection phase of the experiment.

Once the experiment is concluded, each participant will complete an exit survey that incorporates the NASA TLX and System Usability Scale instruments for PWI. When the surveys are completed a research associate will solicit any additional general feedback, ask if the participant experienced any injury or discomfort, and invite them to attend a follow-up session for more in-depth exploration of the HoloLens2. Their responses will be recorded on the exit survey. (5-10 mins)

We conservatively estimate a total time commitment of 45-60 minutes for each participant.

Second Investigation

Following the recruitment, eligibility screening, and consent processes described above, participants are asked to provide basic demographic information, read the NASA TLX instruction sheet, and complete a Behavior Control Survey based on the Adult ADHD Self-Report Scale (ASRSv1.1). Finally, emergency procedures are described, and the participant is given the opportunity to use the restroom. Once the intake process is complete, the participant is ushered to work station 10 where a short orientation process acclimates them to the work area and emergency procedures are described. A research associate will point out the key features of a work cell (work surface, part bins, etc.), describe how to interpret the paperwork instructions, demonstrate typical assembly steps, and answer any relevant questions. Participants practice the station with four vehicles while researchers record results on Data Collection Sheet #2. (10-15 mins)

Next, participants will learn how to use the camera inspection I-4.0 tool. In all cases, the participant will be shown how to use the technology to find errors in the construction on the top of the vehicle and identify which part is incorrect.

In the first phase, participants will be asked to complete the assembly process for as many cars as they can and limit the number of errors produced. The participants will complete the four treatments in the order randomly selected for them, each treatment will last 10 minutes. During that time, we expect that each participant will produce between 3 and 6 cars, based on prior performance data and the 60-second cycle time for which the instructions were designed. Observations will be recorded on Data Collection Sheet #1. (10 mins) We expected approximately 5 minutes between treatments to have participants complete the NASA TLX and System Usability Scale. (Total of 55 minutes of treatment completion time).

Participant performance in all treatments will be recorded on two cameras, a third-person view from a camera mounted nearby, and a head-mounted device. Experimental data will be derived from subsequent analysis of these videos. Participants will not be allowed to ask questions during any data collection phase of the experiment. Once the experiment is concluded, each participant will complete an exit survey that incorporates the NASA TLX and System Usability Scale instruments. When the surveys are completed a research associate will solicit any additional general feedback and ask if the participant experienced any injury or discomfort, and invite them to attend a follow-up session for more in-depth exploration of Augmented Reality. Their responses will be recorded on the exit survey. (5-10 mins) We conservatively estimate a total time commitment of 70-90 minutes for each participant.

13C. List all data collection instruments used in this project, in the order they appear in Appendix C.
(e.g., surveys and questionnaires in the format that will be presented to participants, educational tests, data collection sheets, interview questions, audio/video taping methods etc.)

1. Subject Recruitment Data Sheet: filled out during the screening call; includes the exclusion checklist, participant number, basic demographics (age and gender), and date / time of scheduled trial
2. Code Sheet: collects the personally identifiable data for eligible participants, including name, contact info (phone, email) and subject number
3. Participant Intake Sheet: collects basic demographics and relevant prior experience
4. Data Collection Sheet: consists of general notes from the experiment and data derived from subsequent analysis of video recordings
5. NASA Task Load Index (TLX) instrument.
6. System Usability Scale (SUS) instrument.
7. Behavioral Control Survey based on the Adult ADHD Self-Report Scale (ASRSv1.1)
8. General feedback form to collect open-ended comments and to note any participant injury or discomfort as well as their interest in the follow-up session.

Additionally, video of each session will be recorded as described above, and pictures of the assembled LEGO vehicles will be taken after each task is completed. These items are not included in the Appendix.

13D. Data analysis: Describe how data will be analyzed. If a data collection form (DCF) will be used, submit a copy of the DCF.

In both investigations, the independent variable is treatment type, and the dependent variables are task completion time and number of errors. The dependent variables will be recorded for each car completed in both sessions.

Data will be analyzed with a combination of visual (e.g., box plots) and statistical methods. Methods based on analysis of variance (ANOVA) will be used to test the stated hypotheses. Additional analysis will be done to explore the relationship between other variables of interest, including demographics, mental workload, behavioral control, and system usability with the measured outcomes.

13E. List any drugs, medications, supplements, or imaging agents that participants will ingest/ receive during participation in the study or indicate not applicable (N/A).

n/a

14. Risks & Discomforts: List and describe all the risks participants may encounter in this research including risks from item 6d of this form, in this research. <u>If deception will be part of the study, provide the rationale for the deception, describe the debriefing process, and attach a copy of the debriefing form that will be used as Appendix D.</u> (Examples of possible risks are in section #6C)

1. Physical Discomfort: All participants will be required to wear the HoloLens2 device, regardless of treatment group to control for its effects on user fatigue, etc., and to allow us to record a first-person view of their session. As a result, they may experience mild physical discomfort including neck strain after prolonged use. The limited duration of this study should mitigate this effect.
2. Vestibular and Visual Discomfort: Participants assigned to the HMDAR and HMDMR treatments will experience display technology that may cause mild dizziness, eye strain, and related effects. Owing to the see-through design of the HoloLens2 device these effects are less common and less pronounced than seen in fully immersive Virtual Reality (VR) headsets.
3. Trip and Impact Risk: Any head-mounted display can reduce the wearer's peripheral vision and otherwise impact their natural field of view. Consequently, they may become more susceptible to tripping over or running into things around them. This risk is minimized by the HoloLens2's design, which offers a very wide, minimally obscured field of view. Furthermore, the HL2 is a standalone device, so there is no risk of tripping over a cord. Additionally, the participant is generally stationary in an environment free of obstruction. Finally, the Lean Lab is a clean, organized, safe, and well-lit environment with no history of related hazards.
4. Breach of Confidentiality Risk: All resulting data will be anonymized, and video of each session will be recorded from the first person and top-down angles to prevent participant exposure. That said, subjects could be seen entering, leaving, or during the experiment. All of these create a small possibility that subjects could be identified, inadvertently breaching their confidentiality. Additionally, there is the possibility that the subject code list, which connects each participant's identity with their experimental data, could be obtained. Mitigation methods for this risk are described in section 17 Protection of Data.
5. Psychological Discomfort: Due to the nature of the experiment, some participants may experience mild psychological discomfort induced by its time and performance-based measures. Participants will be told that their objective is to learn to perform the task quickly and error free. Otherwise, no overt pressure is put on the subjects to perform. Given that the outcome of their performance has no impact on their life outside the experiment, any related psychological discomfort should be minimal and short-lived.
6. COVID-19 Exposure: This study will be a Category C study with no High-Risk Procedures or Participants. Precautions will be implemented using the COVID-19 2022 Precautions Matrix to determine appropriate precautions at the time of data collection(s) for a Category C study. All work surfaces and the HMD will be wiped down before and after each participant. Necessary supplies will be made available, including as masks, hand sanitizer (60%+ alcohol), tissues, paper towels, trash baskets, and cleaners / disinfectants. All research participants will follow the University's guidance on self-screening. At the time of this writing, the CDC's COVID-19 community level for Lee County, Alabama is LOW, so participant screening is not required. The Shelby Center for Engineering Technology, where this protocol will be administered, is assigned the highest level of building readiness due to increased air turn-over and filtration. Further details and resources can be found in Appendix D.

15. Precautions / Minimization of Risks

15A. Identify and describe all precautions that will be taken to eliminate or reduce risks listed in items 6.c. and 14. If participants can be classified as a "vulnerable" population, describe additional safeguards that will be used to assure the ethical treatment of vulnerable individuals. **If applicable, <u>submit a copy of any emergency plans/procedures and medical referral lists in Appendix D.</u> (Sample documents can be found online at** https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs **precautions)**

This study does not involve any vulnerable populations. Please see section 14, where the primary mitigations are described for each identified risk. Additionally, all participant activities will be supervised and monitored for relevant symptoms. If any participant experiences dizziness or related vestibular issues, or any other significant but unexpected side-effect, we will suspend the experiment, remove the HMD, have them sit and offer drinking water while assessing the situation. If escalation is required, the emergency plan and contact list is included in Appendix D. During the debriefing all participants will be asked if they were injured or experienced any discomfort during their trials. The debriefing also serves to keep each participant under our supervision long enough to ensure no lingering or delayed effects.

15B. If the internet, mobile apps, or other electronic means will be used to collect data, describe confidentiality and/or security precautions that will be used to protect (or not collect) identifiable data? Include protections used during collection of data, transfer of data, and storage of data. If participant data may be obtained and/or stored by apps during the study, describe.

n/a

15C. Does this research include purchase(s) that involve technology hardware, software or online services?
☐ YES ☒ NO
If YES:

A. **Provide the name of the product** Click or tap here to enter text.
and the manufacturer of the product Click or tap here to enter text.
B. **Briefly describe use of the product in the proposed human subject's research.**
Click or tap here to enter text.

C. **To ensure compliance with AU's Electronic and Information Technology Accessibility Policy, contact AU IT Vendor Vetting team at vetting@auburn.edu to learn the vendor registration process (prior to completing the purchase).**
D. **Include a copy of the documentation of the approval from AU Vetting with the revised submission.**

15D. Additional Safeguards

Will DEXA, pQCT, or other devices which emit radiation be used? ☐ Yes ☒ No
If yes, the IRB will notify the Auburn Department of Risk Management and Safety, who will contact the Alabama Department of Public Health (ADPH) and secure approval. Research which includes device(s) which emit radiation may NOT be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of ADPH approval.

Will a Certificate of Confidentiality (CoC) issued by NIH be obtained ☐ Yes ☒ No If yes, include CoC language in consent documents and include the documentation of CoC approval. Research which includes a CoC may not be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of CoC approval. AU Required CoC Language

Is the study a clinical trial? ☐ Yes ☒ No
If yes, provide the National Clinical Trial (NCT) # Click or tap here to enter text. **and include required clinical trial information in all consent documents. AU Clinical Trial Information**

16. Benefits

16A. List all realistic direct benefits participants can expect by participating in this study. (Compensation is not a benefit) If participants will not directly benefit check here. ☒

There are no direct benefits for participants in this study. It will offer all of them an opportunity to interact with projection and/or head-mounted AR hardware and training methods for the first time. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

16B. List realistic benefits for the general population that may be generated from this study.

First Investigation

Turnover in the workforce and the lack of skilled labor necessitates scalable, efficient training methods. Furthermore, the shift from mass production to mass customization forces operators to contend with wide variance in the assembly steps required at each workstation. Together, these trends demand innovative methods for operator training and support.

Augmented and mixed reality are expected to help fill that need, but it is a fragmented market with a variety of solutions. Few studies explore the relationship between those methods (and the affordances that differentiate them) and corresponding learning rates and transfer. We believe this investigation will make meaningful contributions to that effort, helping to build a cohesive understanding of the utility of these systems and best practices for their application.

Second Investigation

The ultimate goal of this investigation is to develop a smart production system through the integration of Lean and Industry 4.0 (I-4.0) technology. However, still, in the literature, there is a research gap to see how the Lean and I-4.0 technologies are aligned. Through this investigation, this research gap would be mitigated. Additionally, the proposed production model will be transferred to Small and Medium Enterprises (SMEs), and thus millions of people will be benefited.

Additionally, by investigating the different impacts of technology and workplace changes on participants with few or many self-reported behavioral control symptoms, recommendations for future implementation can be made to best suit workers with conditions, such as ADD and ADHD. Designing manufacturing workplaces with an end goal of universal design that will be better suited for a variety of workers will benefit many in the workplace.

17. Protection of Data

17A. Data are collected:

- ☐ **Anonymously with no direct or indirect coding, link, or awareness by key personnel of who participated in the study (skip to item E)**
- ☐ **Confidentially, but without a link to participant's data to any identifying information (collected as "confidential" but recorded and analyzed "anonymous") (Skip to item E).**
- ☒ **Confidentially with collection and protection of linkages to identifiable information.**

17B. If data are collected with identifiers and coded or as coded or linked to identifying information, describe the identifiers and how identifiers are linked to participants' data.

In addition to the consent form, a code list will be maintained that includes identifying data of each participant (name, contact information, and ID number). This will be linked to all other data collection forms by the participant number. The consent forms and code list will be maintained on paper only, to facilitate secure storage and disposal (shredding). The consent form will not include reference to the participant's ID number. Only the code list will directly connect participants to their data.

The video recordings may also allow for participants to be identified, though the first-person recording will not allow a view of their face and the third-person view will focus on the work area. If the recorders do not provide a video-only option, audio from those sessions, which may also provide identifying data, will be stripped from the recordings before storage.

17C. Provide the rationale for need to code participants' data or link the data with identifying information.

Only for the purpose of contacting participants while the protocol is open. Once completed, the code list will be destroyed, making the data anonymous.

17D. Describe how and where identifying data and/or code lists will be stored. (Building, room number, AU BOX?) **Describe how the location where data is stored will be secured. For electronic data, describe security measures. If applicable, describe where IRB-approved and participant signed consent documents will be kept on campus for 3 years after the study ends.**

Signed consent forms and the code list will be kept in a secure, locked file in offices 3301J (first investigation) or 0317 (second investigation) of Shelby Center.

17E. Describe how and where data will be stored (e.g., hard copy, audio/ visual files, electronic data, etc.), and how the location where data is stored is separated from identifying data and will be secured. For electronic data, describe security. Note use of a flash drive or portable hard drive is not appropriate if identifiable data will be stored; rather, identifying participant data must be stored on secured servers.

All electronic data pertaining to the study will be stored on a secured server. Non-identifiable data will be available to other members of the research team.

17F. List the names of all who will have access to participants' data? (If a student PI, the faculty advisor must have full access and be able to produce study data in the case of a federal or institutional audit.)

- Consent forms and code list: Dan O'Leary, Victoria Ballard, Md Monir Hossain, Dr. Richard Sesek, Dr. Gregory Purdy
- Non-identifiable data: full research team, by request

17G. When is the latest date that identifying information or links will be retained and how will that information or links be destroyed? (Check here if only anonymous data will be retained ☒)

December 2023

AUBURN
UNIVERSITY

INDUSTRIAL & SYSTEMS
ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

The Effects of Augmented Instruction on Manufacturing Assembly Training

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. Following an initial phone screening the experiment will be scheduled at your convenience. After a brief orientation you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned one of the following forms of instructional media: paper work instructions (PWI), projected augmented reality (PAR), head-mounted AR (HMDAR), or head-mounted mixed reality (HMDMR). After a 10-minute training session you will be asked to repeat the assembly task from memory for 4 cars. Paper work instructions will remain available for reference as needed. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 45-60 minutes.

This study has some risk of physical and psychological discomfort, including fatigue, dizziness, eyestrain, and performance anxiety. Participants assigned the HMD instructional media are most susceptible to physical discomfort due to the nature of its display system, which can also increase the risk of tripping and impact. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience modern AR training methods, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

You are invited to participate in a research study to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The study is being conducted by Dan O'Leary, Ph.D. Candidate, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are not prone to motion sickness.
2. Have no prior experience with head-mounted or projected Augmented Reality (AR) systems.
3. Have no prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO Lab) as part of INSY 5800/6800 or otherwise.
4. Are age 18 or older.

What will be involved if you participate?

If you decide to participate in this research study, you will be asked to follow a mix of paper and augmented (projected or head-mounted AR) work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will be approximately 45-60 minutes. You will be required to wear a HoloLens2 head-mounted display (HMD) and video of your session will be recorded for later analysis. Another video camera will capture the work area from above. Camera placement is designed to prevent / limit the capture of personally identifiable imagery. Fully redacted versions of these videos, wherein any personally identifiable imagery is removed, will be kept indefinitely. Original recordings will be deleted within 1 year of the protocol's completion.

Are there any risks or discomforts?

The risks associated with participating in this study are identified below.

1. Physical discomfort and/or fatigue related to the weight of the HoloLens2 HMD.
2. Vestibular and/or visual discomfort for participants assigned to the HMD AR instructional methods, which may cause mild dizziness, eye strain, and related effects in some users.
3. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
4. Trip and impact risk due to slightly altered field of view and reduced peripheral vision while wearing the HoloLens2 HMD.
5. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.
6. Exposure to COVID-19 or other respiratory illnesses, such as the flu.

The discomforts identified are considered mild and unlikely. The HoloLens2 is well-balanced and uses a state-of-the-art optical see-through design that limits display-related discomforts. To minimize the risk of tripping and impact, participants are largely stationary in a well-lit area that is free of hazards. The HoloLens2 features a wireless design, which eliminates cables as a source of tripping hazard. Finally, all activities will be supervised, and participants will be continuously monitored for relevant symptoms.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

To mitigate the risk of exposure to COVID-19 and other respiratory illnesses, the research team will follow University policies outlined on the Human Research COVID-19 Precautions page. All work surfaces and equipment will be wiped down before and after each participant, and all necessary supplies (e.g. masks, hand sanitizer) will be made available. The research staff will follow the University's guidance on self-screening. Finally, conditions will be monitored, and precautions adjusted as necessary throughout the data collection process.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to interact with projection and/or head-mounted AR hardware and training methods. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
All volunteers recruited from the Auburn University community will be eligible for up to $100 in participation and performance related prizes. Odds of winning one of the eight available prizes will depend on the number of qualifying participants. Attendance and participation in the end of semester open house event is required for some of the prizes.

Volunteers from the manufacturing industry will be given $40 in cash / gift card(s) for their participation.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

If you have questions about this study, please ask now or contact Dan O'Leary at djo0008@auburn.edu, 407-399-3189, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Participant's signature Date	Investigator obtaining consent Date
Printed Name	Printed Name

AUBURN UNIVERSITY

INDUSTRIAL & SYSTEMS ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

Studying Manufacturing with LEGO® Research

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity, learning rates, and skills transfer for industrial assembly tasks. Following an initial phone screening, the experiment will be scheduled at your convenience. After a brief orientation, you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned an order to complete the following treatments: paper work instructions (PWI), assembly with a pre-completed model for quality checks, an inspection camera for quality checks, and both the pre-completed model and inspection camera. You will be asked to complete four car assemblies for training using the paper work instructions prior to using the prescribed tasks. After the training, each treatment will last 10 minutes for a total of four treatments. Paper work instructions will remain available for reference as needed. Between each task you will be asked to complete two brief surveys about your experience. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 70-90 minutes.

This study has some risk of physical and psychological discomfort, including fatigue and performance anxiety. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience training in the Tiger Motors Lab, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

You are invited to participate in a research study to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity. The study is being conducted by Victoria Ballard and Md Monir Hossain, Ph. D. students, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are age 18 or older.

What will be involved if you participate?
If you decide to participate in this research study, you will be asked to follow work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will -be approximately 70-90 minutes. Video of your session will be recorded for later analysis. Camera placement is designed to prevent / limit the capture of personally identifiable imagery.

Are there any risks or discomforts?
The risks associated with participating in this study are identified below.

1. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
2. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to participate in research in the Tiger Motors Lab. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
All volunteers recruited from the Auburn University community will be eligible for up to $100 in participation and performance related prizes. Odds of winning one of the eight available prizes will depend on the number of qualifying participants. Attendance and participation in the end of semester open house event is required for some of the prizes.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

If you have questions about this study, please ask now or contact Victoria Ballard at victoria.ballard@auburn.edu,360-632-1359, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

____
Participant's signature Date

____
Investigator obtaining consent Date

____
Printed Name

____
Printed Name

Appendix A - Reference List

First Investigation

[1] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355-385, 1997

[2] B. E. Shelton and N. R. Hedley, "Exploring a Cognitive Basis for Learning Spatial Relationships with Augmented Reality," *Technology, Instruction, Cognition and Learning*, vol. 1, no. 4, 2003

[3] C. Chen *et al.*, "Using Microsoft HoloLens to improve memory recall in anatomy and physiology: A pilot study to examine the efficacy of using augmented reality in education," *Journal of Educational Technology Development and Exchange*, vol. 12, no. 1, 2019

[4] S. N. Leonard and R. N. Fitzgerald, "Holographic learning: A mixed reality trial of Microsoft HoloLens in an Australian secondary school," *Research in Learning Technology*, vol. 26, no. 0, 2018

[5] R. T. Azuma and G. Bishop, "Improving Static and Dynamic Registration in an Optical See-through HMD Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques," *SIGGRAPH '94*, pp. 197-204, 1994

[6] B. C. Kress, "Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets," SPIE, 2020.

[7] E. Oztemel and S. Gursev, "Literature Review of Industry 4.0 and Related Technologies," *Journal of Intelligent Manufacturing*, vol. 31, pp. 127-182, 2020

[8] R. T. Azuma, "The road to ubiquitous consumer augmented reality systems," *Human Behavior and Emerging Technologies*, vol. 1, pp. 26-32, 2019

[9] T. Masood and J. Egger, "Augmented reality in support of Industry 4.0—Implementation challenges and success factors," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 181-195, 2019

[10] T. Masood and J. Egger, "Adopting augmented reality in the age of industrial digitalisation," *Computers in Industry*, vol. 115, pp. 103112, 2020

[11] A. D. Kaplan *et al.*, "The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 63, no. 4, pp. 706-726, 2021

[12] S. Borsci *et al.*, "Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance," *Computers in Industry*, vol. 67, pp. 17-26, 2015

[13] J. Jetter *et al.*, "Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits," *Computers in Human Behavior*, vol. 87, pp. 18-33, 2018

[14] O. Danielsson *et al.*, "Augmented reality smart glasses in industrial assembly: Current status and future challenges," *Journal of Industrial Information Integration*, vol. 20, pp. 100175, 2020

[15] S. Büttner *et al.*, "Augmented Reality Training for Industrial Assembly Work - Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020

[16] V. Havard *et al.*, "A use case study comparing augmented reality (AR) and electronic document-based maintenance instructions considering tasks complexity and operator competency level," *Virtual Reality*, vol. 25, no. 4, pp. 999-1014, 2021

[17] D. Parsons and K. MacCallum, "Current Perspectives on Augmented Reality in Medical Education: Applications, Affordances and Limitations," *Advances in Medical Education and Practice*, vol. Volume 12, pp. 77-91, 2021

Second Investigation

[1] A. Giorgi *et al.*, “Wearable Technologies for Mental Workload, Stress, and Emotional State Assessment during Working-Like Tasks: A Comparison with Laboratory Technologies,” *Sensors*, vol. 21, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/s21072332.

[2] H. Devos *et al.*, “Psychometric Properties of NASA-TLX and Index of Cognitive Activity as Measures of Cognitive Workload in Older Adults,” *Brain Sci.*, vol. 10, no. 12, Art. no. 12, Dec. 2020, doi: 10.3390/brainsci10120994.

[3] S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” Nov. 2016, doi: 10.1177/154193120605000909.

[4] “CDC - NORA - Healthy Work Design and Well-Being Sector Agenda,” Jan. 30, 2020. https://www.cdc.gov/nora/councils/hwd/research-agenda.html (accessed Apr. 18, 2022).

[5] “Attention-Deficit/Hyperactivity Disorder (ADHD),” *National Institute of Mental Health (NIMH)*. https://www.nimh.nih.gov/health/statistics/attention-deficit-hyperactivity-disorder-adhd (accessed Apr. 19, 2022).

[6] R. C. Kessler, M. Lane, P. E. Stang, and D. L. V. Brunt, “The prevalence and workplace costs of adult attention deficit hyperactivity disorder in a large manufacturing firm,” *Psychol. Med.*, vol. 39, no. 1, pp. 137–147, Jan. 2009, doi: 10.1017/S0033291708003309.

[7] R. C. Kessler, L. A. Adler, M. J. Gruber, C. A. Sarawate, T. Spencer, and D. L. Van Brunt, “Validity of the World Health Organization Adult ADHD Self-Report Scale (ASRS) Screener in a representative sample of health plan members,” *Int. J. Methods Psychiatr. Res.*, vol. 16, no. 2, pp. 52–65, 2007, doi: 10.1002/mpr.208.

[8] M. P. Ciano, P. Dallasega, G. Orzes, and T. Rossi, “One-to-one relationships between Industry 4.0 technologies and Lean Production techniques: a multiple case study,” *Int. J. Prod. Res.*, vol. 59, no. 5, pp. 1386–1410, Mar. 2021, doi: 10.1080/00207543.2020.1821119.

[9] J. P. Womack and D. T. Jones, “From lean production to lean enterprise.,” *Harv. Bus. Rev.*, vol. 72, no. 2, 1994, Accessed: Jun. 26, 2022. [Online]. Available: https://elibrary.ru/item.asp?id=1991549

[10] V. L. Bittencourt, A. C. Alves, and C. P. Leão, “Lean Thinking contributions for Industry 4.0: a Systematic Literature Review,” *IFAC-Pap.*, vol. 52, no. 13, pp. 904–909, Jan. 2019, doi: 10.1016/j.ifacol.2019.11.310.

[11] D. Kolberg, J. Knobloch, and D. Zühlke, “Towards a lean automation interface for workstations,” *Int. J. Prod. Res.*, vol. 55, no. 10, pp. 2845–2856, May 2017, doi: 10.1080/00207543.2016.1223384.

[12] M. P. Ciano, P. Dallasega, G. Orzes, and T. Rossi, “One-to-one relationships between Industry 4.0 technologies and Lean Production techniques: a multiple case study,” *Int. J. Prod. Res.*, vol. 59, no. 5, pp. 1386–1410, 2021.

[13] L. S. Dalenogare, G. B. Benitez, N. F. Ayala, and A. G. Frank, “The expected contribution of Industry 4.0 technologies for industrial performance,” *Int. J. Prod. Econ.*, vol. 204, pp. 383–394, Oct. 2018, doi: 10.1016/j.ijpe.2018.08.019.

[14] G. Tortorella, R. Sawhney, D. Jurburg, I. C. de Paula, D. Tlapa, and M. Thurer, “Towards the proposition of a Lean Automation framework: Integrating Industry 4.0 into Lean Production,” *J. Manuf. Technol. Manag.*, vol. 32, no. 3, pp. 593–620, Jan. 2020, doi: 10.1108/JMTM-01-2019-0032.

[15] B. Wang, J. Zhao, Z. Wan, J. Ma, H. Li, and J. Ma, “Lean Intelligent Production System and Value Stream Practice,” 2016, doi: 10.12783/DTEM/ICEM2016/4106.

[16] M. Rossini, F. Costa, G. L. Tortorella, and A. Portioli-Staudacher, “The interrelation between Industry 4.0 and lean production: an empirical study on European manufacturers,” *Int. J. Adv. Manuf. Technol.*, vol. 102, no. 9, pp. 3963–3976, Jun. 2019, doi: 10.1007/s00170-019-03441-7.

Appendix B - Recruiting Materials

In-Class Recruiting Script

Hello, Class.

Industrial Engineering graduate students pursuing their PhDs are recruiting participants for a research study. They are investigating the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. These investigations hope to better understand the relationships between those methods, learning effectiveness, and operator performance. A flyer with details of the study will be emailed to each of you. If you are interested, please follow up as described therein.

Email Script

Dear Student,

Please review the attached flyer, which provides details of the study recently described in __class name__. You are invited to participate in a research study on the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. The research team is conducting this study as Ph.D. Candidates under the supervision of Dr. Richard Sesek, Tim Cook Associate Professor in the Department of Industrial and Systems Engineering at Auburn University.

If you would like to participate, simply respond to this email or via text / phone to 407-399-3189. Questions or concerns can be directed to me through the same channels, or you may contact my advisor Dr. Sesek (sesek@auburn.edu).

Thank you for your consideration,

Email Script, Industry

Dear Manufacturing Professional,

You are invited to participate in a research study on the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. Please review the attached flyer for details. The research team is conducting this study as Ph.D. Candidates under the supervision of Dr. Richard Sesek, Tim Cook Associate Professor in the Department of Industrial and Systems Engineering at Auburn University.

If you would like to participate, simply visit the website. Questions or concerns can be directed to the research team at leanmanufacturingteam@auburn.edu, or their faculty advisor Dr. Richard Sesek (sesek@auburn.edu).

Thank you for your consideration,

Confirmation Email

Dear <student name>,

Thank you for your interest in our study, and for taking the time to discuss it with me. I'm happy to confirm that your trial is scheduled as follows:

Date and Time:<date and time>

Location: Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab), in the basement of the Shelby Center for Engineering Technology, room 0317, located at 345 W Magnolia Ave, Auburn, AL 36849

Please arrive on time. We anticipate that it will take 45-90 minutes to complete the session.

If you need to reschedule or have further questions, feel free to respond to this email or call / text me at 407-399-3189.

Thank you for your participation,

Invitation to Open House Event

Dear Participant,

Thank you for volunteering for our research studies. We are writing to invite you to an open house event on <DATE and TIME>. This is an opportunity for you to try some of the methods and technologies that you may not have experienced before, along with a variety of other software on the HoloLens 2 device.

All attendees will have additional chances to win cash or gift cards worth up to $100 by participating in a single experiment designed to test how well you retained the prior training.

Food and drink will also be provided.

Hope to see you there!

The Research Team

Flyers

On the following pages are flyers for both investigations, formatted as posters and slides.

1. First investigation, printed poster format.
2. Second investigation, printed poster format.
3. First investigation, manufacturing recruitment, printed poster format.
4. First investigation, displayed slide format.
5. Second investigation, displayed slide format.

Augmented Reality Research Study

Training methods for tomorrow's workforce, today!

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

This study is open to anyone 18 and older that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in in the Tiger Motors Lean Education Center (aka LEGO® Lab). It takes 45 to 60 minutes to complete.

💰💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰💰

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up: https://aub.ie/TigerMotorsResearch or scan the QR Code. Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

Studying Manufacturing with LEGO(R) Research

Participate in research in Auburn's Tiger Motors Lab!

The Effects of Lean Tools and Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials. Participants will assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours.

💰💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰💰

This study is open to anyone 18 and older.

TIGER MOTORS
LEAN EDUCATION CENTER

AUBURN UNIVERSITY

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up: https://aub.ie/TigerMotorsResearch or scan the QR Code. Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

Manufacturing Volunteers Needed!

Augmented Reality Research Study

The Effects of Augmented Instruction on Manufacturing Assembly Training

Are You Interested in Augmented and Mixed Reality?
Do You Want to experience the latest in Projected and Head-Mounted AR?
Build LEGO® cars in Auburn's famous Lean Education Lab, *for Science!*

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

All volunteers will receive a $40 gift card as thanks for their participation.

This study is open to operators 18 and older, that aren't prone to motion sickness, and have no prior experience with similar AR systems or Auburn's Lean Education Center. For details, or to sign up, head to https://aub.ie/TigerMotorsResearch or scan the QR code below.

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

Augmented Reality Research Study

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰

Open to anyone 18 and older that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in the Tiger Motors Lean Education Center (aka LEGO® Lab).

Scan to Sign Up!

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up at https://aub.ie/TigerMotorsResearch or scan the QR Code.

Please contact the research team with any other questions: leanmanufacturingteam@auburn.edu

Studying Manufacturing with LEGO®

The Effects of Lean Tools & Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials.

💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰

You will be asked to assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours.

This study is open to anyone 18 and older.

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up at https://aub.ie/TigerMotorsResearch or scan the QR Code.

Please contact the research team with any other questions: leanmanufacturingteam@auburn.edu

Scan to Sign Up!

Tiger Motors
Lean Education Center

Appendix C - Data Collection Instruments

See attached, on the pages that follow:

1. Subject Recruitment Data Sheet
2. Code Sheet
3. Participant Intake Sheet
4. Data Collection Sheet #1
5. Data Collection Sheet #2
6. Task Loading Index (NASA TLX)
7. System Usability Scale
8. Behavioral Control Survey
9. General Feedback

Subject Recruitment Data Sheet

Eligibility Checklist:

- ☐ 18 or older
- ☐ Not prone to motion sickness
- ☐ No prior experience with projected or head-mounted augmented reality systems
- ☐ No prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab) as part of INSY 5/6800 or otherwise

If eligible, record name, contact info (phone, email), and subject number in code sheet.

Participant Number: ____________________

Gender: ______________________________

Age: ________________________________

Eligible: ____ I1 ____ I2 ____ Both

Scheduled Trial(s): ____________________

Notes:

Eligibility Checklist:

- ☐ 18 or older
- ☐ Not prone to motion sickness
- ☐ No prior experience with projected or head-mounted augmented reality systems
- ☐ No prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab) as part of INSY 5/6800 or otherwise

If eligible, record name, contact info (phone, email), and participant number in code sheet.

Participant Number: ____________________

Gender: ______________________________

Age: ________________________________

Eligible: ____ I1 ____ I2 ____ Both

Scheduled Trial(s): ____________________

Notes:

Participant Intake Sheet, p1 / 2

Participant #: __________ **Date: __________**

1. Gender:
   - ☐ Female
   - ☐ Male
   - ☐ Other
2. Age: _____
3. Race (select those with which you identify):
   - ☐ American Indian or Alaska Native
   - ☐ Asian
   - ☐ Black or African-American
   - ☐ Native Hawaiian or Other Pacific Islander
   - ☐ White
   - ☐ More than one race
   - ☐ Unknown or not reported
4. Ethnicity (select ONLY one with which you most closely identify):
   - ☐ Hispanic or Latino
   - ☐ Not Hispanic or Latino
   - ☐ Unknown or not reported
5. Country of Origin: ________________________
6. What language do you mainly speak at home?
   - ☐ English
   - ☐ Other
7. What is the highest level of school you have completed or the highest degree you have received?
   - ☐ Less than high school degree
   - ☐ High school degree or equivalent (e.g., GED)
   - ☐ Some college but no degree
   - ☐ Associate degree
   - ☐ Bachelor degree
   - ☐ Graduate degree: ____ Master or ____ PhD

Participant #: __________ **Date: __________**

8. If you are currently pursuing a degree, please complete the following:

   College (e.g. Education or Business): ________________________

   Program (e.g. MS Adult Ed or BS Accounting) : ________________________

9. Which of the following statements best describes your experience building LEGO models?

   ☐ I have little to no experience building LEGO models.

   ☐ I have some experience building LEGO models.

   ☐ I have lots of experience building LEGO models.

   ☐ I consider myself an expert in building LEGO models.

10. Please indicate your level of manufacturing experience

    ☐ I have no experience in manufacturing.

    ☐ I have taken one or more classes in manufacturing.

    ☐ I have held a part-time or temporary position in manufacturing.

    ☐ I have 1 or more years of experience working in manufacturing.

11. Do you normally wear corrective lenses? ____ Glasses ____ Contacts ____ Neither

    If yes, do you plan to wear them during this experiment? ____ Yes ____ No

12. Are you color blind? ____ Yes ____ No

13. Do you have any other condition that might affect your performance in this study?

    ____ Yes ____ No

14. What is your height? ____ feet ____ inches

15. Have you ever run an event of 5 kilometers or more? ____ Yes ____ No

16. How did you learn about this study? ________________________

Code Sheet

Part. #	Date	Name	Email	Phone	Assigned	Notes
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	

CONFIDENTIAL

Data Collection Sheet #1

Participant #: __________ **Date: __________**

First Investigation
Circle Training Treatment:
PWI / PAR / HMDAR / HMDMR

Second Investigation	
Treatment Number	**Treatment**
1 / 2 / 3 / 4	Control / Lean / I-4.0 / Lean+I-4.0

Car #	TCT	Errors Made		Uncorrected Error Types			PWI Ref Count	Trial Notes
		Corrected	Uncorrected	Sel	Pos	Rot		
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Observer Initials: ________

Data Collection Sheet #2

Participant #: __________ **Date: __________**

First Investigation
Circle Training Treatment:
PWI / PAR / HMDAR / HMDMR

Second Investigation
Training with
Paper Work Instructions

Car #	**TCT**	**Errors Made**		**Uncorrected Error Types**			**PWI Ref Count**	**Trial Notes**
		Corrected	**Uncorrected**	**Sel**	**Pos**	**Rot**		
1								
2								
3								
4								

General Notes:

Observer Initials: ________

Task Loading Index, p1 / 2

Participant #: __________ **Invest / Treat: __________** **Date: __________**

Sources of Workload Evaluation

Consider the following definitions:

Title	Range	Description
Mental Demand	Low / High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low / High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low / High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Perfect / Failure	How successful do you think you were in accomplishing the goals of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low / High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Low / High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

For each of the following pairs, circle the word that represents the more important contributor to workload for the specific task(s) you performed in this experiment.

Effort or **Performance**	**Temporal Demand** or **Frustration**	**Physical Demand** or **Performance**	**Temporal Demand** or **Mental Demand**	**Mental Demand** or **Physical Demand**
Temporal Demand or **Effort**	**Physical Demand** or **Frustration**	**Frustration** or **Effort**	**Performance** or **Mental Demand**	**Effort** or **Physical Demand**
Performance or **Frustration**	**Physical Demand** or **Temporal Demand**	**Performance** or **Temporal Demand**	**Mental Demand** or **Effort**	**Frustration** or **Mental Demand**

Task Loading Index, p2 / 2

Participant #: __________ **Invest / Treat: __________** **Date: __________**

Workload Rating Scales

For each of the following 6 questions, consider the assembly task you just completed. Record your immediate response to each item by putting an "X" at the point which matches your experience.

1. Mental Demand — How mentally demanding was the task?

Very Low — Very High

2. Physical Demand — How physically demanding was the task?

Very Low — Very High

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

Very Low — Very High

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

Perfect — Failure

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

Very Low — Very High

6. Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low — Very High

System Usability Scale

Participant #: __________ **Invest / Treat: __________** **Date: __________**

For each of the following 10 questions, consider the assembly task you just completed. Record your immediate response to each item by circling the number that you feel best represents your experience.

		Strongly Agree				Strongly Disagree
1	I think that I would like to use this system frequently.	1	2	3	4	5
2	I found the system unnecessarily complex.	1	2	3	4	5
3	I thought the system was easy to use.	1	2	3	4	5
4	I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5	I found the various functions in this system were well integrated.	1	2	3	4	5
6	I thought there was too much inconsistency in this system.	1	2	3	4	5
7	I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8	I found the system very cumbersome to use.	1	2	3	4	5
9	I felt very confident using the system.	1	2	3	4	5
10	I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Behavioral Control Survey

Participant #: __________ **Date: __________**

Please answer the questions below, rating yourself on each of the criteria shown using the scale on the right side of the page. As you answer each question, place an X in the box that best describes how you have felt and conducted yourself over the past 6 months.	Never	Rarely	Sometimes	Often	Very Often
1. How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?					
2. How often do you have difficulty getting things in order when you have to do a task that requires organization?					
3. How often do you have problems remembering appointments or obligations?					
4. When you have a task that requires a lot of thought, how often do you avoid or delay getting started?					
5. How often do you fidget or squirm with your hands or feet when you have to sit down for a long time?					
6. How often do you feel overly active and compelled to do things, like you were driven by a motor?					
7. How often do you make careless mistakes when you have to work on a boring or difficult project?					
8. How often do you have difficulty keeping your attention when you are doing boring or repetitive work?					
9. How often do you have difficulty concentrating on what people say to you, even when they are speaking to you directly?					
10. How often do you misplace or have difficulty finding things at home or at work?					
11. How often are you distracted by activity or noise around you?					
12. How often do you leave your seat in meetings or other situations in which you are expected to remain seated?					
13. How often do you feel restless or fidgety?					
14. How often do you have difficulty unwinding and relaxing when you have time to yourself?					
15. How often do you find yourself talking too much when you are in social situations?					
16. When you're in a conversation, how often do you find yourself finishing the sentences of the people you are talking to, before they can finish them themselves?					
17. How often do you have difficulty waiting your turn in situations when turn taking is required?					
18. How often do you interrupt others when they are busy?					

General Feedback

Participant #: __________ **Date: __________**

Please share with us any other feedback you have regarding this experiment.

For Research Associate Only

Follow-up? _____ Injury? _____ Discomfort? _____ Initial: _____

Appendix D - Emergency Plan, Contact List, and COVID Resources

Emergency Action Plan

In Case of Emergency ***DIAL 911***

For non-emergency assistance:

Service	On-Campus	Off-Campus
Ambulance (EMS)	9-749-8504	334-749-8504
City of Auburn Police	9-501-3100	334-501-3100
Auburn Medical Pavilion	9-364-3000	334-364-3000
East Alabama Medical Center, Opelika	9-749-3411	334-749-3411

Research Team Contact List:

Contact	Phone	Email
Dan O'Leary, Principal Investigator	407-399-3189 (cell)	djo0008@auburn.edu
Dr. Richard Sesek, Faculty Advisor	334-728-1438 (cell)	rfs0006@auburn.edu
Victoria Ballard, Graduate Student	360-632-1359 (cell)	vzb0024@auburn.edu
Dr. Gregory Harris, Faculty Advisor	334-844-1407 (office)	gah0015@auburn.edu
Dr. John Evans, Faculty Advisor	334-844-1418 (office)	evansjl@auburn.edu
Tom Devall, Tiger Motors Director	334-740-3905 (office)	tld0017@auburn.edu
Industrial & Systems Engineering Department	334-844-4340 (main office)	insy@eng.auburn.edu

Lab Location and Access:

Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab), Basement, Shelby Center, Auburn University, room 0317. Street address: 345 W Magnolia Ave, Auburn, AL 36849.

Elevator access: exit the lab and turn left

Stairwell access: exit the lab, turn left, proceed around the elevator in either direction. Stairwell entrance is on the inside wall behind the elevator.

Emergency exit: exit the lab and turn right. Continue to exit at ground level.

Emergency Equipment:

First aid kit, eye wash and shower station are present, as are fire extinguisher and alarm pull.

COVID-19 Resources

[CDC COVID-19 Data Tracker for Lee County, Alabama]

University Policies for Research Exposure and Related Resources:

- Human Research COVID-19 Precautions
- COVID-19 Guidance on Self Screening
- AU Facilities COVID Building Readiness Status Page

Auburn University Screening Protocol (source):

All research participants should be screened remotely (by phone or Zoom) for fever, cough, and flu-like symptoms the day before, with a repeat screening at the time of an in-person visit. Appropriate screening questions might include the following, which could be modified to fit your participant population and the location of in-person interactions:

1. Do you have a fever or Respiratory Symptoms? Symptoms include fever, acute respiratory infection, persistent cough, sore throat, fatigue and shortness of breath, or sudden loss of taste or smell with or without a fever.
2. Are you waiting on COVID-19 test results?
3. Have you been asked to self-isolate by your doctor?
4. In the past three weeks, have you visited another state, country, or facility with a substantial or high community COVID-19 level (see CDC COVID-19 Community Levels)?
5. Health/Vaccination Status - Do you have underlying medical conditions, or are you unvaccinated?

Precautions Matrix:

COVID-19 PRECAUTIONS MATRIX	CATEGORY A HIGH-RISK PROCEDURES*	CATEGORY B HIGH-RISK PARTICIPANTS**	CATEGORY C NO HIGH-RISK PROCEDURES OR PARTICIPANTS
HIGH COVID-19 COMMUNITY LEVEL	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS PPE: RESEARCH PERSONNEL WEAR N-95 OR KN95; EYE PROTECTION; GLOVES FOR DIRECT CONTACT; PARTICIPANTS WEAR FACE COVERINGS	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS PPE: RESEARCH PERSONNEL WEAR N-95 OR KN95; EYE PROTECTION; GLOVES FOR DIRECT CONTACT; PARTICIPANTS WEAR FACE COVERINGS	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES
MEDIUM COVID-19 COMMUNITY LEVEL	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES
LOW COVID-19 COMMUNITY LEVEL	FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES

*HIGH-RISK PROCEDURES ARE DEFINED AS ANY PROCEDURES THAT INCUR A SIGNIFICANT OR INCREASED RISK OF EXPOSURE, SUCH AS THROUGH FREQUENT OR SUSTAINED CLOSE CONTACT BETWEEN INVESTIGATORS AND PARTICIPANTS; SPECIMEN COLLECTION FROM PARTICIPANTS; OR ACTIVITIES INVOLVING INCREASED RESPIRATORY OUTPUT SUCH AS EXERCISE STUDIES.

**HIGH-RISK PARTICIPANTS INCLUDE PEOPLE AT HIGHER RISK OF SEVERE ILLNESS FROM SARS COV-2 INFECTION, INCLUDING PEOPLE WHO ARE UNVACCINATED, OLDER ADULTS, OR PEOPLE WITH CERTAIN MEDICAL CONDITIONS.

Appendix E - CITI Training Documentation

See attached, on the pages that follow:

1. Dan O'Leary (3)
2. Victoria Ballard (9)
3. Gregory Harris (1)
4. Richard Sesek (3)
5. Gregory Purdy (3)
6. Md Monir Hossain (6)
7. Diego Roberto Caputo Rodriguez (1)
8. Alex Barras (1)
9. Brown Teague (1)
10. Carson Tillery (1)
11. Kralyn Thomas (1)
12. Yen-Ting Guo (1)
13. Yuqing "Lucie" Wang (1)

Clean Modified Forms for Restamp

AUBURN
UNIVERSITY

INDUSTRIAL & SYSTEMS
ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

The Effects of Augmented Instruction on Manufacturing Assembly Training

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. Following an initial phone screening the experiment will be scheduled at your convenience. After a brief orientation you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned one of the following forms of instructional media: paper work instructions (PWI), projected augmented reality (PAR), head-mounted AR (HMDAR), or head-mounted mixed reality (HMDMR). After a 10-minute training session you will be asked to repeat the assembly task from memory for 4 cars. Paper work instructions will remain available for reference as needed. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 45-60 minutes.

This study has some risk of physical and psychological discomfort, including fatigue, dizziness, eyestrain, and performance anxiety. Participants assigned the HMD instructional media are most susceptible to physical discomfort due to the nature of its display system, which can also increase the risk of tripping and impact. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience modern AR training methods, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

Participant's Initials: ______

You are invited to participate in a research study to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The study is being conducted by Dan O'Leary, Ph.D. Candidate, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are not prone to motion sickness.
2. Have no prior experience with head-mounted or projected Augmented Reality (AR) systems.
3. Have no prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO Lab) as part of INSY 5800/6800 or otherwise.
4. Are age 18 or older.

What will be involved if you participate?
If you decide to participate in this research study, you will be asked to follow a mix of paper and augmented (projected or head-mounted AR) work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will be approximately 45-60 minutes. You will be required to wear a HoloLens2 head-mounted display (HMD) and video of your session will be recorded for later analysis. Another video camera will capture the work area from above. Camera placement is designed to prevent / limit the capture of personally identifiable imagery. Fully redacted versions of these videos, wherein any personally identifiable imagery is removed, will be kept indefinitely. Original recordings will be deleted within 1 year of the protocol's completion.

Are there any risks or discomforts?
The risks associated with participating in this study are identified below.

1. Physical discomfort and/or fatigue related to the weight of the HoloLens2 HMD.
2. Vestibular and/or visual discomfort for participants assigned to the HMD AR instructional methods, which may cause mild dizziness, eye strain, and related effects in some users.
3. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
4. Trip and impact risk due to slightly altered field of view and reduced peripheral vision while wearing the HoloLens2 HMD.
5. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.
6. Exposure to COVID-19 or other respiratory illnesses, such as the flu.

The discomforts identified are considered mild and unlikely. The HoloLens2 is well-balanced and uses a state-of-the-art optical see-through design that limits display-related discomforts. To minimize the risk of tripping and impact, participants are largely stationary in a well-lit area that is free of hazards. The HoloLens2 features a wireless design, which eliminates cables as a source of tripping hazard. Finally, all activities will be supervised, and participants will be continuously monitored for relevant symptoms.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

To mitigate the risk of exposure to COVID-19 and other respiratory illnesses, the research team will follow University policies outlined on the Human Research COVID-19 Precautions page. All work surfaces and equipment will be wiped down before and after each participant, and all necessary supplies (e.g. masks, hand sanitizer) will be made available. The research staff will follow the University's guidance on self-screening. Finally, conditions will be monitored, and precautions adjusted as necessary throughout the data collection process.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to interact with projection and/or head-mounted AR hardware and training methods. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
All volunteers recruited from the Auburn University community will be eligible for up to $100 in participation and performance related prizes. Odds of winning one of the eight available prizes will depend on the number of qualifying participants. Attendance and participation in the end of semester open house event is required for some of the prizes.

Volunteers from the manufacturing industry will be given $40 in cash / gift card(s) for their participation.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

If you have questions about this study, please ask now or contact Dan O'Leary at djo0008@auburn.edu, 407-399-3189, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Participant's signature Date	Investigator obtaining consent Date
Printed Name	Printed Name

The Auburn University Institutional Review Board has approved this Document for use from 04/07/2023 to -------------
Protocol # 22-538 EP 2301

INDUSTRIAL & SYSTEMS ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

Studying Manufacturing with LEGO® Research

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity, learning rates, and skills transfer for industrial assembly tasks. Following an initial phone screening, the experiment will be scheduled at your convenience. After a brief orientation, you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned an order to complete the following treatments: paper work instructions (PWI), assembly with a pre-completed model for quality checks, an inspection camera for quality checks, and both the pre-completed model and inspection camera. You will be asked to complete four car assemblies for training using the paper work instructions prior to using the prescribed tasks. After the training, each treatment will last 10 minutes for a total of four treatments. Paper work instructions will remain available for reference as needed. Between each task you will be asked to complete two brief surveys about your experience. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 70-90 minutes.

This study has some risk of physical and psychological discomfort, including fatigue and performance anxiety. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience training in the Tiger Motors Lab, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

The Auburn University Institutional Review Board has approved this Document for use from
04/07/2023 to -------------
Protocol # 22-538 EP 2301

You are invited to participate in a research study to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity. The study is being conducted by Victoria Ballard and Md Monir Hossain, Ph. D. students, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are age 18 or older.

What will be involved if you participate?
If you decide to participate in this research study, you will be asked to follow work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will -be approximately 70-90 minutes. Video of your session will be recorded for later analysis. Camera placement is designed to prevent / limit the capture of personally identifiable imagery.

Are there any risks or discomforts?
The risks associated with participating in this study are identified below.

1. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
2. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to participate in research in the Tiger Motors Lab. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
All volunteers recruited from the Auburn University community will be eligible for up to $100 in participation and performance related prizes. Odds of winning one of the eight available prizes will depend on the number of qualifying participants. Attendance and participation in the end of semester open house event is required for some of the prizes.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

If you have questions about this study, please ask now or contact Victoria Ballard at victoria.ballard@auburn.edu,360-632-1359, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Participant's signature	Date	Investigator obtaining consent	Date
Printed Name		Printed Name	

The Auburn University Institutional Review Board has approved this Document for use from
04/07/2023 to -------------
Protocol # 22-538 EP 2301

Appendix B - Recruiting Materials

In-Class Recruiting Script

Hello, Class.

Industrial Engineering graduate students pursuing their PhDs are recruiting participants for a research study. They are investigating the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. These investigations hope to better understand the relationships between those methods, learning effectiveness, and operator performance. A flyer with details of the study will be emailed to each of you. If you are interested, please follow up as described therein.

Email Script

Dear Student,

Please review the attached flyer, which provides details of the study recently described in __class name__. You are invited to participate in a research study on the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. The research team is conducting this study as Ph.D. Candidates under the supervision of Dr. Richard Sesek, Tim Cook Associate Professor in the Department of Industrial and Systems Engineering at Auburn University.

If you would like to participate, simply respond to this email or via text / phone to 407-399-3189. Questions or concerns can be directed to me through the same channels, or you may contact my advisor Dr. Sesek (sesek@auburn.edu).

Thank you for your consideration,

Email Script, Industry

Dear Manufacturing Professional,

You are invited to participate in a research study on the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. Please review the attached flyer for details. The research team is conducting this study as Ph.D. Candidates under the supervision of Dr. Richard Sesek, Tim Cook Associate Professor in the Department of Industrial and Systems Engineering at Auburn University.

If you would like to participate, simply visit the website. Questions or concerns can be directed to the research team at leanmanufacturingteam@auburn.edu, or their faculty advisor Dr. Richard Sesek (sesek@auburn.edu).

Thank you for your consideration,

Confirmation Email

Dear <student name>,

Thank you for your interest in our study, and for taking the time to discuss it with me. I'm happy to confirm that your trial is scheduled as follows:

Date and Time:<date and time>

Location: Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab), in the basement of the Shelby Center for Engineering Technology, room 0317, located at 345 W Magnolia Ave, Auburn, AL 36849

Please arrive on time. We anticipate that it will take 45-90 minutes to complete the session.

If you need to reschedule or have further questions, feel free to respond to this email or call / text me at 407-399-3189.

Thank you for your participation,

Invitation to Open House Event

Dear Participant,

Thank you for volunteering for our research studies. We are writing to invite you to an open house event on <DATE and TIME>. This is an opportunity for you to try some of the methods and technologies that you may not have experienced before, along with a variety of other software on the HoloLens 2 device.

All attendees will have additional chances to win cash or gift cards worth up to $100 by participating in a single experiment designed to test how well you retained the prior training.

Food and drink will also be provided.

Hope to see you there!

The Research Team

Flyers

On the following pages are flyers for both investigations, formatted as posters and slides.

1. First investigation, printed poster format.
2. Second investigation, printed poster format.
3. First investigation, manufacturing recruitment, printed poster format.
4. First investigation, displayed slide format.
5. Second investigation, displayed slide format.

Augmented Reality Research Study

Training methods for tomorrow's workforce, today!

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

This study is open to anyone 18 and older that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in in the Tiger Motors Lean Education Center (aka LEGO® Lab). It takes 45 to 60 minutes to complete.

💰💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰💰

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up: https://aub.ie/TigerMotorsResearch or scan the QR Code. Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

The Auburn University Institutional Review Board has approved this Document for use from 04/07/2023 to -------------
Protocol # 22-538 EP 2301

Studying Manufacturing with LEGO(R) Research

Participate in research in Auburn's Tiger Motors Lab!

The Effects of Lean Tools and Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials. Participants will assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours.

💰💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰💰

This study is open to anyone 18 and older.

TIGER MOTORS
LEAN EDUCATION CENTER

AUBURN
UNIVERSITY

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up: https://aub.ie/TigerMotorsResearch or scan the QR Code. Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

The Auburn University Institutional Review Board has approved this Document for use from
04/07/2023 to -------------
Protocol # 22-538 EP 2301

Manufacturing Volunteers Needed!

Augmented Reality Research Study

The Effects of Augmented Instruction on Manufacturing Assembly Training

Are You Interested in Augmented and Mixed Reality?
Do You Want to experience the latest in Projected and Head-Mounted AR?
Build LEGO® cars in Auburn's famous Lean Education Lab, *for Science!*

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

All volunteers will receive a $40 gift card as thanks for their participation.

This study is open to operators 18 and older, that aren't prone to motion sickness, and have no prior experience with similar AR systems or Auburn's Lean Education Center. For details, or to sign up, head to https://aub.ie/TigerMotorsResearch or scan the QR code below.

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Contact the research team at leanmanufacturingteam@auburn.edu with any other questions.

Scan to Sign Up!

The Auburn University Institutional Review Board has approved this Document for use from 04/07/2023 to -------------
Protocol # 22-538 EP 2301

Augmented Reality Research Study

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰

Open to anyone 18 and older that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in the Tiger Motors Lean Education Center (aka LEGO® Lab).

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up at https://aub.ie/TigerMotorsResearch or scan the QR Code.

Scan to Sign Up!

Microsoft HoloLens

LightGuide

Please contact the research team with any other questions:
leanmanufacturingteam@auburn.edu

The Auburn University Institutional Review Board has approved this Document for use from 04/07/2023 to -------------
Protocol # 22-538 EP 2301

Studying Manufacturing with LEGO®

The Effects of Lean Tools & Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials.

💰 **Participants are eligible for up to $100 in cash / gift card prizes!** 💰

You will be asked to assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours.

This study is open to anyone 18 and older.

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

Sign up at https://aub.ie/TigerMotorsResearch or scan the QR Code.

Please contact the research team with any other questions:
leanmanufacturingteam@auburn.edu

Scan to Sign Up!

TIGER MOTORS
LEAN EDUCATION CENTER

The Auburn University Institutional Review Board has approved this Document for use from
04/07/2023 to -------------
Protocol # 22-538 EP 2301

Appendix C - Data Collection Instruments

See attached, on the pages that follow.

Only the highlighted items changed and thus require restamp. The others are not included.

1. Subject Recruitment Data Sheet
2. Code Sheet
3. Participant Intake Sheet
4. Data Collection Sheet #1
5. Data Collection Sheet #2
6. Task Loading Index (NASA TLX)
7. System Usability Scale
8. Behavioral Control Survey
9. General Feedback

Participant Intake Sheet, p1 / 2

Participant #: __________ **Date: __________**

1. Gender:
   - ☐ Female
   - ☐ Male
   - ☐ Other
2. Age: _____
3. Race (select those with which you identify):
   - ☐ American Indian or Alaska Native
   - ☐ Asian
   - ☐ Black or African-American
   - ☐ Native Hawaiian or Other Pacific Islander
   - ☐ White
   - ☐ More than one race
   - ☐ Unknown or not reported
4. Ethnicity (select ONLY one with which you most closely identify):
   - ☐ Hispanic or Latino
   - ☐ Not Hispanic or Latino
   - ☐ Unknown or not reported
5. Country of Origin: ________________________
6. What language do you mainly speak at home?
   - ☐ English
   - ☐ Other
7. What is the highest level of school you have completed or the highest degree you have received?
   - ☐ Less than high school degree
   - ☐ High school degree or equivalent (e.g., GED)
   - ☐ Some college but no degree
   - ☐ Associate degree
   - ☐ Bachelor degree
   - ☐ Graduate degree: ____ Master or ____ PhD

Participant Intake Sheet, p2 / 2

Participant #: __________ **Date: __________**

8. If you are currently pursuing a degree, please complete the following:

   College (e.g. Education or Business): _______________________

   Program (e.g. MS Adult Ed or BS Accounting) : _______________________

9. Which of the following statements best describes your experience building LEGO models?

   ☐ I have little to no experience building LEGO models.

   ☐ I have some experience building LEGO models.

   ☐ I have lots of experience building LEGO models.

   ☐ I consider myself an expert in building LEGO models.

10. Please indicate your level of manufacturing experience

    ☐ I have no experience in manufacturing.

    ☐ I have taken one or more classes in manufacturing.

    ☐ I have held a part-time or temporary position in manufacturing.

    ☐ I have 1 or more years of experience working in manufacturing.

11. Do you normally wear corrective lenses? ____ Glasses ____ Contacts ____ Neither

    If yes, do you plan to wear them during this experiment? ____ Yes ____ No

12. Are you color blind? ____ Yes ____ No

13. Do you have any other condition that might affect your performance in this study?

    ____ Yes ____ No

14. What is your height? ____ feet ____ inches

15. Have you ever run an event of 5 kilometers or more? ____ Yes ____ No

16. How did you learn about this study? _______________________

Task Loading Index, p1 / 2

Participant #: __________ **Invest / Treat: __________** **Date: __________**

Sources of Workload Evaluation

Consider the following definitions:

Title	Range	Description
Mental Demand	Low / High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low / High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low / High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Perfect / Failure	How successful do you think you were in accomplishing the goals of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low / High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Low / High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

For each of the following pairs, circle the word that represents the more important contributor to workload for the specific task(s) you performed in this experiment.

Effort or **Performance**	**Temporal Demand** or **Frustration**	**Physical Demand** or **Performance**	**Temporal Demand** or **Mental Demand**	**Mental Demand** or **Physical Demand**
Temporal Demand or **Effort**	**Physical Demand** or **Frustration**	**Frustration** or **Effort**	**Performance** or **Mental Demand**	**Effort** or **Physical Demand**
Performance or **Frustration**	**Physical Demand** or **Temporal Demand**	**Performance** or **Temporal Demand**	**Mental Demand** or **Effort**	**Frustration** or **Mental Demand**

Task Loading Index, p2 / 2

Participant #: __________ **Invest / Treat: __________** **Date: __________**

Workload Rating Scales

For each of the following 6 questions, consider the assembly task you just completed. Record your immediate response to each item by putting an “X” at the point which matches your experience.

1. Mental Demand — How mentally demanding was the task?

   Very Low … Very High

2. Physical Demand — How physically demanding was the task?

   Very Low … Very High

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

   Very Low … Very High

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

   Perfect … Failure

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

   Very Low … Very High

6. Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

   Very Low … Very High

All Current IRB Stamped Documents

AUBURN UNIVERSITY INSTITUTIONAL REVIEW BOARD for RESEARCH INVOLVING HUMAN SUBJECTS

PROTOCOL REVIEW FORM
FULL BOARD or EXPEDITED REVIEW

For assistance, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu** Web Address: **http://www.auburn.edu/research/vpr/ohs**
Submit completed form and supporting materials as one PDF through the IRB Submission Page
Handwritten forms are not accepted. Where links are found hold down the control button (Ctrl) then click the link.

1. Proposed Start Date of Study:1/11/2023 Today's Date: January 4, 2023
Submission Status (Check One): ☐ New ☒ Revisions (to address IRB Review Comments)
Proposed Review Category (Check One): ☐ Full Board (greater than minimal risk) ☒ Expedited
If Expedited, Indicate Category(ies) ((Link to Expedited Category Review Sheet) 6

2. Project Title: The Effects of Augmented Instruction on Manufacturing Assembly Training

3. Principal Investigator (PI): Dan O'Leary Degree(s): BS Mech Eng, MS Eng Mgmt
Rank/Title: Graduate Student Department/School: Industrial & Systems Engineering
Role/responsibilities in this project: Organize and conduct research, perform data collection and analysis
Preferred Phone Number: 407-399-3189 AU Email: djo0008@auburn.edu

Faculty Advisor Principal Investigator (if applicable): Richard Sesek
Rank/Title: Associate Professor Department/School: Industrial & Systems Engineering
Role/responsibilities in this project: Supervise and advise the design and execution of the experiment
Preferred Phone Number: 334-728-1438 AU Email: rfs0006@auburn.edu

Department Head: Gregory Harris Department/School: Industrial & Systems Engineering
Preferred Phone Number: 334-844-1407 AU Email: gah0015@auburn.edu
Role/responsibilities in this project: Dissertation co-chair and primary project advisor

4. Funding Support: ☒ N/A ☐ Internal External Agency: n/a Pending ☐ Received ☐
For federal funding, list funding agency and grant number (if available): n/a

5. a) List any contractors, sub-contractors, and other entities associated with this project: n/a

b) List any other AU IRB approved protocols associated with this study and describe the association: n/a

c) List any other institutions associated with this study and submit a copy of their IRB approval(s): n/a

Protocol Packet Checklist

Check all applicable boxes. A completed checklist is required.

☒ **Protocol Review Form** (All required signatures included and all sections completed)
(Examples of appended documents are found on the website: https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)

☒ **CITI Training Certificates** for key personnel

☒ **Consent Form or Information Letter** and any releases (audio, video or photo) that participants will review and/or sign

☒ **Appendix A** "Reference List"

☒ **Appendix B** if e-mails, flyers, advertisements, social media posts, generalized announcements or scripts, etc., will be used to recruit participants.

☒ **Appendix C** if data collection sheets, surveys, tests, other recording instruments, interview scripts, etc. will be used for data collection. Attach documents in the order they are listed in item 13c.

Continued on Page 2

☒ **Appendix D** if they study will use a debriefing form or will include emergency plans/ procedures and medical referral lists. (A referral list may be attached to the consent document.)

The Auburn University Institutional Review Board has approved this Document for use from 01/28/2023 to ------------
Protocol # 22-538 EP 2301

☐ **Appendix E** if research is being conducted at sites other than Auburn University or in cooperation with other entities. A **permission letter** from the site/ program director must be included indicating their cooperation or involvement in the project. NOTE: If the proposed research is a multi-site project, involving investigators or participants at other academic institutions, hospitals or private research organizations, a letter of **IRB approval** from each entity is required prior to initiating the project.

☐ **Appendix F** Written evidence of approval by the host country, local IRB or institutions if research is conducted outside the United States

6. General Research Project Characteristics

6A. Research Methodology

Check all descriptions that best apply to the research methodology.

Data Source(s): ☒ New Data ☐ Existing Data

Will recorded data directly or indirectly identify participants?
☒ Yes ☐ No

Data collection will involve the use of:

☒ Educational Tests (cognitive diagnostic, aptitude, etc.)
☒ Interview
☒ Observation
☐ Locations or Tracking Measures
☐ Physical / Physiological Measures or Specimens
☒ Surveys / Questionnaires
☐ Other: Click or tap here to enter text.

☒ Internet / Electronic
☒ Audio
☒ Video
☐ Photos
☐ Digital Images
☐ Private records or files

6B. Participant Information

Check all descriptors that apply to the TARGET population.
(link to definition of target population)
☐ Males ☐ Females ☐ AU students

Vulnerable Populations
☐ Pregnant Women/Fetuses ☐ Prisoners ☐ Institutionalized
☐ Children and / or Adolescents (under age 18 in AL; if minor participants, at least 2 adults must be present during all research procedures that include the minors)

Persons with:
☐ Economic Disadvantages ☐ Physical Disabilities
☐ Educational Disadvantages ☐ Intellectual Disabilities

Will participants be compensated? ☐ Yes ☒ No

6C. Risks to Participants

Identify all risks participants might encounter in this research.

☒ Breach of Confidentiality* ☐ Coercion
☐ Deception ☐ Physical
☐ Psychological ☐ Social
☐ None
☒ Other (COVID-19, other medical):
COVID-19 Exposure; Discomfort, including possibility of mild nausea, see section 14

*Note that if the investigator is using or accessing confidential or identifiable data, reach of confidentiality is always a risk.

6D. Corresponding Approval/ Oversight

- **Does the study include participant exposure to radiation?** ☐ **Yes** ☒ **No**
  If yes indicate: ☐ **DEXA** ☐ **PQCT** ☐ **Other**

- **Is IBC Approval required for this study?**
  ☐ **Yes** ☒ **No**

  If yes, BUA # Click or tap here to enter text. **Expiration Date** Click or tap to enter a date.

- **Is IACUC Approval required for this study?**
  ☐ **Yes** ☒ **No**

  If yes, PRN # Click or tap here to enter text. **Expiration Date** Click or tap to enter a date.

- **Does this study involve the Auburn University MRI Center?**
  ☐ **Yes** ☒ **No**

  Which MRI(s) will be used for this project? (Check all that apply)
  ☐ **3T** ☐ **7T**

Continued on Page 3

Does any portion of this project require review by the MRI Safety Advisory Council?

☐ **Yes** ☒ **No**

Signature of one MRI Center Representative:____
Required for all projects involving the AU MRI Center
Appropriate MRI Center Representatives:
Dr. Thomas S. Denney, Director AU MRI Center
Dr. Ron Beyers, MR Safety Officer

7. Project Assurances

7A. Principal Investigator's Assurances

1. I certify that all information provided in this application is complete and correct.
2. I understand that, as Principal Investigator, I have ultimate responsibility for the conduct of this study, the ethical performance this project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the Auburn University IRB.
3. I certify that all individuals involved with the conduct of this project are qualified to carry out their specified roles and responsibilities and are in compliance with Auburn University policies regarding the collection and analysis of the research data.
4. I agree to comply with all Auburn policies and procedures, as well as with all applicable federal, state, and local laws regarding the protection of human subjects, including, but not limited to the following:
   a. Conducting the project by qualified personnel according to the approved protocol
   b. Implementing no changes in the approved protocol or consent form without prior approval from the Office of Research Compliance
   c. Obtaining the legally effective informed consent from each participant or their legally responsible representative prior to their participation in this project using only the currently approved, stamped consent form
   d. Promptly reporting significant adverse events and / or effects to the Office of Research Compliance in writing within 5 working days of the occurrence.
5. If I will be unavailable to direct this research personally, I will arrange for a co-investigator to assume direct responsibility in my absence. This person has not been named as co-investigator in this application, or I will advise ORC, by letter, in advance of such arrangements.
6. I agree to conduct this study only during the period approved by the Auburn University IRB.
7. I will prepare and submit a renewal request and supply all supporting documents to the Office of Research Compliance before the approval period has expired if it is necessary to continue the research project beyond the time period approved by the Auburn University IRB.
8. I will prepare and submit a final report upon completion of this research project.

My signature indicates I have read, understand and agree to conduct this research project in accordance with the assurances listed above.

____Dan O'Leary_________________	____________________________	1/4/2023
Principal Investigator Name	**Principal Investigator Signature**	**Date**

7B. Faculty Advisor / Sponsor's Assurances

1. I have read the protocol submitted for this project for content, clarity, and methodology.
2. By my signature as faculty advisor / sponsor on this research application, I certify that the student or guest investigator is knowledgeable about the regulations and policies governing research with human subjects and has sufficient training and experience to conduct this particular study in accord with the approved protocol.
3. I agree to meet with the investigator on a regular basis to monitor study progress. Should problems arise during the course of the study, I agree to be available, personally, to supervise the investigator in solving them.
4. I assure that the investigator will promptly report significant incidents and / or adverse events and / or effects to the ORC in writing within 5 working days of the occurrence.
5. If I will be unavailable, I will arrange for an alternate faculty sponsor to assume responsibility during my absence, and I will advise the ORC by letter of such arrangements. <u>If the investigator is unable to fulfill requirements for submission of renewals, modifications or the final report, I will assume that responsibility.</u>

____Richard Sesek___________________	____________________________	1/4/2023
Faculty Advisor / Sponsor Name	**Faculty Advisor Signature**	**Date**

Continued on Page 4

7C. Department Head's Assurance

By my signature as department head, I certify that I will cooperate with the administration in the application and enforcement of all Auburn University policies and procedures, as well as all applicable federal, state, and local laws regarding the protection and ethical treatment of human participants by researchers in my department

Gregory Harris	Gregory A. Harris	1/4/23
Department Head Name	**Department Head Signature**	**Date**

8. Project Overview:

8A. A summary of relevant research findings leading to this research proposal:
(Cite source; include a "Reference List" as ***Appendix A****.)*

Augmented Reality (AR) systems "combine real and virtual, are interactive in real time, and are registered in 3-D" [1]. By realistically integrating informative and/or interactive virtual objects in our view of the world, AR aims to enhance the users' interaction with and perception of it. Its essential affordance is the direct and natural manipulation of virtual objects in everyday surroundings. Relative to metaphorical digital interfaces, this is thought to improve the uptake of knowledge by reducing the overall cognitive load and better distributing it across multiple sensory pathways [2]. AR-assisted learners demonstrate improved perception, performance, and understanding of spatial concepts, with outcomes correlated to the amount of physical engagement involved [3]. As a result, AR is thought to be well-suited for task-related learning. Using untethered, hands-free devices with optical see-through head-mounted displays, AR can continuously enhance the user's actions in the real world [4]. These benefits have broad industrial applications.

In manufacturing, operator support has been a common application of AR research and development since the early 1990s [5]. It is also seen as a source of innovative operator training methods required to meet rapidly increasing demand for skilled labor due to high retirement rates, global expansion, and increasing specialization [6]. Manufacturing support, training, and related applications have been identified in the areas of assembly, maintenance, operations, quality control, safety, design, visualization, logistics, and marketing [7].

Despite great potential, the adoption of AR is slowed by technical, market, and other important social and legal obstacles [8]. To successfully transition from research projects and proof of concepts and gain widespread adoption in manufacturing, AR must demonstrate a worthwhile return on investment [9; 10]. But AR remains a highly fragmented market, including a diverse selection of screen-based, projected, and head-mounted technologies [6]. Studies show that the efficacy of these systems varies with the task type, technology used, application design, and other factors [11]. Thus, the success rate of AR adoption in industry would be improved by frameworks for strategic decision making based on quantified benefits in various scenarios [12–14]. Research in this area is young but accelerating. Most of it focuses on efficiency (task time) and accuracy (error count). These are relevant but incomplete measures for assessing training outcomes, where the learning rate and transfer effectiveness must also be considered [15]. This study extends prior work [16] to explore the relationship between a variety of AR technologies and their underlying affordances [17] and learning outcomes for manufacturing assembly operations. By controlling for the task type and application design we hope to better understand the relative value of these systems, filling in important gaps that can lead to a cohesive framework for successful adoption.

8B. A brief summary/abstract of the study methodology, including design, population, and variables of interest.
(350 word maximum, in language understandable to someone who is not familiar with your area of study. Note this summary/abstract can be used to prepare the concise summary in the consent document.):

This experiment will be conducted in the Tiger Motors Lean Education Center, which simulates automotive manufacturing best practices using LEGO® cars. Participants will act as operators assembling the SUV car at station 8. This process has been used thousands of times in INSY 5/6800 without significant incident.

Figure 1 - Work Station 8

Figure 2- LEGO Speedster Assembly

A population of 40-60 adults will be recruited from Auburn University. Candidates with experience using head-mounted or projected AR or building cars in the Lean Lab will be excluded. Participants in this between-groups design will experience a single level of the Instructional Media Type (IMT) treatment, with increasingly augmented work instructions:

1. Paper Work Instructions (PWI): traditional printed instructions (control)
2. Projector Augmented Reality (PAR): interactive instructions projected on the work surface via the LightGuide system with a stationary model
3. Head-Mounted Display AR (HMDAR): interactive instructions presented in the user's field of view using the HoloLens2 (HL2) HMD with a stationary model
4. HMD Mixed Reality (HMDMR): extends the third treatment by leveraging advanced capabilities of the HL2, allowing for more natural interactions and movement of the model

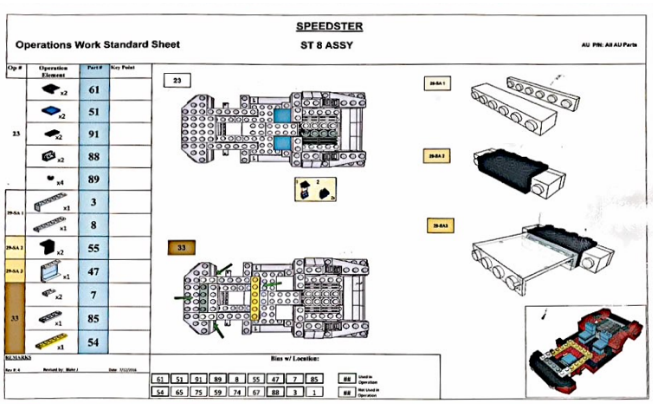


Figure 3 – Paper Work Instructions for Station 8

Figure 4- LightGuide Work Instructions

Participant groups will be set randomly. We hypothesize that HDMR will outperform other treatments in accuracy-based performance measures, as well as learning rate and transfer. In contrast, we expect participants assigned the PWI treatment to have the best times.

Figure 5 – HoloLens2 Wireless, See-Through Design

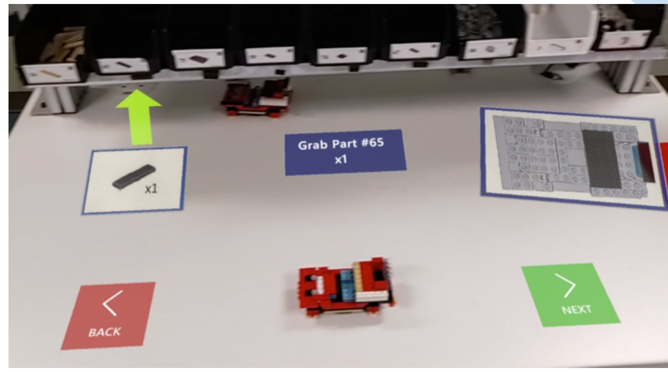


Figure 6- HoloLens2 Work Instructions, 1st Person View

First, participants will be shown how to interpret paper work instructions and use them to construct a sample LEGO assembly. Next, those assigned to an AR treatment level are given a brief introduction to its operation. Questions are allowed throughout this process.

The hypotheses are then tested in two phases. The first compares the effects of instructional media on the speed (task completion time) and accuracy (number and type of corrected and uncorrected errors) with which participants perform each repetition of the task. These measures are tracked for each assembly completed in the 10-minute session, allowing us to assess learning rates.

During the second phase, participants repeat the task four times in the control condition while the same measures are observed. Their results in each phase will be analyzed to compare transfer effectiveness between treatments.

9. Purpose

9A. State the purpose of the study and all research questions or aims. (Include a sentence that begins, "The purpose of this study is…")
The purpose of this study is to measure the effect of instructional media type (IMT) on learning rates and skills transfer for industrial assembly tasks. The first phase will help us understand how each IMT affects the operator's learning rate (time or cycles to learn the process) and ultimate measures of performance (speed and accuracy). The second will help assess how learning transfer varies with each treatment.

9B. Describe how results of this study will be used? (e.g., presentation? publication? thesis? dissertation?)
The data collected during this project will be used for thesis and dissertations, scholarly publications and presentations, and grant proposals.

10. Key Personnel. Describe responsibilities as specifically as possible. Include information on research training or certifications related to this project. **To determine key personnel see decision tree at [https://cws.auburn.edu/OVPR/pm/compliance/irb/training.](https://cws.auburn.edu/OVPR/pm/compliance/irb/training) Submit a copy of CITI training documentation for all key personnel.** (For additional personnel, add lines as needed).

To determine Auburn University HIPAA – covered entities click link to HIPAA Policy.

If any key personnel have a formal association with institutions/entities involved in the study (for example is an employee or supervisor at the site research will occur), describe that affiliation. For all non-AU affiliated key personnel, submit a copy of their IRB approval.

Principal Investigator: Dan O'Leary **Rank/Title:** Graduate Student
Email Address: djo0008@auburn.edu **Degree(s):** BS ME, MS Eng Mgmt
Dept / Affiliation: Industrial & Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**
Roles / Responsibilities: Overall responsibility for the project, including design and administration of experiments, coordinating recruitment, obtaining consent, and data collection and analysis.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have

influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 8/26/2025

Individual: Richard Sesek **Rank/Title:** Associate Professor
Email Address: rfs0006@auburn.edu **Degree(s):** BS, MS, MPH, PhD
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: Advise, oversee, and assist with experiment design, IRB review process, obtaining consent, conducting trials, data collection and analysis.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 4/25/2023
Choose a course Expiration Date

Individual: Gregory Harris **Rank/Title:** Associate Professor
Email Address: gah0015@auburn.edu **Degree(s):** PhD
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: Dissertation co-chair and primary advisor
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 5/12/2024
Choose a course Expiration Date

Individual: Victoria Ballard **Rank/Title:** Graduate Student
Email Address: vzb0024@auburn.edu **Degree(s):** BS CHE, MS CivE
Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes** ☐ **No** ☒
Roles / Responsibilities: Lab manager, design and conduct research
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**
- Plan for IRB approval for non-AU affiliated personnel? **n/a**
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: **n/a**
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 2/9/2025
Choose a course Expiration Date

Revised 07/12/2022

Individual: Md Monir Hossain **Rank/Title:** Graduate Student

Email Address: mzh0116@auburn.edu **Degree(s):**BS BE, MS TM, MS ISE

Dept. / Affiliation: Industrial and Systems Engineering **HIPAA Covered Entity? Yes ☐ No ☒**

Roles / Responsibilities: Lab assistant, design and conduct research

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: **n/a**

- Plan for IRB approval for non-AU affiliated personnel? **n/a**

- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No

- If yes, briefly describe the potential or real conflict of interest: **n/a**

- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.

- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 8/29/2025

Choose a course Expiration Date

11. Location of research.

11A. List all locations where data collection will occur. If applicable, attach permission letters as Appendix E. (School systems,
organizations, businesses, buildings and room numbers, servers for web surveys, etc.) **Be as specific as possible.** (See sample letters at https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)

Data collection will take place at the Lean Lab in the basement of the Shelby Center for Engineering Technology, room 0317, located at 345 W Magnolia Ave, Auburn, AL 36849

11B. Will study data be stored within a HIPAA covered facility? Yes ☐ No ☒

If yes, which facility(ies) (To determine AU HIPPA covered entities, go to VII of the HIPPA Hybrid Entity Policy)**:**
n/a

12. Participants (If minor participants, at least 2 adults must be present during all research procedures that include the minors.)

12A. Describe the targeted/ intended participant population for the study. Include the anticipated number of participants and inclusion and exclusion criteria and the procedures to ensure more than 1 adult is present during all research procedures which include the minor.

☐ Check here if existing data will be used and describe the population from whom data was collected including the number of data files.

☐ Check here if permission to access existing data is required and submit a copy of the agreement to access.

Between 40 and 60 subjects will be recruited from the Auburn University community. Potential participants will be screened for exclusion based on the following: 1. Under 18 years of age 2. Prone to motion sickness 3. Prior experience with head-mounted or projected AR systems 4. Prior experience building cars in the Lean Lab as part of INSY 5800/6800 or otherwise Active recruiting efforts will focus on freshman and sophomore engineering students in Industrial & Systems Engineering (ISE), as they are accessible and are likely to meet all requirements.

12B. Describe, step-by-step in lay language all procedures to recruit participants. Include in Appendix B a copy of all e-mails, flyers, advertisements, recruiting scripts, invitations, etc., that will be used to invite people to participate. (See sample documents at https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs)

Students and Faculty will be recruited using flyers distributed around the Auburn University campus. Additionally, ISE students will be recruited via in-class announcements and the distribution of emails. Copies of each are included in Appendix B. Interested participants will be instructed to contact the PI for more information. In the call that follows, the PI will: 1. Briefly explain the study, recapping and elaborating on the recruiting materials 2. Explain the exclusion criteria and identify relevant issues for the candidate 3. Set expectations for participant involvement, including time commitment and tasks 4. Answer any questions the candidate has regarding participation in the study If the candidate is ready and willing to proceed, their information will be collected using the Subject Recruitment Data Sheet provided in Appendix C. They will be assigned a unique participant ID and a date and time for data collection. If interest in the study exceeds capacity, additional participants will be thanked for their interest and informed that enrollment is limited. They will be given the option to remain "waitlisted" if additional participants or follow-up studies are required.

12C. Minimum number of participants required to validate the study? 40

Number of participants expected to enroll? About 50

Provide the rationale for the number of participants. Appropriate for the desired power given the number of treatments.

Is there a limit to the number of participants that will be included in the study?
☐ **No** ☒ **Yes, the number is** 60, due to time constraints

12D. Describe the process to compensate, amount and method of compensation and/or incentives for participants. [AU Procurement and Business Services (PBS) policies]
(benefits to participants are NOT compensation)

If participants will not be compensated, check here: ☒
Indicate the amount of compensation per procedure and in total: Click or tap here to enter text.
Indicate the type of compensation: ☐ Monetary ☐ Incentives
☐ Raffle or Drawing incentive (Include the chances of winning.)
☐ Extra Credit (State the value)
☐ Other

Describe how compensation will be distributed (USPS, email, etc.): Click or tap here to enter text.

13. Project Design & Methods

13A. Describe, step-by-step, all procedures and methods that will be used to consent participants. If a waiver is being requested, indicate the waiver, and describe how the study meets the criteria for the waiver. If minors will be enrolled describe the process to obtain parental/ legally authorized guardian permission.

☐ **Waiver of Consent (including using existing data)**
☐ **Waiver of Documentation of Consent (use of Information Letter)**
☐ **Waiver of Parental Permission (for college students 18 years or younger)**

As each participant arrives, they will be welcomed and given brief introductions to members of the team administering the study. We will then ask them to review the consent document, encouraging them to ask any questions they have. After a verbal confirmation that the participant has read and is satisfied with the terms of this document, we will ask that they sign and date it.

13B. In lay language, understandable by someone not familiar with the area of study, describe the complete research design and methods that will be used to address the purpose. Include a clear description of who, when, where and how data will be collected. Include specific information about participants' time and effort.

Following the recruitment, eligibility screening, and consent processes described above, a short orientation process acclimates the participant to the work area and emergency procedures are described. A research associate will point out the key features of a work cell (work surface, part bins, etc.), describe how to interpret the paperwork instructions, demonstrate typical assembly steps, and answer any relevant questions. (5-10 mins)

Next, participants assigned to any AR IMT (PAR, HMDAR, or HMDMR) will receive a brief demonstration of its basic operation. In all cases, the participant will be shown how to use the appropriate forward and back triggers, and how the system signals instructions and feedback related to part bin and placement. PAR and HMDAR users will be instructed that the model must remain in the fixture. HMDMR users will understand that the model can be freely manipulated during assembly. (5-10 mins)

Once orientation and training are complete, the experiment is conducted in two phases. Regardless of IMT assigned, all participants will wear the HL2 during both phases to control for its effects and allow us to record each session from their POV.

In the first phase, participants will be asked to complete the assembly process for as many cars as they can, while learning the steps and limiting the number of errors produced. This phase will be conducted with the support of the assigned IMT and will last 10 minutes. During that time, we expect that each participant will produce between 3 and 6 cars, based on prior performance data and the 60-second takt time for which the instructions were designed. (10 mins)

Following a short break to reset the workstation, the second phase will begin. In this phase each participant will build 4 more cars using only paper work instructions. Their stated goal will be to deliver error-free results quickly, while referencing the instructions only when necessary. (5-10 mins)

Participant performance in both phases will be recorded on two cameras, one first-person view from onboard the HL2, and one third-person view from a camera mounted nearby. Experimental data will be derived from subsequent analysis of these videos. Participants will not be allowed to ask questions during either data collection phase of the experiment.

Once the experiment is concluded, each participant will complete an exit survey that incorporates the NASA TLX and System Usability Scale instruments, along with the Adult ADHD Self-Report Scale (ASRSv1.1). It also includes a section for open-ended feedback. When the survey is collected a research associate will ask if the participant experienced any injury and if they are interested in attending a follow-up session for more in-depth exploration of the HoloLens2. Their responses will be recorded on the exit survey. (5-10 mins)

We conservatively estimate a total time commitment of 45-60 minutes for each participant.

13C. List all data collection instruments used in this project, in the order they appear in Appendix C.
(e.g., surveys and questionnaires in the format that will be presented to participants, educational tests, data collection sheets, interview questions, audio/video taping methods etc.)

1. Subject Recruitment Data Sheet: filled out during the screening call; includes the exclusion checklist, participant number, basic demographics (age and gender), and date / time of scheduled trial
2. Code Sheet: collects the personally identifiable data for eligible participants, including name, contact info (phone, email) and subject number
3. Data Collection Sheet: consists of general notes from the experiment and data derived from subsequent analysis of video recordings
4. Exit Survey: incorporates the NASA TLX and System Usability Scale instruments, open-ended feedback, and area for research associate to indicate answers about participant injury and interest in follow-up session

13D. Data analysis: Describe how data will be analyzed. If a data collection form (DCF) will be used, submit a copy of the DCF.

In both phases of this study, the independent variable is treatment type, and the dependent variables are task completion time and number of errors. The dependent variables will be recorded for each car completed in both sessions.

Data will be analyzed with a combination of visual (e.g., box plots) and statistical methods. Methods based on analysis of variance (ANOVA) will be used to test the stated hypotheses. Additional analysis will be done to explore the relationship between other variables of interest, including demographics, mental workload, behavioral control, and system usability with the measured outcomes.

13E. List any drugs, medications, supplements, or imaging agents that participants will ingest/ receive during participation in the study or indicate not applicable (N/A).

n/a

14. Risks & Discomforts: List and describe all the risks participants may encounter in this research including risks from item 6d of this form, in this research. <u>If deception will be part of the study, provide the rationale for the deception, describe the debriefing process, and attach a copy of the debriefing form that will be used as Appendix D.</u> (Examples of possible risks are in section #6C)

1. Physical Discomfort: All participants will be required to wear the HoloLens2 device, regardless of treatment group to control for its effects on user fatigue, etc., and to allow us to record a first-person view of their

session. As a result, they may experience mild physical discomfort including neck strain after prolonged use. The limited duration of this study should mitigate this effect.

2. Vestibular and Visual Discomfort: Participants assigned to the HMDAR and HMDMR treatments will experience display technology that may cause mild dizziness, eye strain, and related effects. Owing to the see-through design of the HoloLens2 device these effects are less common and less pronounced than seen in fully immersive Virtual Reality (VR) headsets.
3. Trip and Impact Risk: Any head-mounted display can reduce the wearer's peripheral vision and otherwise impact their natural field of view. Consequently, they may become more susceptible to tripping over or running into things around them. This risk is minimized by the HoloLens2's design, which offers a very wide, minimally obscured field of view. Furthermore, the HL2 is a standalone device, so there is no risk of tripping over a cord. Additionally, the participant is generally stationary in an environment free of obstruction. Finally, the Lean Lab is a clean, organized, safe, and well-lit environment with no history of related hazards.
4. Breach of Confidentiality Risk: All resulting data will be anonymized, and video of each session will be recorded from the first person and top-down angles to prevent participant exposure. That said, subjects could be seen entering, leaving, or during the experiment. All of these create a small possibility that subjects could be identified, inadvertently breaching their confidentiality. Additionally, there is the possibility that the subject code list, which connects each participant's identity with their experimental data, could be obtained. Mitigation methods for this risk are described in section 17 Protection of Data.
5. Psychological Discomfort: Due to the nature of the experiment, some participants may experience mild psychological discomfort induced by its time and performance-based measures. Participants will be told that their objective is to learn to perform the task quickly and error free. Otherwise, no overt pressure is put on the subjects to perform. Given that the outcome of their performance has no impact on their life outside the experiment, any related psychological discomfort should be minimal and short-lived.
6. COVID-19 Exposure: This study will be a Category C study with no High-Risk Procedures or Participants. Precautions will be implemented using the COVID-19 2022 Precautions Matrix to determine appropriate precautions at the time of data collection(s) for a Category C study. All work surfaces and the HMD will be wiped down before and after each participant. Necessary supplies will be made available, including as masks, hand sanitizer (60%+ alcohol), tissues, paper towels, trash baskets, and cleaners / disinfectants. All research participants will follow the University's guidance on self-screening. At the time of this writing, the CDC's COVID-19 community level for Lee County, Alabama is LOW, so participant screening is not required. The Shelby Center for Engineering Technology, where this protocol will be administered, is assigned the highest level of building readiness due to increased air turn-over and filtration. Further details and resources can be found in Appendix D.

15. Precautions / Minimization of Risks

15A. Identify and describe all precautions that will be taken to eliminate or reduce risks listed in items 6.c. and 14. If participants can be classified as a "vulnerable" population, describe additional safeguards that will be used to assure the ethical treatment of vulnerable individuals. **If applicable, submit a copy of any emergency plans/procedures and medical referral lists in Appendix D. (Sample documents can be found online at** https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs **precautions)**

This study does not involve any vulnerable populations. Please see section 14, where the primary mitigations are described for each identified risk. Additionally, all participant activities will be supervised and monitored for relevant symptoms. If any participant experiences dizziness or related vestibular issues, or any other significant but unexpected side-effect, we will suspend the experiment, remove the HMD, have them sit and offer drinking water while assessing the situation. If escalation is required, the emergency plan and contact list is included in Appendix D. During the debriefing all participants will be asked if they were injured or experienced any discomfort during their trials. The debriefing also serves to keep each participant under our supervision long enough to ensure no lingering or delayed effects.

15B. If the internet, mobile apps, or other electronic means will be used to collect data, describe confidentiality and/or security precautions that will be used to protect (or not collect) identifiable data? Include protections used during collection of data, transfer of data, and storage of data. If participant data may be obtained and/or stored by apps during the study, describe.

n/a

15C. Does this research include purchase(s) that involve technology hardware, software or online services?

☐ **YES** ☒ **NO**

If YES:

A. **Provide the name of the product** Click or tap here to enter text.
and the manufacturer of the product Click or tap here to enter text.

B. **Briefly describe use of the product in the proposed human subject's research.**
Click or tap here to enter text.

C. **To ensure compliance with AU's Electronic and Information Technology Accessibility Policy, contact AU IT Vendor Vetting team at vetting@auburn.edu to learn the vendor registration process (prior to completing the purchase).**

D. **Include a copy of the documentation of the approval from AU Vetting with the revised submission.**

15D. Additional Safeguards

Will DEXA, pQCT, or other devices which emit radiation be used? ☐ Yes ☒ No
If yes, the IRB will notify the Auburn Department of Risk Management and Safety, who will contact the Alabama Department of Public Health (ADPH) and secure approval. Research which includes device(s) which emit radiation may NOT be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of ADPH approval.

Will a Certificate of Confidentiality (CoC) issued by NIH be obtained ☐ Yes ☒ No If yes, include CoC language in consent documents and include the documentation of CoC approval. Research which includes a CoC may not be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of CoC approval. AU Required CoC Language

Is the study a clinical trial? ☐ Yes ☒ No
If yes, provide the National Clinical Trial (NCT) # Click or tap here to enter text. **and include required clinical trial information in all consent documents. AU Clinical Trial Information**

16. Benefits

16A. List all realistic direct benefits participants can expect by participating in this study. (Compensation is not a benefit) If participants will not directly benefit check here. ☒

There are no direct benefits for participants in this study. It will offer all of them an opportunity to interact with projection and/or head-mounted AR hardware and training methods for the first time. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

16B. List realistic benefits for the general population that may be generated from this study.

Turnover in the workforce and the lack of skilled labor necessitates scalable, efficient training methods. Furthermore, the shift from mass production to mass customization forces operators to contend with wide variance in the assembly steps required at each workstation. Together, these trends demand innovative methods for operator training and support.

Augmented and mixed reality are expected to help fill that need, but it is a fragmented market with a variety of solutions. Few studies explore the relationship between those methods (and the affordances that differentiate them) and corresponding learning rates and transfer. We believe this study will make meaningful contributions to that effort, helping to build a cohesive understanding of the utility of these systems and best practices for their application.

17. Protection of Data

17A. Data are collected:

☐ **Anonymously with no direct or indirect coding, link, or awareness by key personnel of who participated in the study (skip to item E)**

☐ **Confidentially, but without a link to participant's data to any identifying information (collected as "confidential" but recorded and analyzed "anonymous") (Skip to item E).**

☒ **Confidentially with collection and protection of linkages to identifiable information.**

17B. If data are collected with identifiers and coded or as coded or linked to identifying information, describe the identifiers and how identifiers are linked to participants' data.

In addition to the consent form, a code list will be maintained that includes identifying data of each participant (name, contact information, and ID number). This will be linked to all data collection forms by the participant number. The consent forms and code list will be maintained on paper only, to facilitate secure storage and disposal (shredding). The consent form will not include reference to the participant's ID number. Only the code list will directly connect participants to their data.

The video recordings may also allow for participants to be identified, though the first-person recording will not allow a view of their face and the third-person view will focus on the work area. If the recorders do not provide a video-only option, audio from those sessions, which may also provide identifying data, will be stripped from the recordings before storage.

17C. Provide the rationale for need to code participants' data or link the data with identifying information.

Only for the purpose of contacting participants while the protocol is open. Once completed, the code list will be destroyed, making the data anonymous.

17D. Describe how and where identifying data and/or code lists will be stored. (Building, room number, AU BOX?) **Describe how the location where data is stored will be secured. For electronic data, describe security measures. If applicable, describe where IRB-approved and participant signed consent documents will be kept on campus for 3 years after the study ends.**

Signed consent forms and the code list will be kept in a secure, locked file in office 3301J of Shelby Center.

17E. Describe how and where data will be stored (e.g., hard copy, audio/ visual files, electronic data, etc.), and how the location where data is stored is separated from identifying data and will be secured. For electronic data, describe security. Note use of a flash drive or portable hard drive is not appropriate if identifiable data will be stored; rather, identifying participant data must be stored on secured servers.

All electronic data pertaining to the study will be stored on a secured server. Non-identifiable data will be available to other members of the research team.

17F. List the names of all who will have access to participants' data? (If a student PI, the faculty advisor must have full access and be able to produce study data in the case of a federal or institutional audit.)

- Consent forms and code list: Dan O'Leary, Gregory Harris
- Non-identifiable data: full research team, by request

17G. When is the latest date that identifying information or links will be retained and how will that information or links be destroyed? (Check here if only anonymous data will be retained ☒)

August 2023

AUBURN UNIVERSITY HUMAN RESEARCH PROTECTION PROGRAM (HRPP)

REQUEST for MODIFICATION

For Information or help completing this form, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: IRBAdmin@auburn.edu

- **Federal regulations require IRB approval before implementing proposed changes.**
- **Change means any change, in content or form, to the protocol, consent form, or any supportive materials (such as the investigator's Brochure, questionnaires, surveys, advertisements, etc.). See Item 4 for more examples.**

1. Today's Date	2/6/2023

2. Principal Investigator (PI) Name: Dan O'Leary			
PI's Title:	Instructor / PhD Candidate	**Faculty PI (if PI is a student):**	Dr. Richard Sesek
Department:	Industrial & Systems Eng	**Department:**	Industrial & Systems Eng
Phone:	407-399-3189	**Phone:**	334-728-1438
AU-E-Mail:	djo0008@auburn.edu	**AU E-Mail:**	rfs0006@auburn.edu
Contact person who should receive copies of IRB correspondence (Optional):	Click or tap here to enter text.	**Department Head Name:**	Dr. Gregory Harris
Phone:	Click or tap here to enter text.	**Phone:**	334-844-1407
AU E-Mail:	Click or tap here to enter text.	**AU E-Mail:**	gah0015@auburn.edu

3. AU IRB Protocol Identification		
3.a. Protocol Number: 22-538		
3.b. Protocol Title: The Effects of Augmented Instruction on Manufacturing Assembly Training		
3. c. Current Status of Protocol – For active studies, check ONE box at left; provide numbers and dates where applicable		
☒	**Study has not yet begun; no data has been entered or collected**	
☐	**In progress If YES, number of data/participants entered:** Click or tap here to enter text.	**Current Approval Dates From:** 1/30/2023
☐	**Is this modification request being made in conjunction with/as a result of protocol renewal?** ☐ **YES** ☐ **NO**	
☐	**Adverse events since last review If YES, describe:** Click or tap here to enter text.	**To:** Click or tap to enter a date.
☐	**Data analysis only**	
☐	**Funding Agency and Grant Number:** Click or tap here to enter text.	**AU Funding Information:** Click or tap here to enter text.
☐	**List any other institutions and/ or AU approved studies associated with this project:** Click or tap here to enter text.	

The Auburn University Institutional Review Board has approved this Document for use from
02/09/2023 to --------------
Protocol # 22-538 EP 2301

4. Types of Change **Mark all that apply, and describe the changes in item 5**	
☒	**Change in Key Personnel** List the name(s) of personnel being added to or removed from the study and attach a copy of the CITI documentation for personnel being added to the study. Adding: Dr. Gregory Purdy, Diego Caputo Rodriguez, Alex Barras, David "Brown" Teague, Carson Tillery
☐	**Additional Sites or Change in Sites, including AU classrooms, etc.** Attach permission forms for new sites.
☐	**Change in methods for data storage/ protection or location of data/ consent documents**
☐	**Change in project purpose or project questions**
☐	**Change in population or recruitment** Attach new or revised recruitment materials as needed; both highlighted version & clean copy for IRB approval stamp
☒	**Change in study procedure(s)** Attach new or revised consent documents as needed; both highlighted revised copy & clean copy for IRB approval stamp No change is required to the consent documents.
☒	**Change in data collection instruments/forms (surveys, data collection forms)** Attach new forms as needed; both highlighted version & clean copy for IRB approval stamp Attached.
☐	**Other** (BUAs, DUAs, etc.) Indicate the type of change in the space below, and provide details in the Item 5.c. or 5.d. as applicable. Include a copy of all affected documents, with revisions highlighted as applicable. Click or tap here to enter text.

5. Description and Rationale
5.a. For each item marked in Question #4 describe the requested change(s) to your research protocol, and the rationale for each.
Needed added team members to help run the protocol. Minor changes to streamline procedure. Added an expanded demographics form. Asking participants to repeat the NASA TLX and SUS instruments after each phase (twice total, one additional time).
5.b. Briefly list (numbered or bulleted) **the activities that have occurred up to this point, particularly those that involved participants.**
Only initial recruiting. No trials run or scheduled yet.
5.c. Does the requested change affect participants, such as procedures, risks, costs, benefits, etc.
No. Added surveys may add a little time but that was offset by streamlined procedure.
5.d. Attach a copy of all "IRB stamped" documents currently used. (Information letters, consent forms, flyers, etc.)
Attached.
5.e. List all revised documents and attach two copies of the revised documents – one copy which highlights the revisions and one clean copy of the revised documents for the IRB approval stamp.
Attached.

6. Signatures
Principal Investigator: ______________________________
Faculty Advisor PI, if applicable: ______________________________

Version Date: Click or tap to enter a date.

AUBURN UNIVERSITY HUMAN RESEARCH PROTECTION PROGRAM (HRPP)

REQUEST for MODIFICATION

For Information or help completing this form, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu**

- **Federal regulations require IRB approval before implementing proposed changes.**
- **Change means any change, in content or form, to the protocol, consent form, or any supportive materials (such as the investigator's Brochure, questionnaires, surveys, advertisements, etc.). See Item 4 for more examples.**

1. Today's Date	2/20/2023

2. Principal Investigator (PI) Name: Dan O'Leary			
PI's Title:	Instructor / PhD Candidate	**Faculty PI (if PI is a student):**	Dr. Richard Sesek
Department:	Industrial & Systems Eng	**Department:**	Industrial & Systems Eng
Phone:	407-399-3189	**Phone:**	334-728-1438
AU-E-Mail:	djo0008@auburn.edu	**AU E-Mail:**	rfs0006@auburn.edu
Contact person who should receive copies of IRB correspondence (Optional):	Click or tap here to enter text.	**Department Head Name:**	Dr. Gregory Harris
Phone:	Click or tap here to enter text.	**Phone:**	334-844-1407
AU E-Mail:	Click or tap here to enter text.	**AU E-Mail:**	gah0015@auburn.edu

3. AU IRB Protocol Identification		
	3.a. Protocol Number: 22-538	
	3.b. Protocol Title: The Effects of Augmented Instruction on Manufacturing Assembly Training	
	3. c. Current Status of Protocol – For active studies, check ONE box at left; provide numbers and dates where applicable	
☐	**Study has not yet begun; no data has been entered or collected**	
☒	**In progress If YES, number of data/participants entered:** 2 trials run, others scheduled	**Current Approval Dates From:** 1/30/2023
☐	**Is this modification request being made in conjunction with/as a result of protocol renewal?** ☐ **YES** ☐ **NO**	
☐	**Adverse events since last review If YES, describe:** Click or tap here to enter text.	**To:** Click or tap to enter a date.
☐	**Data analysis only**	
☐	**Funding Agency and Grant Number:** Click or tap here to enter text.	**AU Funding Information:** Click or tap here to enter text.
☐	**List any other institutions and/ or AU approved studies associated with this project:** Click or tap here to enter text.	

The Auburn University Institutional Review Board has approved this Document for use from
02/21/2023 to -------------
Protocol # 22-538 EP 2301

4. Types of Change **Mark all that apply, and describe the changes in item 5**	
☒	**Change in Key Personnel** List the name(s) of personnel being added to or removed from the study and attach a copy of the CITI documentation for personnel being added to the study. Adding: Kralyn Thomas, Yen-Ting Guo, and Lucie Wang
☐	**Additional Sites or Change in Sites, including AU classrooms, etc.** Attach permission forms for new sites.
☒	**Change in methods for data storage/ protection or location of data/ consent documents** Added location for storage of consent forms for 2nd investigation.
☒	**Change in project purpose or project questions** Added 2nd investigation using similar methods to explore other augmentations.
☒	**Change in population or recruitment** Attach new or revised recruitment materials as needed; both highlighted version & clean copy for IRB approval stamp Expanded target number of participants in the same population. See revised protocol for details.
☒	**Change in study procedure(s)** Attach new or revised consent documents as needed; both highlighted revised copy & clean copy for IRB approval stamp Updated procedures and added separate consent for 2nd investigation. Consent for 1st investigation unchanged.
☒	**Change in data collection instruments/forms (surveys, data collection forms)** Attach new forms as needed; both highlighted version & clean copy for IRB approval stamp Reformatted to support both investigations. No material changes to data collected. Attached.
☐	**Other** (BUAs, DUAs, etc.) Indicate the type of change in the space below, and provide details in the Item 5.c. or 5.d. as applicable. Include a copy of all affected documents, with revisions highlighted as applicable. Click or tap here to enter text.

5. Description and Rationale
5.a. For each item marked in Question #4 describe the requested change(s) to your research protocol, and the rationale for each.
Expanded scope of the experiment to include a second, directly related investigation. Needed added team members to help run the protocol.
5.b. Briefly list (numbered or bulleted) **the activities that have occurred up to this point, particularly those that involved participants.**
Recruiting ongoing, two trial runs, additional scheduled. All those will continue to utilize the methods and forms previously approved. This modification creates no material change in the first investigation. Once the modification is approved we will revise our recruiting methods as described and begin running trials for the 2nd.
5.c. Does the requested change affect participants, such as procedures, risks, costs, benefits, etc.
Not for the first investigation. The 2nd will affect the participants recruited for it as described in the corresponding Informed Consent document.
5.d. Attach a copy of all "IRB stamped" documents currently used. (Information letters, consent forms, flyers, etc.)
Attached.
5.e. List all revised documents and attach two copies of the revised documents – one copy which highlights the revisions and one clean copy of the revised documents for the IRB approval stamp.
Attached.

6. Signatures
Principal Investigator: ______________________
Faculty Advisor PI, if applicable: ______________________

Version Date: 2/20/2023

AUBURN UNIVERSITY

INDUSTRIAL & SYSTEMS ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

The Effects of Augmented Instruction on Manufacturing Assembly Training

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. Following an initial phone screening the experiment will be scheduled at your convenience. After a brief orientation you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned one of the following forms of instructional media: paper work instructions (PWI), projected augmented reality (PAR), head-mounted AR (HMDAR), or head-mounted mixed reality (HMDMR). After a 10-minute training session you will be asked to repeat the assembly task from memory for 4 cars. Paper work instructions will remain available for reference as needed. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 45-60 minutes.

This study has some risk of physical and psychological discomfort, including fatigue, dizziness, eyestrain, and performance anxiety. Participants assigned the HMD instructional media are most susceptible to physical discomfort due to the nature of its display system, which can also increase the risk of tripping and impact. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience modern AR training methods, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

You are invited to participate in a research study to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The study is being conducted by Dan O'Leary, Ph.D. Candidate, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are not prone to motion sickness.
2. Have no prior experience with head-mounted or projected Augmented Reality (AR) systems.
3. Have no prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO Lab) as part of INSY 5800/6800 or otherwise.
4. Are age 18 or older.

What will be involved if you participate?
If you decide to participate in this research study, you will be asked to follow a mix of paper and augmented (projected or head-mounted AR) work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will be approximately 45-60 minutes. You will be required to wear a HoloLens2 head-mounted display (HMD) and video of your session will be recorded for later analysis. Another video camera will capture the work area from above. Camera placement is designed to prevent / limit the capture of personally identifiable imagery. Fully redacted versions of these videos, wherein any personally identifiable imagery is removed, will be kept indefinitely. Original recordings will be deleted within 1 year of the protocol's completion.

Are there any risks or discomforts?
The risks associated with participating in this study are identified below.

1. Physical discomfort and/or fatigue related to the weight of the HoloLens2 HMD.
2. Vestibular and/or visual discomfort for participants assigned to the HMD AR instructional methods, which may cause mild dizziness, eye strain, and related effects in some users.
3. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
4. Trip and impact risk due to slightly altered field of view and reduced peripheral vision while wearing the HoloLens2 HMD.
5. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.
6. Exposure to COVID-19 or other respiratory illnesses, such as the flu.

The discomforts identified are considered mild and unlikely. The HoloLens2 is well-balanced and uses a state-of-the-art optical see-through design that limits display-related discomforts. To minimize the risk of tripping and impact, participants are largely stationary in a well-lit area that is free of hazards. The HoloLens2 features a wireless design, which eliminates cables as a source of tripping hazard. Finally, all activities will be supervised, and participants will be continuously monitored for relevant symptoms.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

To mitigate the risk of exposure to COVID-19 and other respiratory illnesses, the research team will follow University policies outlined on the Human Research COVID-19 Precautions page. All work surfaces and equipment will be wiped down before and after each participant, and all necessary supplies (e.g. masks, hand sanitizer) will be made available. The research staff will follow the University's guidance on self-screening. Finally, conditions will be monitored, and precautions adjusted as necessary throughout the data collection process.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to interact with projection and/or head-mounted AR hardware and training methods. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
No compensation is offered for your participation.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

Participant's Initials: ______

The Auburn University Institutional Review Board has approved this Document for use from 01/28/2023 to ------------- Protocol # 22-538 EP 2301

If you have questions about this study, please ask now or contact Dan O'Leary at djo0008@auburn.edu, 407-399-3189, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Participant's signature	Date	Investigator obtaining consent	Date
Printed Name		Printed Name	

The Auburn University Institutional Review Board has approved this Document for use from
01/28/2023 to -------------
Protocol # 22-538 EP 2301

AUBURN
UNIVERSITY

INDUSTRIAL & SYSTEMS
ENGINEERING

(NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.)

INFORMED CONSENT
for a Research Study entitled

Studying Manufacturing with LEGO® Research

Concise Summary

You are being asked to take part in a research study. This research study is voluntary, meaning you do not have to take part in it. The procedures, risks, and benefits are fully described further in the consent form. The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity, learning rates, and skills transfer for industrial assembly tasks. Following an initial phone screening, the experiment will be scheduled at your convenience. After a brief orientation, you will be asked to learn a simulated manufacturing assembly task – building model "cars" with LEGO® bricks. For this phase you will be randomly assigned an order to complete the following treatments: paper work instructions (PWI), assembly with a pre-completed model for quality checks, an inspection camera for quality checks, and both the pre-completed model and inspection camera. You will be asked to complete four car assemblies for training using the paper work instructions prior to using the prescribed tasks. After the training, each treatment will last 10 minutes for a total of four treatments. Paper work instructions will remain available for reference as needed. Between each task you will be asked to complete two brief surveys about your experience. Finally, you will be asked to complete a survey with questions about the experience and related personal traits. The entire process will take 70-90 minutes.

This study has some risk of physical and psychological discomfort, including fatigue and performance anxiety. Finally, all of your personally identifiable data is carefully secured to protect against the risk of a breach of confidentiality. Your safety and privacy is our utmost priority, and steps have been taken to mitigate all known risks.

Beyond the opportunity to experience training in the Tiger Motors Lab, there are no direct benefits to you for participating in this study. The researchers will benefit from a greater understanding of this emerging field that could potentially benefit the community. The alternative is to not participate in this study.

You are invited to participate in a research study to measure the effect of Lean Tools and Industry 4.0 Technologies on productivity. The study is being conducted by Victoria Ballard and Md Monir Hossain, Ph. D. students, under the direction of Dr. Richard Sesek, Tim Cook Associate Professor in the Auburn University Department of Industrial and Systems Engineering. You were selected as a possible participant because you meet all the following qualifications:

1. Are age 18 or older.

What will be involved if you participate?
If you decide to participate in this research study, you will be asked to follow work instructions to build LEGO car models in a realistic manufacturing setting. Your total time commitment will -be approximately 70-90 minutes. Video of your session will be recorded for later analysis. Camera placement is designed to prevent / limit the capture of personally identifiable imagery.

Are there any risks or discomforts?
The risks associated with participating in this study are identified below.

1. Psychological discomfort may be experienced by those prone to anxiety when encountering time and performance-based measures.
2. Participant confidentiality may be breached if identifying data is compromised or participants are observed entering, leaving, or taking part in the experiment.

Confidentiality of the study data is of utmost importance. All research personnel are trained in research ethics and are aware of procedures to protect the confidentiality of participants and associated data. Paper files with personally identifiable information will be secured in an office that only the PI and Faculty Advisor have access to. Electronic data, including video recordings, will be maintained on a password-protected computer accessible only to the research team.

Are there any benefits to yourself or others?
There are no direct benefits from participating in this study. However, it is a unique opportunity for eligible participants to participate in research in the Tiger Motors Lab. This may lead them to a greater appreciation for the benefits and opportunities these technologies offer.

Will you receive compensation for participating?
No compensation is offered for your participation.

Are there any costs?
There is no cost for you to participate in this study. Auburn University has not provided for any payment if you are harmed as a result of participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is completely voluntary. If you choose to withdraw, your data can be withdrawn as long as it is identifiable. Your decision about whether or not to participate or to stop participating will not jeopardize your future relations with Auburn University, the Department of Industrial and Systems Engineering or any member of the research team.

Your privacy will be protected. Any information obtained in connection with this study will remain confidential. Information obtained through your participation may be used in a variety of capacities, including fulfillment of educational requirements, publication in professional journals, and/or presentation at professional meetings. In any case, your identity will not be revealed, and your information will remain private.

If you have questions about this study, please ask now or contact Victoria Ballard at victoria.ballard@auburn.edu,360-632-1359, or Dr. Richard Sesek at rfs0006@auburn.edu, 334-728-1438. A copy of this document will be given to you to keep.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

____________________	____________________
Participant's signature Date	Investigator obtaining consent Date
____________________	____________________
Printed Name	Printed Name

The Auburn University Institutional Review Board has approved this Document for use from
02/21/2023 to -------------
Protocol # 22-538 EP 2301

Appendix B - Recruiting Materials

In-Class Recruiting Script

Hello, Class.

Industrial Engineering graduate students pursuing their PhDs are recruiting participants for a research study. They are investigating the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. These investigations hope to better understand the relationships between those methods, learning effectiveness, and operator performance. A flyer with details of the study will be emailed to each of you. If you are interested, please follow up as described therein.

Email Script

Dear Student,

Please review the attached flyer, which provides details of the study recently described in class name. You are invited to participate in a research study on the effectiveness of Mixed Reality, Lean, and Industry 4.0 methods for operator training and support in manufacturing. The research team is conducting this study as Ph.D. Candidates under the supervision of Dr. Richard Sesek, Tim Cook Associate Professor in the Department of Industrial and Systems Engineering at Auburn University.

If you would like to participate, simply respond to this email or via text / phone to 407-399-3189. Questions or concerns can be directed to me through the same channels, or you may contact my advisor Dr. Sesek (sesek@auburn.edu).

Thank you for your consideration,

Confirmation Email

Dear <student name>,

Thank you for your interest in our study, and for taking the time to discuss it with me. I'm happy to confirm that your trial is scheduled as follows:

Date and Time:<date and time>

Location: Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab), in the basement of the Shelby Center for Engineering Technology, room 0317, located at 345 W Magnolia Ave, Auburn, AL 36849

Please arrive on time. We anticipate that it will take 45-90 minutes to complete the session.

If you need to reschedule or have further questions, feel free to respond to this email or call / text me at 407-399-3189.

Thank you for your participation,

Flyer

On the following pages are flyers for both investigations, formatted as posters and slides.

Augmented Reality Research Study

Training methods for tomorrow's workforce, today!

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

This study is open to anyone 18 and older, that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab)as part of INSY 5/6800 or otherwise.

Microsoft HoloLens

LightGuide

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

If you are interested in participating or have questions, please contact Dan O'Leary (djo0008@auburn.edu, 407-399-3189), or scan the QR code to generate an email.

Scan to Email!

The Auburn University Institutional Review Board has approved this Document for use from
01/28/2023 to -------------
Protocol # 22-538 EP 2301

Studying Manufacturing with LEGO(R) Research

Participate in research in Auburn's Tiger Motors Lab!

The Effects of Lean Tools and Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials. Participants will assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours. No compensation for the study, but you will get to build LEGO cars in the world-famous Auburn Tiger Motors Lean Education Center (AKA LEGO Lab!).

This study is open to anyone 18 and older.

TIGER MOTORS
LEAN EDUCATION CENTER

AUBURN
UNIVERSITY

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

If you are interested in participating or have questions, please contact Dan O'Leary (djo0008@auburn.edu, 407-399-3189), or scan the QR code to generate an email.

Scan to Email!

The Auburn University Institutional Review Board has approved this Document for use from
02/21/2023 to -------------
Protocol # 22-538 EP 2301

Augmented Reality Research Study

The Effects of Augmented Instruction on Manufacturing Assembly Training

Interested in Augmented and Mixed Reality?
Want to experience the latest in Projected and Head-Mounted AR?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of augmented instruction on learning rates and skills transfer for industrial assembly tasks. The effect of projected (LightGuide) and head-mounted (HoloLens2) augmented reality methods will be compared with paper-based materials for instruction and support.

This study is open to anyone 18 and older, that isn't prone to motion sickness, has no prior experience with head-mounted or projected AR systems, and hasn't worked in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab)as part of INSY 5/6800 or otherwise.

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

If you are interested in participating or have questions, please contact Dan O'Leary (djo0008@auburn.edu, 407-399-3189), or scan the included QR code to generate an email.

Scan to Email!

Microsoft HoloLens

LightGuide

The Auburn University Institutional Review Board has approved this Document for use from 02/21/2023 to -------------
Protocol # 22-538 EP 2301

Studying Manufacturing with LEGO®

The Effects of Lean Tools & Industry 4.0 Technology on Manufacturing Assembly Performance

Want to help the future of manufacturing research?
Want to use the latest vision inspection equipment and play with LEGO?
You may be eligible to participate in an important study!

The purpose of this study is to measure the effect of Lean Tools and Industry 4.0 Technology on industrial assembly tasks. The effect of a model check piece, camera inspection technology, and a combination of the two will be compared with paper-based materials.

Participants will assemble one station of LEGO vehicles in four scenarios. The time for completion is approximately 1.5 hours. No compensation for the study, but you will get to build LEGO cars in the world-famous Auburn Tiger Motors Lean Education Center (AKA LEGO Lab)!

This study is open to anyone 18 and older.

Conducted by graduate students in the Department of Industrial & Systems Engineering at Auburn University.

If you are interested in participating or have questions, please contact Dan O'Leary (djo0008@auburn.edu, 407-399-3189), or scan the included QR code to generate an email.

Scan to Email!

TIGER MOTORS
LEAN EDUCATION CENTER

The Auburn University Institutional Review Board has approved this Document for use from 02/21/2023 to -------------
Protocol # 22-538 EP 2301

Subject Recruitment Data Sheet

Eligibility Checklist:

- ☐ 18 or older
- ☐ Not prone to motion sickness
- ☐ No prior experience with projected or head-mounted augmented reality systems
- ☐ No prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab) as part of INSY 5/6800 or otherwise

If eligible, record name, contact info (phone, email), and subject number in code sheet.

Participant Number: ____________________

Gender: ______________________________

Age: ________________________________

Eligible: ____ I1 ____ I2 ____ Both

Scheduled Trial(s): ____________________

Notes:

Eligibility Checklist:

- ☐ 18 or older
- ☐ Not prone to motion sickness
- ☐ No prior experience with projected or head-mounted augmented reality systems
- ☐ No prior experience building cars in the Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab) as part of INSY 5/6800 or otherwise

If eligible, record name, contact info (phone, email), and participant number in code sheet.

Participant Number: ____________________

Gender: ______________________________

Age: ________________________________

Eligible: ____ I1 ____ I2 ____ Both

Scheduled Trial(s): ____________________

Notes:

Participant Intake Sheet, p1 / 2

Participant #: __________ **Date: __________**

1. Gender:
   - ☐ Female
   - ☐ Male
   - ☐ Other
2. Age: _____
3. Race (select those with which you identify):
   - ☐ American Indian or Alaska Native
   - ☐ Asian
   - ☐ Black or African-American
   - ☐ Native Hawaiian or Other Pacific Islander
   - ☐ White
   - ☐ More than one race
   - ☐ Unknown or not reported
4. Ethnicity (select ONLY one with which you most closely identify):
   - ☐ Hispanic or Latino
   - ☐ Not Hispanic or Latino
   - ☐ Unknown or not reported
5. Country of Origin: _______________________
6. What language do you mainly speak at home?
   - ☐ English
   - ☐ Other
7. What is the highest level of school you have completed or the highest degree you have received?
   - ☐ Less than high school degree
   - ☐ High school degree or equivalent (e.g., GED)
   - ☐ Some college but no degree
   - ☐ Associate degree
   - ☐ Bachelor degree
   - ☐ Graduate degree: ____ Master or ____ PhD

Participant Intake Sheet, p2 / 2

Participant #: __________ **Date: __________**

8. If you are currently pursuing a degree, please complete the following:

   College (e.g. Education or Business): ________________________

   Program (e.g. MS Adult Ed or BS Accounting) : ________________________

9. Which of the following statements best describes your experience building LEGO models?

   ☐ I have little to no experience building LEGO models.

   ☐ I have some experience building LEGO models.

   ☐ I have lots of experience building LEGO models.

   ☐ I consider myself an expert in building LEGO models.

10. Please indicate your level of manufacturing experience

   ☐ I have no experience in manufacturing.

   ☐ I have taken one or more classes in manufacturing.

   ☐ I have held a part-time or temporary position in manufacturing.

   ☐ I have 1 or more years of experience working in manufacturing.

Code Sheet

Part. #	Date	Name	Email	Phone	Assigned	Notes
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	
					1 2	

CONFIDENTIAL

Data Collection Sheet #1

Participant #: __________ **Date:** __________

First Investigation
Circle Training Treatment:
PWI / PAR / HMDAR / HMDMR

Second Investigation	
Treatment Number	**Treatment**
1 / 2 / 3 / 4	Control / Lean / I-4.0 / Lean+I-4.0

Car #	TCT	Errors Made		Uncorrected Error Types			PWI Ref Count	Trial Notes
		Corrected	Uncorrected	Sel	Pos	Rot		
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Observer Initials: ________

Data Collection Sheet #2

Participant #: __________ **Date:** __________

First Investigation
Circle Training Treatment:
PWI / PAR / HMDAR / HMDMR

Second Investigation
Training with
Paper Work Instructions

Car #	TCT	Errors Made		Uncorrected Error Types			PWI Ref Count	Trial Notes
		Corrected	Uncorrected	Sel	Pos	Rot		
1								
2								
3								
4								

General Notes:

Observer Initials: ________

Task Loading Index, p1 / 2

Participant #: __________   **Invest / Treat: __________**   **Date: __________**

Sources of Workload

Consider the following definitions:

Title	Range	Description
Mental Demand	Low / High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low / High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low / High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good / Poor	How successful do you think you were in accomplishing the goals of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low / High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Low / High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

For each of the following pairs, circle the word that represents the more important contributor to workload for the specific task(s) you performed in this experiment.

Effort or **Performance**	**Temporal Demand** or **Frustration**	**Physical Demand** or **Performance**	**Temporal Demand** or **Mental Demand**	**Mental Demand** or **Physical Demand**
Temporal Demand or **Effort**	**Physical Demand** or **Frustration**	**Frustration** or **Effort**	**Performance** or **Mental Demand**	**Effort** or **Physical Demand**
Performance or **Frustration**	**Physical Demand** or **Temporal Demand**	**Performance** or **Temporal Demand**	**Mental Demand** or **Effort**	**Frustration** or **Mental Demand**

Participant #: __________ **Invest / Treat: __________** **Date: __________**

Workload Rating Sheet

For each of the following 6 questions, consider the assembly task you just completed. Record your immediate response to each item by circling the number that you feel best represents your experience.

1. How mentally demanding was the task?

1	2	3	4	5	6	7
Very Low						Very High

2. How physically demanding was the task?

1	2	3	4	5	6	7
Very Low						Very High

3. How hurried or rushed was the pace of the task?

1	2	3	4	5	6	7
Very Low						Very High

4. How successful were you in accomplishing what you were asked to do?

1	2	3	4	5	6	7
Perfect						Failure

5. How hard did you have to work to accomplish your level of performance?

1	2	3	4	5	6	7
Very Low						Very High

6. How insecure, discouraged, irritated, stressed, and annoyed were you?

1	2	3	4	5	6	7
Very Low						Very High

System Usability Scale

Participant #: __________ **Invest / Treat:** __________ **Date:** __________

For each of the following 10 questions, consider the assembly task you just completed. Record your immediate response to each item by circling the number that you feel best represents your experience.

		Strongly Agree				Strongly Disagree
1	I think that I would like to use this system frequently.	1	2	3	4	5
2	I found the system unnecessarily complex.	1	2	3	4	5
3	I thought the system was easy to use.	1	2	3	4	5
4	I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5	I found the various functions in this system were well integrated.	1	2	3	4	5
6	I thought there was too much inconsistency in this system.	1	2	3	4	5
7	I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8	I found the system very cumbersome to use.	1	2	3	4	5
9	I felt very confident using the system.	1	2	3	4	5
10	I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Behavioral Control Survey

Participant #: __________ **Date: __________**

Please answer the questions below, rating yourself on each of the criteria shown using the scale on the right side of the page. As you answer each question, place an X in the box that best describes how you have felt and conducted yourself over the past 6 months.	Never	Rarely	Sometimes	Often	Very Often
1. How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?					
2. How often do you have difficulty getting things in order when you have to do a task that requires organization?					
3. How often do you have problems remembering appointments or obligations?					
4. When you have a task that requires a lot of thought, how often do you avoid or delay getting started?					
5. How often do you fidget or squirm with your hands or feet when you have to sit down for a long time?					
6. How often do you feel overly active and compelled to do things, like you were driven by a motor?					
7. How often do you make careless mistakes when you have to work on a boring or difficult project?					
8. How often do you have difficulty keeping your attention when you are doing boring or repetitive work?					
9. How often do you have difficulty concentrating on what people say to you, even when they are speaking to you directly?					
10. How often do you misplace or have difficulty finding things at home or at work?					
11. How often are you distracted by activity or noise around you?					
12. How often do you leave your seat in meetings or other situations in which you are expected to remain seated?					
13. How often do you feel restless or fidgety?					
14. How often do you have difficulty unwinding and relaxing when you have time to yourself?					
15. How often do you find yourself talking too much when you are in social situations?					
16. When you're in a conversation, how often do you find yourself finishing the sentences of the people you are talking to, before they can finish them themselves?					
17. How often do you have difficulty waiting your turn in situations when turn taking is required?					
18. How often do you interrupt others when they are busy?					

General Feedback

Participant #: __________ **Date: __________**

Please share with us any other feedback you have regarding this experiment.

For Research Associate Only

Follow-up? _____ Injury? _____ Discomfort? _____ Initial: _____

Appendix D - Emergency Plan, Contact List, and COVID Resources

Emergency Action Plan

In Case of Emergency ***DIAL 911***

For non-emergency assistance:

Service	On-Campus	Off-Campus
Ambulance (EMS)	9-749-8504	334-749-8504
City of Auburn Police	9-501-3100	334-501-3100
Auburn Medical Pavilion	9-364-3000	334-364-3000
East Alabama Medical Center, Opelika	9-749-3411	334-749-3411

Research Team Contact List:

Contact	Phone	Email
Dan O'Leary, Principal Investigator	407-399-3189 (cell)	djo0008@auburn.edu
Dr. Richard Sesek, Faculty Advisor	334-728-1438 (cell)	rfs0006@auburn.edu
Victoria Ballard, Graduate Student	360-632-1359 (cell)	vzb0024@auburn.edu
Dr. Gregory Harris, Faculty Advisor	334-844-1407 (office)	gah0015@auburn.edu
Dr. John Evans, Faculty Advisor	334-844-1418 (office)	evansjl@auburn.edu
Tom Devall, Tiger Motors Director	334-740-3905 (office)	tld0017@auburn.edu
Industrial & Systems Engineering Department	334-844-4340 (main office)	insy@eng.auburn.edu

Lab Location and Access:

Tiger Motors Lean Education Center (Lean Lab, aka LEGO® Lab), Basement, Shelby Center, Auburn University, room 0317. Street address: 345 W Magnolia Ave, Auburn, AL 36849.

Elevator access: exit the lab and turn left

Stairwell access: exit the lab, turn left, proceed around the elevator in either direction. Stairwell entrance is on the inside wall behind the elevator.

Emergency exit: exit the lab and turn right. Continue to exit at ground level.

Emergency Equipment:

First aid kit, eye wash and shower station are present, as are fire extinguisher and alarm pull.

COVID-19 Resources

[CDC COVID-19 Data Tracker for Lee County, Alabama]

University Policies for Research Exposure and Related Resources:

- Human Research COVID-19 Precautions
- COVID-19 Guidance on Self Screening
- AU Facilities COVID Building Readiness Status Page

Auburn University Screening Protocol (source):

All research participants should be screened remotely (by phone or Zoom) for fever, cough, and flu-like symptoms the day before, with a repeat screening at the time of an in-person visit. Appropriate screening questions might include the following, which could be modified to fit your participant population and the location of in-person interactions:

1. Do you have a fever or Respiratory Symptoms? Symptoms include fever, acute respiratory infection, persistent cough, sore throat, fatigue and shortness of breath, or sudden loss of taste or smell with or without a fever.
2. Are you waiting on COVID-19 test results?
3. Have you been asked to self-isolate by your doctor?
4. In the past three weeks, have you visited another state, country, or facility with a substantial or high community COVID-19 level (see CDC COVID-19 Community Levels)?
5. Health/Vaccination Status - Do you have underlying medical conditions, or are you unvaccinated?

Precautions Matrix:

COVID-19 PRECAUTIONS MATRIX	CATEGORY A HIGH-RISK PROCEDURES*	CATEGORY B HIGH-RISK PARTICIPANTS**	CATEGORY C NO HIGH-RISK PROCEDURES OR PARTICIPANTS
HIGH COVID-19 COMMUNITY LEVEL	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS PPE: RESEARCH PERSONNEL WEAR N-95 OR KN95; EYE PROTECTION; GLOVES FOR DIRECT CONTACT; PARTICIPANTS WEAR FACE COVERINGS	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS PPE: RESEARCH PERSONNEL WEAR N-95 OR KN95; EYE PROTECTION; GLOVES FOR DIRECT CONTACT; PARTICIPANTS WEAR FACE COVERINGS	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES
MEDIUM COVID-19 COMMUNITY LEVEL	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES	SCREENING PROTOCOLS FOR PARTICPANTS AND INVESTIGATORS FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES
LOW COVID-19 COMMUNITY LEVEL	FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES	FOLLOW AU COVID-19 GUIDELINES

*HIGH-RISK PROCEDURES ARE DEFINED AS ANY PROCEDURES THAT INCUR A SIGNIFICANT OR INCREASED RISK OF EXPOSURE, SUCH AS THROUGH FREQUENT OR SUSTAINED CLOSE CONTACT BETWEEN INVESTIGATORS AND PARTICIPANTS; SPECIMEN COLLECTION FROM PARTICIPANTS; OR ACTIVITIES INVOLVING INCREASED RESPIRATORY OUTPUT SUCH AS EXERCISE STUDIES.

**HIGH-RISK PARTICIPANTS INCLUDE PEOPLE AT HIGHER RISK OF SEVERE ILLNESS FROM SARS COV-2 INFECTION, INCLUDING PEOPLE WHO ARE UNVACCINATED, OLDER ADULTS, OR PEOPLE WITH CERTAIN MEDICAL CONDITIONS.