Fakultät für Mathematik und Informatik
Institut für Informatik
Mensch-Computer-Interaktion

Masterarbeit

zur Erlangung des Grades
Master of Science (M.Sc.)
im Studiengang Mensch-Computer Interaktion
an der Universität Würzburg

Evaluating the Effect of AR-Assistance on Task Performance in Manual Liquid Handling Scenarios: A Comparative Study

vorgelegt von
Sascha Lange
Matrikelnummer: 2609220

am December 20, 2023

Betreuer/Prüfer:
Dr. Martin Fischbach, Informatik IX, Universität Würzburg
Dr. Florian Heinrich, Virtual and Augmented Reality Group,
Otto-von-Guericke-Universität Magdeburg

To those who persist and stay the course

—

your efforts are not in vain.

Zusammenfassung

Manuelles pipettieren ist ein gängiger, aber fehleranfälliger Prozess in biochemischen Laboren, bei dem LaborantInnen wiederholt kleine Flüssigkeitsmengen in verschiedene Behälter, z.B. Mikrotiterplatten, übertragen. In dieser Studie wird ein Augmented Reality Pipetting Assistant System (ARPAS) als potenzielle Lösung zur Reduzierung von Fehlern und Zeitaufwand bei Pipettieraufgaben vorgestellt. Die Wirksamkeit von ARPAS wurde mit einem traditionellen Papierprotokoll und einer Tablet-Anwendung in einer Studie mit 48 Laborfachkräften aus der Life Science Industrie verglichen. Diese drei Methoden - *Paper*, *Tablet* und *AR* - wurden im Hinblick auf die Ausführungszeit der Aufgabe, die Anzahl der Fehler und die subjektive Arbeitsbelastung bewertet. In der Studie wurde eine innovative Messung der Fluoreszenzintensität mit Uranin verwendet, um die Pipettiergenauigkeit genau zu bewerten. Die Studie zeigte, dass die *Tablet*-Methode in Bezug auf die Geschwindigkeit am effizientesten war. Darüber hinaus reduzierten sowohl die *Tablet*- als auch die *AR*-Methode die Fehlerquote erheblich, wobei die *Tablet*-Methode im Vergleich zur *Papier*-Methode zusätzlich auch die subjektive Arbeitsbelastung senkte. Obwohl die *Tablet*-Methode am effektivsten war, unterstreicht die vergleichbare Genauigkeit der *AR*-Methode, trotz der begrenzten Exposition der Teilnehmer gegenüber der Augmented- und Virtual-Reality-Technologie, das Potenzial von Augmented-Reality-basierten Assistenzsystemen in Laborsituationen.

Abstract

Manual liquid handling, or pipetting, is a common yet error-prone process in biochemical laboratories, in which technicians repetitively transfer small volumes of liquid into different containers, e.g. microplates. This study introduces an *Augmented Reality Pipetting Assistant System* (ARPAS) as a potential solution to reduce errors and time spent in pipetting tasks. The effectiveness of ARPAS was compared with a traditional paper protocol and a tablet application in a between-subject study involving 48 life-science laboratory professionals. These three methods - *Paper*, *Tablet*, and *AR* - were evaluated in terms of task execution time, error count, and subjective workload. The study innovatively used fluorescent intensity measurement with uranine to accurately assess pipetting accuracy. The study showed that the *Tablet* method was the most efficient in terms of speed. Additionally, both the *Tablet* and *AR* methods significantly reduced errors, with the *Tablet* method also lowering subjective workload compared to the *Paper* method. Although the *Tablet* method was the most effective, the comparable accuracy of the *AR* method, despite limited participant exposure to augmented and virtual reality technology, highlights the potential of Augmented Reality-based assistance systems in laboratory settings.

Contents

List of Figures

List of Tables

Acronyms

ANOVA Analysis of Variance. 59, 62–64, 67–69, 71

ANSI American National Standards Institute. 38

API Application Programming Interface. 20, 82

AR Augmented Reality. 2, 3, 5, 8–12, 14, 16, 17, 20, 27, 33, 35, 51, 79–83, 85, 87, 88

eLN Electronic Laboratory Notebook. 2, 3, 6

FoV Field of View. 37, 76, 77, 82, 84

HCI Human-Computer Interaction. 4, 6, 36, 79, 80, 83

HMD Head-Mounted Display. 3, 8, 9, 11, 12, 20, 36, 52, 54, 75–77, 80–82, 84, 85, 87, 88

OST Optical See-Through. 3, 8, 9, 12, 37, 84

PBS Phosphate-Buffered Saline. 38–40

pLN Paper Laboratory Notebook. 2

Q-Q plots Quantile-Quantile plots. 62, 67, 70, 71

RFU Relative Fluorescence Unit. 47, 49

SAR Spatial Augmented Reality. 8, 9

SLAS Society for Laboratory Automation and Screening. 38

SUS System Usability Scale. 29, 51, 55, 56, 58, 61, 69–72

TLX NASA Task Load Index. 11, 16, 51, 55, 57, 65–69, 80

TUI Tangible User Interface. 7, 85, 87

UI User Interface. 8, 11–13, 22, 23, 79–81

VR Virtual Reality. 8, 33, 35, 51, 80, 81

VST Video See-Through. 8, 82, 84, 88

1 Introduction

Research laboratories in the Life Science industry discover new medicine through accurate, reliable, and reproducible science. This process is reliant on liquid handling, the transferal of exactly defined amounts of liquids, e.g. blood samples, reagents or solvents by means of a pipette. Manual liquid handling, despite growing automation efforts in the industry (Kong et al., 2012), is still a key and critical process in the laboratory. Lab technicians need high concentration, good preparation, and years of practice to avoid pipetting mistakes (Swangnetr et al., 2018).

Despite all preparations, errors in the liquid handling process do occur. Unfortunately, those mistakes are often not discovered until the end of the analytical pipeline, rendering the results unreliable at best and useless at worst. Insufficient data can lead to an expensive and time-consuming rerun of the experiment, keeping the researchers from doing more progressive work (Tegally et al., 2020).

While lab automation with robots is often considered a very accurate and reliable solution to this problem, the upfront investment, high maintenance cost and inflexible programming cut out small sized research laboratories with frequently changing experiments (Holland & Davies, 2020).

Non-automated laboratory testing, a complex multistep process (Da Rin, 2009), relies on manual liquid handling in both pre- and analytical phases. Continuous pipetting, due to high exerted workload (Swangnetr et al., 2018), can lead to lab technician fatigue, impacting performance and increasing error likelihood (Yung et al., 2017). Fatigue, often being caused by monotony (Kim et al., 2009), can result in short-term cognitive and physical degradation (Techera et al., 2016), adversely affecting not just pipetting but the entire analytical process.

To counteract this, humans come up with a variety of cognitive artifacts as support in their work: post-it's, paper notes or specific placement of objects etc. (Norman, 1991). In the laboratory context the most common form of cognitive artifact is

Figure 1.1. A Phd student keeping notes in a lab notebook. Source: https://commons.wikimedia.org/wiki/File:Lab_Notebook.jpg (Last accessed: 27.02.2023)

the Paper Laboratory Notebook (pLN) (Figure 1.1). It's purpose is to keep a permanent and detailed record of the materials, procedures, results obtained and any observations made by a scientist during an experiment (Bird et al., 2013).

The laboratory notebook is an important tool for the scientist in order to execute experiments accurately. For a long time only existent in paper, nowadays the industry shifted towards accepting the Electronic Laboratory Notebook (eLN) to reap its benefits (Gerlach et al., 2020). Although eLN's have many positive features, one major drawback they share with their paper counterparts is the spatial disconnection between where the information is stored and where it is needed. In the case of pipetting, scientists have to refocus from their sample they are working on, towards the notebook in order to make sure they are following the documented protocol. Reducing this information gap, a study by Tang et al., 2003 revealed the effectiveness of Augmented Reality (AR) in an assembly task, by overlaying 3D instructions on the actual work piece. This approach led to fewer errors and a lower mental load than traditional printed manuals or monitor-based displays.

In recent years, researchers have explored possibilities to extend the functionality of eLN's with wearable devices such as Google Glass (Hu et al., 2015) or the Apple Watch (Guerrero et al., 2016). The goal was to provide important information to the lab technician with as little friction and spatial separation as possible. Besides getting protocol information in the field of view or on the lab technicians wrist, a combination of both wearable devices was used to extend input modalities and control experiment documentation (Scholl, Wille, et al., 2015).

A pipetting-specific example of bridging the gap between information storage and application using AR is the work of Hile et al., 2004. A projector-based pipetting assistance with microplate and pipette tracking capabilities was developed, which was an early exploration of the idea of AR pipetting assistance. Motivated by this work, previous personal work looked at ways how AR on an Optical See-Through (OST) Head-Mounted Display (HMD) could assist users to perform manual liquid handling task faster, with less errors and less subjective workload compared to the use of paper-based experiment protocols (Lange & Niebling, 2022).

Building upon this previous system, this thesis explored the research question: can an AR-based assistance system be effectively used in manual liquid handling scenarios by expert personal? A comprehensive user study was conducted, comparing the proposed AR solution with a state-of-the-art, screen-based pipetting assistance system and a traditional paper-based protocol. Through this methodology this thesis contributed to the growing body of evidence supporting AR as a valuable tool in modern laboratory practice.

This thesis is structured as follows:

Chapter 1 provides the introduction, detailing the motivation, objectives and overarching research question.

Chapter 2 reviews related work in laboratory digitalization, AR task assistance, and the Pipette Show web application. It revisits the previous personal work on AR pipetting assistance (ARPAS v1), discussing its influence on the improved system developed in this thesis.

Chapter 3 details the development of ARPAS v2, focusing on its advancements over ARPAS v1. It outlines key improvements made in response to the evaluation feedback from ARPAS v1, aligning with this thesis' overall research objective and methodology.

Chapter 4 details the research methods, including the formation of hypotheses, study design, participant description, materials and apparatus used, and the procedure adopted. This chapter also presents all relevant measurements, such as task execution time, error count, subjective workload, usability scores, and user experience metrics. The evaluation section illustrates both statistical and qualitative data analysis methods used.

Chapter 4 outlines the research methodology, including hypothesis development, study design, sample description, and the materials and procedures used. The key measurements are explained including task execution time, error count, subjective workload, usability scores, and user experience metrics. The chapter concludes with an overview of both statistical and qualitative data analysis methods applied in the evaluation.

Chapter 5 presents the study results, including both descriptive and inferential analysis. It covers task performance metrics, questionnaire and qualitative interview responses, providing a comprehensive overview of the results.

Chapter 6 engages in a discussion, summarising and interpreting the study's results. It integrates these insights with previous research, acknowledges the study's limitations, and suggests possible technical and methodological directions for future work.

Chapter 7 provides a comprehensive summary of the research, its broader implications, and its contributions to the field of Human-Computer Interaction (HCI).

2 Related Work

This chapter offers an in-depth review of the research related to this work. It begins by introducing the concept of computer assistance systems in laboratory environments, before delving into the use of AR for task assistance across various industries. The chapter also discusses the application of a recently published work, Pipette Show, as it is applied in this study's experimental setup. Finally, the chapter focuses on previous personal work on AR pipetting assistance, the foundation for the software system developed in this research. The aim is to clearly delineate the scope and functionality of prior work, establishing a distinct boundary between what has been previously accomplished and what is newly implemented in this study.

2.1 Challenges of Manual Pipetting in Laboratory Practice

In the Total Testing Process (TTP) of laboratories, manual pipetting is a critical part of the pre-analytical phase (Da Rin, 2009), where precision and accuracy are paramount for reliable results. Studies show that pipetting involves high subjective workload (Swangnetr et al., 2018; Swangnetr et al., 2012) and is impacted by cognitive and physical fatigue, affecting performance. This could contribute to the significant error rates, ranging from 46 to 68%, observed in the pre-analytical phase of the TTP (Plebani, 2006). Despite the advent of automation, manual pipetting is indispensable, particularly in smaller-scale experiments or when automation is not accessible, as in many developing countries or acedemic reseach laboratories (Falk et al., 2022).

Variability in manual pipetting, a critical factor in experimental accuracy, is influenced by operator technique, environmental conditions, and equipment quality. This variability has significant implications, for example in clinical laboratories, where pipetting precision directly affects diagnostic accuracy and clinical decisions (Guan et al., 2023). Although often under-discussed, studies reveal that both intra- and inter-individual imprecision in manual pipetting is inversely related to the volume dispensed (Lippi et al., 2017). This underscores the necessity for rigorous training and regular skill assessment to ensure high standards in laboratory practices.

Proper pipetting technique, including practices like tip pre-wetting and careful aspiration, is crucial for accurate results, particularly with different liquid types (EppendorfAG, 2015). Comprehensive training and adherence to these techniques are essential for maintaining high lab standards. Training programs, using simple instructional materials and practical exercises, have shown to enhance skills and significantly improve pipetting accuracy (Yamamoto et al., 2014). Innovative approaches, like digital pipetting badges, have also been explored to improve hands-on laboratory skills. These methods not only enhance technical proficiency but also boost confidence and knowledge in pipetting among practitioners (Towns et al., 2015).

2.2 Digitalization Efforts in the Laboratory Environment

While the intricacies of manual pipetting and the associated challenges underscore the need for precision and skill in laboratory practices, they also highlight an opportunity for technological advancements. One such advancement is the integration of eLN's. Today's eLN's have extended beyond mere digital replacements of paper notebooks. They not only facilitate the capture and search for experimental procedures but also incorporate multiple heterogeneous streams of user activity and web information. This integration offers a balanced approach to recording activities, providing both discipline and flexibility (Dirnagl & Przesdzing, 2016; Kanza et al., 2017). While there are many specialised eLN's available (Rubacha et al., 2011), also general office software such as Microsoft OneNote can be implemented as an eLN. Praised for its flexibility in data acquisition, data presentation and the ability to run on desktop an mobile devices, it is a great example of an eLN serving as the central hub for laboratory information (Guerrero et al., 2019; Schneikart & Mayrhofer, 2022).

Parallel to the advances in eLN's and their progress of adaptation, HCI research focused on an even more integrated and interactive approach to laboratory work: the advent of smart tabletop surfaces. With early advances made two decades ago by Labscape (Arnstein et al., 2002), a first exploration of ubiquitous computing, providing researchers information wherever needed. Evolutions of this idea specifically targeting the biology laboratory and bio molecular domain were first the G-nome Surfer (later, G-nome Surfer 2.0), a tabletop interface for collaborative data exploration of genomic data using multi-touch and a Tangible User Interface (TUI) (Shaer et al., 2010; Shaer et al., 2011). The work of Echtler et al., 2010, BioTISCH, set out to close the physical gap between the storage and access of relevant information in the wet lab. BioTISCH allows users to display data about reagents and assists them through ongoing procedures by emphasising the steps in progress. This interactive tabletop system facilitates the digital management of protocols right at the bench. It includes a virtual calculator and keyboard, enabling users to conveniently recalculate the concentrations and volumes of reagents needed for pipetting. The introduction of the BioTISCH concept showcased the possibilities of using tabletop systems in experimental research labs, indicating potential areas for further enhancement.

Lastly, eLabBench developed by Tabard et al., 2011, is another variation of the smart table top system, designed to aid researchers in molecular biology laboratories, which was evaluated in a field trial. The eLabBench provides access to relevant experiment information and enables users to annotate and alter these. Besides enabling users to enter data directly via keyboard and mouse, it employed several new interaction mechanisms, among them: hand written annotations, pictures of the bench via a top mounted camera, machine tags to learn about lab machines and the tracking of tube racks (Tabard et al., 2012). While overall targeting general application throughout the whole experimentation process, all works showcased the potential of digital assistance as a direct implementation in the laboratory environment.

2.3 Augmented Reality for Task Assistance

The eLabBench project, along with related initiatives, utilised large screens to enhance laboratory research activities. In contrast, the study by Hile et al., 2004 adopted a projector-based system to aid specifically in pipetting tasks. This system was an extension of the Labscape project, developed by Arnstein et al., 2002, which provided a framework for organising experimental plans. Hile et al., 2004 developed a system employing vision techniques to track specially marked microplates and pipettes in real-time on a lab bench. The tracked pipette tip could be used as a pointer, serving as a TUI, directly interacting with the projected information on the microplate and workbench. This setup enabled accurate tracking of liquid dispensing into specific wells and monitoring of pipetting progress. While user feedback was generally positive, the system's technical performance was constrained, achieving only 4 frames per second when displaying the User Interface (UI). Additionally, the visual tracking of the pipette often interfered with the natural holding position preferred by laboratory workers.

Apart from the work by Borriello, 2006, which elaborates on the real-world application of Hile et al.'s project, research on AR for pipetting task assistance remains scarce. Other studies have concentrated on general experimental support in wet labs, such as using smart glasses like Google Glass for documenting and recalling experiment steps (Scholl, Wille, et al., 2015), and tracking reagent tubes (Scholl, Schultes, et al., 2015). However, the potential of AR for educational scenarios in the laboratory setting has caught the interest of researchers (Barrow et al., 2019). Research has demonstrated that AR technology can enhance the use of instruments and overall experiment experience by providing direct and relevant feedback (An et al., 2019; Kapp et al., 2022).

Given the limited research of AR task assistance in the laboratory domain, it's crucial to explore adjacent areas, particularly medical and surgical applications, to understand the application of AR in high-precision settings. Venkatesan et al., 2021 provide an extensive overview of AR and Virtual Reality (VR) technologies across various biomedical fields, including data visualisation, surgery, and education. In surgery, the potential of AR is particularly notable, with positive outcomes observed in superimposing planned surgical paths for both human and robot-assisted procedures, as detailed in studies by Iqbal et al., 2021; Schreiter et al., 2022; Schwenderling et al., 2021. In the realm of VR, Ali et al., 2022 demonstrated that using arrow-textual aids in a virtual chemistry lab significantly enhanced student performance, reducing task execution time, error rates, and subjective workload.

In exploring AR display technologies for task assistance, Baumeister et al., 2017 compared four displays, finding Spatial Augmented Reality (SAR) using a projector to outperform Video See-Through (VST), OST HMDs as well as a traditional monitor-based display. This advantage of in-situ SAR was also evident in assembly tasks (Büttner et al., 2016; Funk et al., 2016) and medical procedures like needle insertion (Heinrich et al., 2020), confirming its effectiveness across various applications. While OST HMDs may not perform as well as other systems, their wearability and mobility offer significant advantages. Unlike projector-based SAR which requires calibrated setups, OST HMDs are a versatile AR display technology not constrained by such requirements.

Nearly two decades following the work of Hile et al., which inspired previous personal work, detailed in section 2.5, building on the presented findings, especially the benefits of AR task assistance demonstrated in medical applications, as well as the advantages of providing relevant information in laboratory training scenarios, this thesis aims to assess whether ARPAS v2 (see chapter 3) can be effectively used as localisation assistance and information provider to improve pipetting task performance.

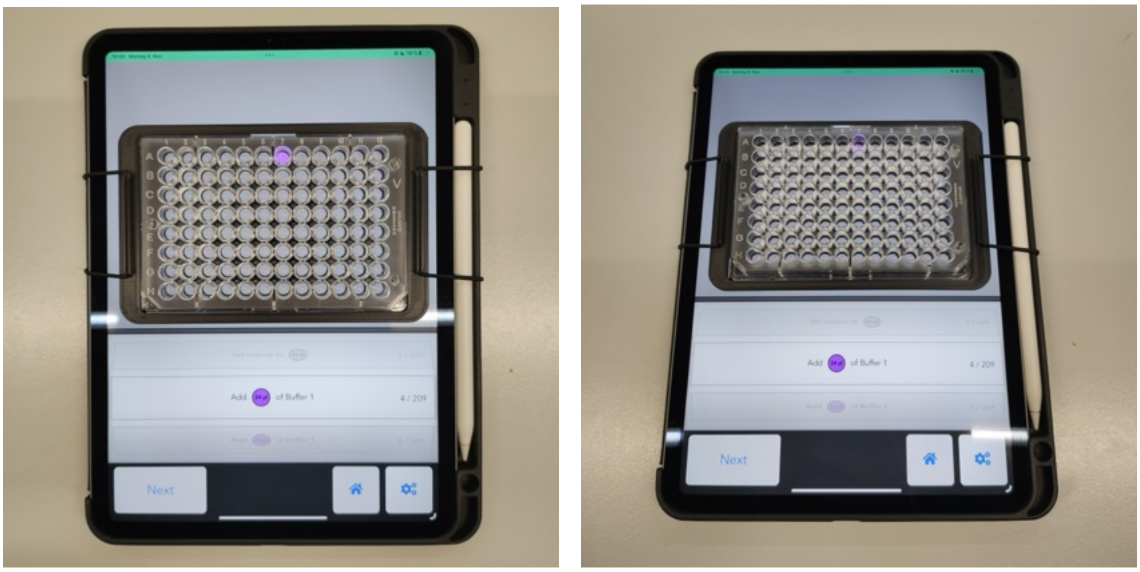
2.4 Applied Work: Pipette Show - An Open Source Web Application to Support Pipetting into Microplates

This section underscores the importance of Pipette Show, a Vue.js-based open-source web application developed by Falk et al., 2022. The application stands out for its open-source nature, contrasting with existing commercial solutions like PlatR by Biosistemika ("PlatR Pipetting Aid," 2022) and TRACKMAN Connected by Gilson ("TRACKMAN Connected," 2023).

Pipette Show was developed in response to the challenges associated with manual pipetting, particularly the risk of errors and the time-intensive process of creating efficient work plans. The application's main feature is its capability to create and execute detailed pipetting plans while offering visual guidance by back lighting relevant wells of a microplate placed on a tablet screen (Figure 2.1). This guidance is instrumental in reducing errors and increasing the efficiency of pipetting tasks, thereby enhancing the reproducibility and reliability of experimental results.

In this thesis, the evaluation of various assistance methods for manual pipetting tasks incorporates Pipette Show as a key example of screen-based pipetting assistance software. Its open-source nature offers a cost-effective, flexible, and adaptable alternative to commercial solutions. Pipette Show's ability to enhance the accuracy and efficiency of manual pipetting, using a tablet-based interface familiar to many users in the general population, positions it as a significant tool in this study. By incorporating Pipette Show as the *Tablet* condition in the study design (outlined in section 4.2), the research gains a more comprehensive scope. This inclusion complements the traditional paper protocol, which represent the conventional method of pipetting, and the advanced AR assistance system (chapter 3) proposed in this work. Pipette Show represents a practical and accessible solution, bridging the gap between traditional methods and cutting-edge technologies in manual pipetting, thereby enriching the study's comparative analysis of task assistance methods.

The Pipette Show Build module ("Pipette Show Build," 2022) played a crucial role in the development of the experimental protocol, as detailed in subsection 4.4.5. Its user-friendly interface facilitated the rapid iteration and refinement of the protocol to meet the updated length and complexity requirements, significantly streamlining the study's preparation phase. The module's export feature, allowing protocols to be saved in a JSON structured format, enabled seamless integration with the ARPAS v2 application. This integration not only ensured data consistency across the experiment setup but also aided in maintaining uniformity in the pipetting instructions across different assistance methods. Furthermore, the module's step-by-step protocol representation provided the structure of the printed paper protocol, thereby standardising the instructions for all study conditions and ensuring equitable textual step information for participants, irrespective of the assistance method employed.



(a) Top view of interface displaying the well illumination technique for pipetting assistance.

(b) Backward-tilted perspective of the interface, emphasising the effect of viewing angle on well illumination.

Figure 2.1. Pipette Show displayed on an 11-inch iPad Pro, used by participants in the *Tablet* condition.

2.5 Previous Personal Work: Augmented Reality Pipetting Assistance System v1

This section explores the development of the 'Augmented Reality Pipetting Assistance System' (ARPAS v1), undertaken during the *HCI Projekt* course from December 2021 to March 2022, supervised by Prof. Dr. Niebling (Lange & Niebling, 2022). Inspired by Hile et al., 2004, the primary objective was to modernise previous augmentation approaches using a mobile HMD platform, thus validating the feasibility of a mobile AR system for pipetting tasks.

The project encompassed the conceptualisation of an intuitive 3D UI to display protocol information and highlight well positions on a microplate, along with the creation of a 3D-printable microplate carrier frame. A Unity application, leveraging the Vuforia Engine ("Vuforia Engine," 2023) and Microsoft Reality Toolkit ("MRTK Unity," 2023) on the HoloLens 2 platform, was developed and subsequently evaluated in a small-scale user study. This study involved comparing the AR system with a traditional paper protocol among microbiology PhD students from the Biocenter at the Julius-Maximilians University Würzburg ("Biocenter, JMU," 2023).

The research question focused on the impact of AR assistance on task performance in manual liquid handling. To quantitatively assess this, three key performance metrics were selected:

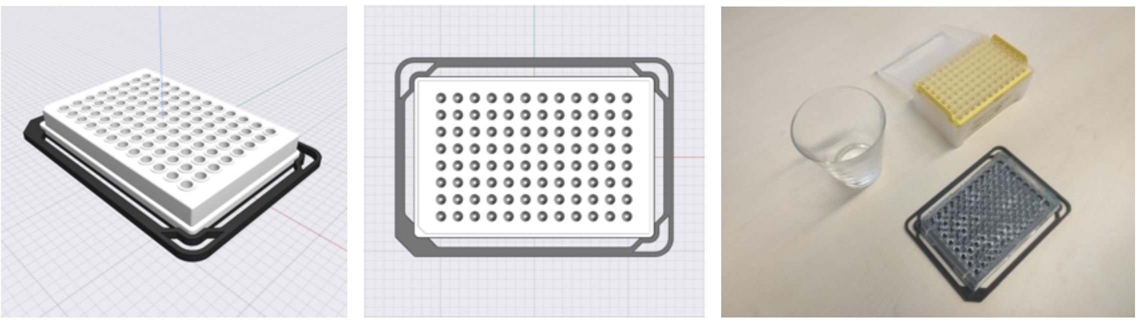
- *Task Execution Time*: The duration each participant took to complete the test experiment, measured in minutes and seconds.
- *Error Count*: The number of misplacement errors each participant committed during the test (e.g., missing wells or skipping steps).
- *Task Load*: The subjective task load of each participant, assessed using the raw NASA Task Load Index (TLX) scores.

These metrics were chosen to provide a comprehensive assessment of the two methods as well as offering insights into their potential advantages in a laboratory setting. Reducing task execution time facilitates more time for result analysis and research progression, while minimising errors ensures the robustness of experimental data. Importantly, the inclusion of subjective task load measurement offers a holistic understanding of the assistance methods' impact. A potential reduction in task load is especially beneficial in professional settings, as it aids in conserving concentration and energy, thereby positively affecting overall work quality and productivity.

2.5.1 Implementation and Use

The design of ARPAS v1 embraced a mobile and unobtrusive approach, crucial for integration into the dynamic and often congested laboratory environment. In departing from the stationary setups of past research, notably that of Hile et al., 2004, this version capitalises on the flexibility of HMD technology, specifically the HoloLens 2, which is renowned for its use in AR research and OST capabilities.

A novel tracking solution was devised by combining a 3D-printed black frame with the microplate (Figure 2.2), addressing the challenge of the plate's transparency and reflectivity regarding visual tracking. This frame, designed through iterative prototyping and practical evaluation, provided a stable and recognisable object target for the HoloLens 2, utilising the Vuforia Engines Model Target ("Vuforia Model Targets," 2023) feature effectively. The chosen frame design balanced the need for a complex, rotationally stable shape with minimal physical intrusion (Figure 2.2b), ensuring ease of adoption within the lab setting.



(a) 3D model of the microplate with its tracking frame, designed as a Model Target for HoloLens 2.

(b) Top view of the Model Target highlighting the distinctive frame outline.

(c) Real-world view of the 3D-printed frame with a transparent microplate, displayed on a tabletop.

Figure 2.2. 3D-printed frame for microplate tracking as a Model Target with HoloLens 2, shown in both rendered and real-world views.

The decision to implement the Mixed Reality Toolkit was driven by its comprehensive UI component library and effective integration of hand tracking and voice recognition, facilitating a more rapid development cycle. The framework's rich documentation and pre-built examples were invaluable in expediting the creation of a user-friendly interface for both touch and voice interactions.

Upon application launch, users encounter an informational screen outlining the control methods and features available, designed to make ARPAS v1 accessible without direct author instruction. The primary interaction uses hand tracking with a holographic palm menu, from which users select pipetting protocols. The application's environment scanning initiates upon protocol selection, anchoring the assistance UI to the physical Model Target upon recognition and signaling successful detection audibly (Figure 2.3b).

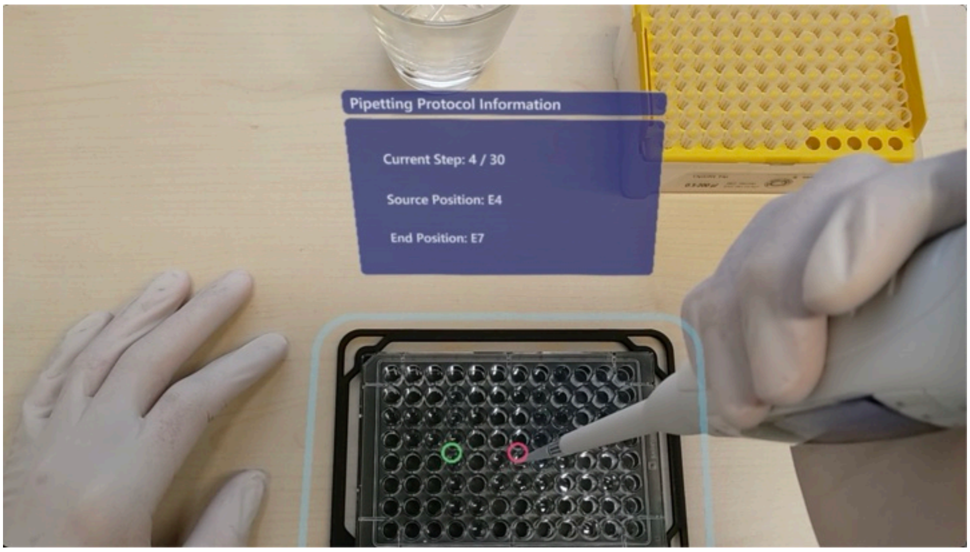
Protocol navigation is voice-activated, enabling users to command "Next" and "Back" to traverse protocol steps, with visual and auditory feedback confirming each action. This system of dual-feedback ensures clear communication between the user and the application, fostering a seamless experience while operating the electronic pipette (Figure 2.3a). Performance evaluations confirmed a consistent 60 frames per second, ensuring a smooth user experience without perceptible lag or stutter.

ARPAS v1's innovation lies in the synergy between the custom-designed microplate frame and the object-based tracking enabled by the Vuforia Engnine on the HoloLens 2. This combination of hardware and software advancements forms the cornerstone of a system that enhances the accuracy and efficiency of manual pipetting tasks with a minimal footprint in the lab environment.

2.5.2 Evaluation

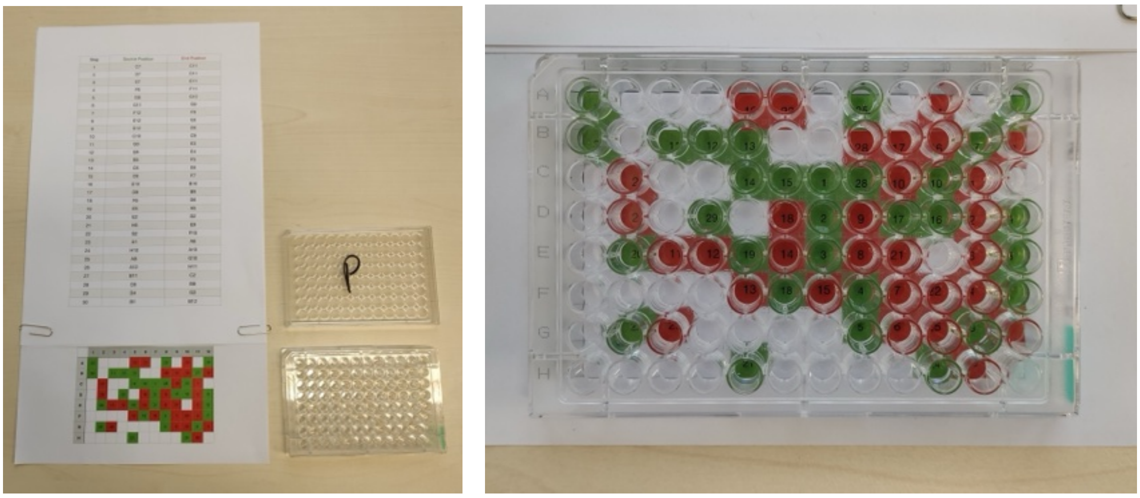
This study formulated three hypotheses to investigate whether there are measurable differences in task execution time, error count, and task load between the two methods. In the experiment, a microplate was set up with 30 wells, each pre-filled with 100 µL of water, designated as 'Source Positions' in the protocol. Participants were instructed to transfer 50 µL from these source positions to specified 'Goal Positions' with the Sartorius Picus NxT electronic pipette (subsection 4.4.3), as outlined in the protocol. The objective was to achieve a microplate with 50 µL of liquid in 60 distinct wells. The accuracy of this task was verified visually against a color-coded plate layout scheme, as illustrated in Figure 2.4b.

(a) Author using ARPAS v1 with Picus NxT electronic pipette, seated at desk wearing HoloLens 2.



(b) First-person view through HoloLens 2, displaying AR augmentations: green ring for source position, red ring for target position, and current step information on the microplate.

Figure 2.3. First and third-person views illustrating the use of ARPAS v1, showing both the AR interface and user interaction with the system.



(a) Experiment protocol for evaluation, detailing transfer instructions (green source to red goal positions) and schematic layout of a 96-well microplate.

(b) Error analysis entailed comparing the microplate against a schematic where each colored field should hold 50µL of liquid. Errors were noted for missing liquid in colored fields or liquid presence in uncolored areas.

Figure 2.4. Printed experiment protocol featuring a microplate schema alongside a microplate placed onto the colored schema for error analysis.

This within-subject study, limited to 9 participants (6 female) due to COVID-19 restrictions, was conducted with microbiology PhD students who regularly perform pipetting. All participants completed the experiment task using both the *Paper* method (printed instructions, Figure 2.4a) and the *AR* method (augmented instructions via ARPAS v1 on HoloLens 2, Figure 2.3). The order of methods was randomised for each participant to mitigate learning effects. To ensure comparability in task complexity while minimising potential learning transfer, distinct yet equally challenging protocol layouts were used for each method.

Participants received training to familiarise themselves with the equipment and procedures. The experiment supervisor recorded task execution times and later counted pipetting errors using a colored plate layout schema. Post-experiment, participants completed questionnaires on task load and usability and participated in semi-structured interviews to discuss laboratory practices and error sources.

2.5.3 Results

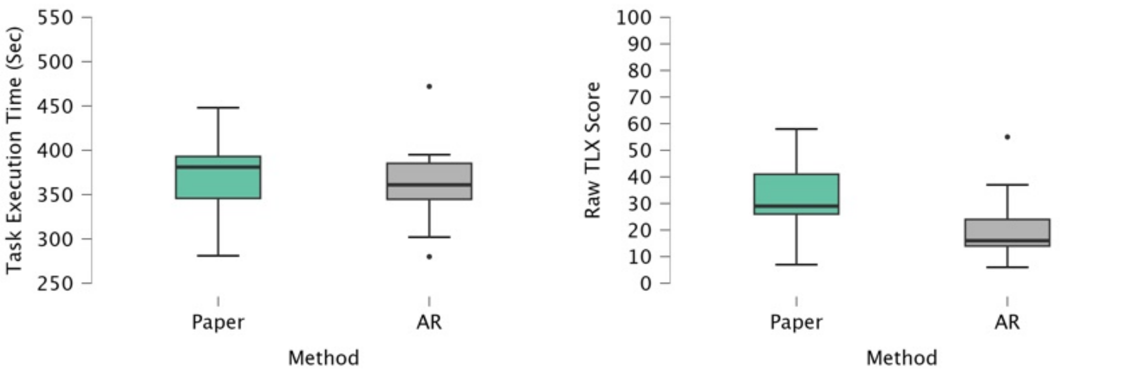
Due to the small sample size and paired nature of the study, inferential statistical analysis was not conducted as it would not yield statistically significant results.

Table 2.1. Descriptive Statistics of Task Execution Time, NASA TLX and Error Count

	Time (Sec)		Raw TLX Score		Error Count	
	Paper	AR	Paper	AR	Paper	AR
Samples	9	9	9	9	9	9
Median	388.00	360.00	29.00	16.00	1.00	0.00
Mean	391.44	345.33	30.77	22.11	0.55	0.00
STD	76.39	78.24	15.66	15.18	0.52	0.00
Sum	–	–	–	–	5.0	0.0

Consequently, statements about the hypotheses are derived from descriptive analysis and should be considered suggestive rather than conclusive. Descriptive statistics indicated a negligible difference in task execution time between the *Paper* and *AR* methods, with averages of 6 : 31 minutes and 5 : 45 minutes, respectively, and similar standard deviations around 77 seconds (Table 2.1, Figure 2.5). However, the *AR* method appeared to reduce task load, as suggested by lower TLX scores (22.11 compared to 30.77 for the *Paper* method). Notably, error count results demonstrated a clear advantage for the *AR* method, which resulted in 0 errors across participants, versus the *Paper* method where 5 errors were recorded in total.

These findings, while derived from a small, paired sample, provided early indications of the benefits of AR in manual pipetting tasks, particularly in error reduction and potential task load alleviation. The next section interprets the ARPAS v1 results, incorporating qualitative feedback and acknowledging the study's constraints.



(a) Boxplot comparing task execution times: mean values and variability across groups.

(b) Boxplot comparing raw TLX scores: mean values and variability across groups.

Figure 2.5. Boxplot of task execution time and task load.

2.5.4 Interpretation, Limitations and Implications for ARPAS v2

The initial findings indicated a potential for AR-assisted pipetting to minimise errors and subjective workload without compromising speed compared to conventional methods. Notably, 7 out of 9 participants were new to AR, yet they adapted quickly to the AR system within the study's time-frame. However, the protocol duration was shorter than typical lab sessions, which were stated to be around 15 to 45 minutes by the participants. Key sources of pipetting errors identified in qualitative interviews — well localisation and concentration fatigue — correlate with the observed benefits of AR assistance.

Feedback from participants on ARPAS v1 highlighted two main issues: the voice control mechanism and tracking stability. Users found the repeated use of "Next" and "Back" voice commands to be somewhat tedious and noted occasional unreliability, with commands needing to be repeated for recognition. Furthermore, tracking inconsistencies were observed, as participants experienced shifts in augmentations when the Model Target's tracking was lost mid experiment. This occasionally led to reduced confidence in the system's overall reliability.

The study's limitations, which this thesis aims to address, include:

- *Sample Size and Study Design:* A larger participant pool with independent group testing is necessary for more generalisable results.
- *Participant Expertise:* Focusing on professionals with extensive pipetting responsibilities could better demonstrate the system's utility.
- *Protocol Duration:* To assess the impact of concentration fatigue over time, the pipetting task should be extended to mirror real-life laboratory sessions.
- *Real-world Relevance:* Adapting the protocol to reflect actual laboratory procedures, such as including inter-container liquid transfers, would improve the study's practical applicability.
- *User Experience Enhancements:* Refinements in tracking performance, well highlighting, system feedback, and control are essential for a more intuitive AR system especially for novice users.

This thesis expands on ARPAS v1 by incorporating a larger sample with professional background, extending protocol duration, aligning tasks with real-world lab activities, and enhancing the user experience. The introduction of Pipette Show (section 2.4) as an additional assistant method offers a broader comparative framework. The forthcoming chapter details the development and implementation of ARPAS v2, informed by the insights and technological limitations from the previous iteration described here. The methodological limitations are addressed in the improved study design and evaluation methods described in chapter 4.

3 Augmented Reality Pipetting Assistance System v2

This chapter delves into the detailed functionality, interface design, technical architecture, and user interaction flow of the evolved 'Augmented Reality Pipetting Assistance System' (ARPAS v2). It aims to provide a clear understanding of how ARPAS v2 represents an advancement and deviation from its predecessor.

The evaluation of ARPAS v1 highlighted three key areas for improvement in its subsequent iteration. Firstly, enhancing tracking stability was crucial to ensure consistent functionality and bolster system reliability. Secondly, the user control mechanism required redesigning to overcome the challenges of the previous version's unreliable and repetitive voice controls. Lastly, the method of augmenting the target well on the microplate needed refinement to strike a balance between being distinct yet not overly obtrusive or distracting.

3.1 Adaptation of the Pipette Show Protocol Structure

Another key focus for ARPAS v2 was enhancing real-world applicability by transitioning from intra-container to inter-container liquid transfers, mirroring typical laboratory setups where reagents are stored in separate containers from the microplate used for analysis. The Pipette Show project, detailed in section 2.4, was identified as an ideal solution for this evolution. Its Build module enables the straightforward creation of experiment protocols with customisable substances and plate layouts.

Protocols exported in the proprietary *.pip* format are easily converted to JSON, aligning with JSON's structure. Since the protocol structure is not detailed in Falk et al., 2022, analysis and reverse engineering were employed to understand its format. This analysis led to the development of software modules for the ARPAS v2 code base, enabling the conversion of Pipette Show JSON protocols into C# class structures. These structures enable ARPAS v2's functionality, displaying relevant step information and highlighting the correct well on the microplate. By adapting the Pipette Show protocol, ARPAS v2 integrates straightforward protocol creation with advanced AR visualisation on a mobile HMD. The specific protocol used in ARPAS v2's evaluation is discussed in subsection 4.4.5.

In the current version of the system, JSON protocols are stored in the *StreamingAssets* folder of the Unity project, integrating them directly into the application. While not addressed in this work, the future development of a webservice to supply protocols via an Application Programming Interface (API) is both feasible and recommended. Such an enhancement would allow seamless interaction with the Pipette Show Build module, creating valuable synergies for future versions.

3.2 Marker Based Tracking With Updated Frame Design

While ARPAS v1 used a Model Target for tracking the microplate on a 3D printed frame, the approach faced tracking stability issues, potentially influenced by varying ambient light conditions, tabletop contrasts, and user movements.

To address these challenges, ARPAS v2 switched to a more traditional, marker-based tracking using the Vuforia Engine's Image Target feature ("Vuforia Image Targets," 2023). In early development tests, this method showed increased reliability under various conditions. These conditions included different ambient lighting, quick head movements, and instances where the user's hands and forearms partially obscured the Image Target. However, this shift required a careful consideration of the marker's size and placement within the workspace. The goal was to ensure effective tracking without hindering the pipetting process.

The initial frame design was refined to incorporate a set of eight neodymium magnets (Figure 3.1a), enabling multiple frames to be effortlessly connected in various orientations. One frame holds the microplate, while the other contains the Image Target marker, sized identically to the microplate (Figure 3.1b, Figure 3.1c). This design not only facilitates seamless integration of the new marker-based tracking system but also maintains the continuity in the overall product design. Additionally, the magnetic attachment system was developed with future scalability in mind. It allows for potential support of multiple microplates by enabling the frames to be arranged in a flexible, grid-like pattern, paving the way for enhanced system extensibility (Figure 3.1d, Figure 3.1e).

(a) In-print frame with exposed neodym magnet array.

(b) Two frames attached in horizontal orientation, upper frame with Image Target marker.

(c) Microplate placed on top of connected frames ready for use in ARPAS v2.

(d) Potential grid configuration for two microplates.

(e) Potential grid configuration for three microplates.

Figure 3.1. Updated tracking frame design: printing phase and layout configurations with finished frames.

3.3 Improvements of the User Interface and Application Control

ARPAS v2's UI was significantly revamped from its predecessor (see Figure 2.3b). The update eliminated voice control for navigating the pipetting protocol, replacing it with two holographic buttons for forward and backward navigation (Figure 3.3a). Voice control was not implemented as a fallback mechanism to maintain consistency and comparability among participants' interactions. Its inclusion could have introduced unnecessary confusion and variability. This approach prevented a split in the participant sample where the use of different interaction methods within or between participants could have led to varied results.

Button placement could be inverted via a palm menu option, to serve right- and left handed users equally, a function also found in Pipette Show. A compact step counter was also introduced, displaying both numerical progress and current step details. Positioned above the microplate, it is angled backwards for optimal visibility when users look down at their workspace. This design mirrors the utilitarian and space-efficient UI of Pipette Show (Figure 2.1), ensuring consistency and facilitating a meaningful comparative analysis.

Incorporating the Pipette Show protocol structure, ARPAS v2 mirrors its tablet counterpart's four main UI states: displaying the protocol's name at the start (Figure 3.3a), showing the selected substance for the next transfer phase (Figure 3.3b), indicating the volume for upcoming transfers (Figure 3.4a), and highlighting target wells during transfer steps with a colored indicator at each well's center (Figure 3.4b).

A significant UI enhancement was the redesign of the well indicators on the microplate, a key feature for precisely highlighting target wells. During early development, various indicator styles were created and tested. This process involved experimenting with different features, such as placement (inside or outside the well), shape (vertical, lateral, small, or large, solid, hollow), resulting in six distinct styles: disc, capsule, cone, pin, ball, and half-cylinder, all depicted in Figure 3.2. These were assessed on the HoloLens 2, focusing on their visibility within the microplate, the extent to which they either obscured the presence of liquid or the pipette tip when inserted into the well.

This exploratory testing led to the selection of the disc-style indicators for further implementation. Contrasting with the ring-style indicators of ARPAS v1, which were placed atop each well, the new disc-style indicators are positioned unobtrusively at the center bottom of each well and occupy a smaller footprint. This change addressed feedback from participants who found the previous ring indicators, though highly visible, to obstruct the view of the pipette tip during insertion, making it challenging to judge the tip's position within the well. The new disc indicators, with their reduced size and placement at the bottom, eliminate this issue, offering a clear view while maintaining visual recognisability. Comparative top and side views are shown in Figure 3.2b and Figure 3.2c) respectively.

3.4 Summary

This section explored the significant enhancements made in ARPAS v2, which were driven by insights from the evaluation of its predecessor, ARPAS v1 (section 2.5), and the integration of external software capabilities of Pipette Show (section 2.4).

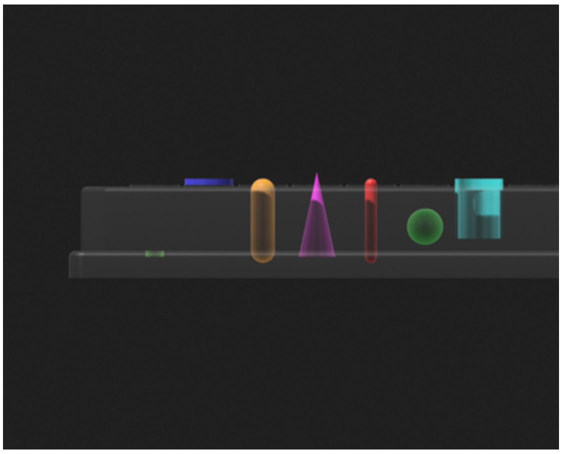
ARPAS v2 key improvements include:

Enhanced Tracking Stability: Transitioning from Model Targets to a more reliable marker-based tracking system using the Vuforia Engines's Image Target feature. This change was influenced by the need for consistent tracking performance in diverse environmental conditions.

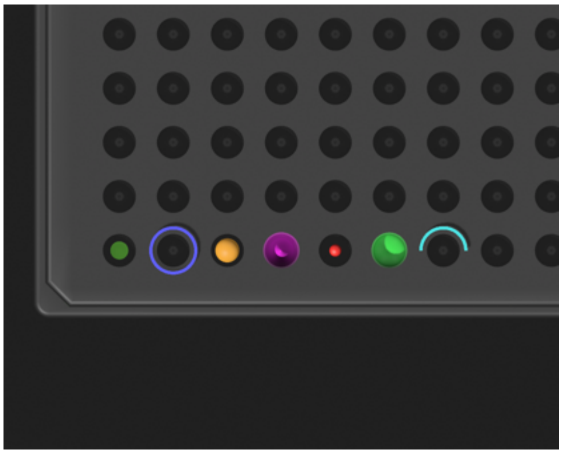
Improved User Interface: The UI redesign involved replacing voice controls with holographic buttons and introducing an informative step counter. This new interface was inspired by the user-friendly and space-efficient design of the Pipette Show, ensuring ease of use and less visual obstruction.

Integration with Pipette Show: The Pipette Show protocol structure was adopted for its inter-container liquid transfers. This alignment with an existing, accessible protocol creation tool not only simplified the process of protocol generation but also set a foundation for future synergies and potential integration of the two systems.

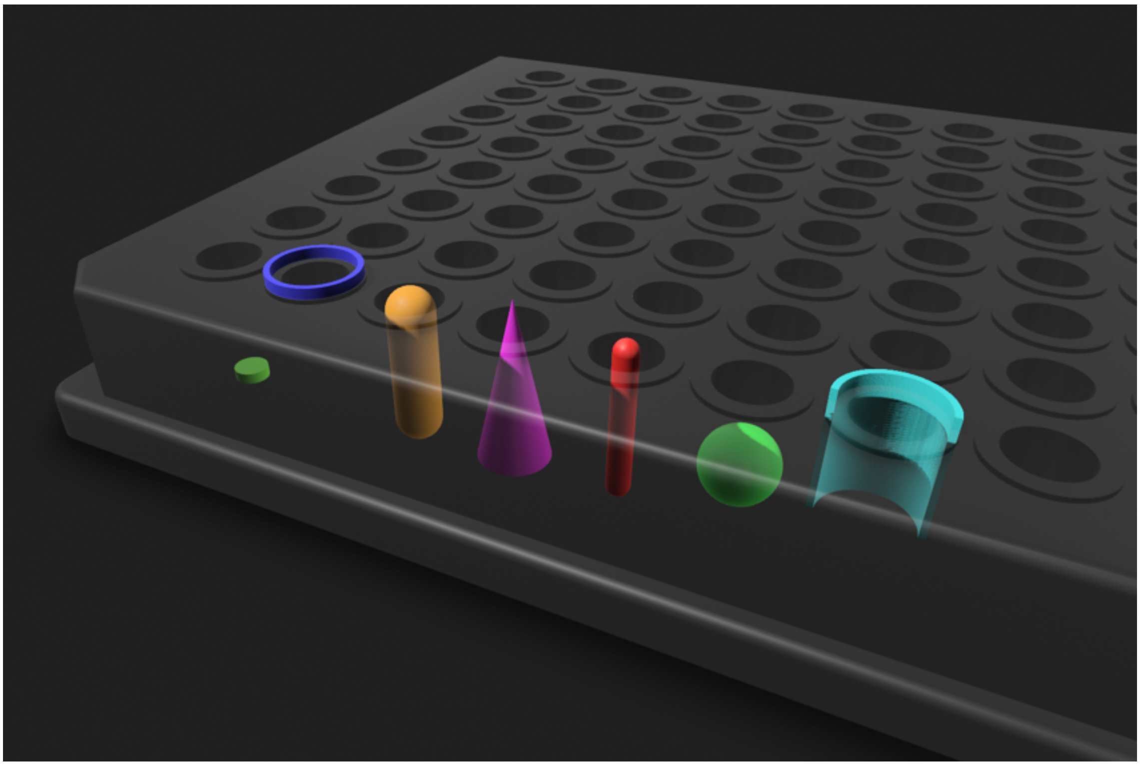
Refinement of Well Indicators: Addressing user feedback, the well indicators were redesigned from ring-style to subtle disc-style, placed at the bottom of each well.



(a) Side view: Disc and ring indicators remain within their respective lateral planes, while the other indicators extend across all three dimensions, resulting in a more prominent overall appearance.



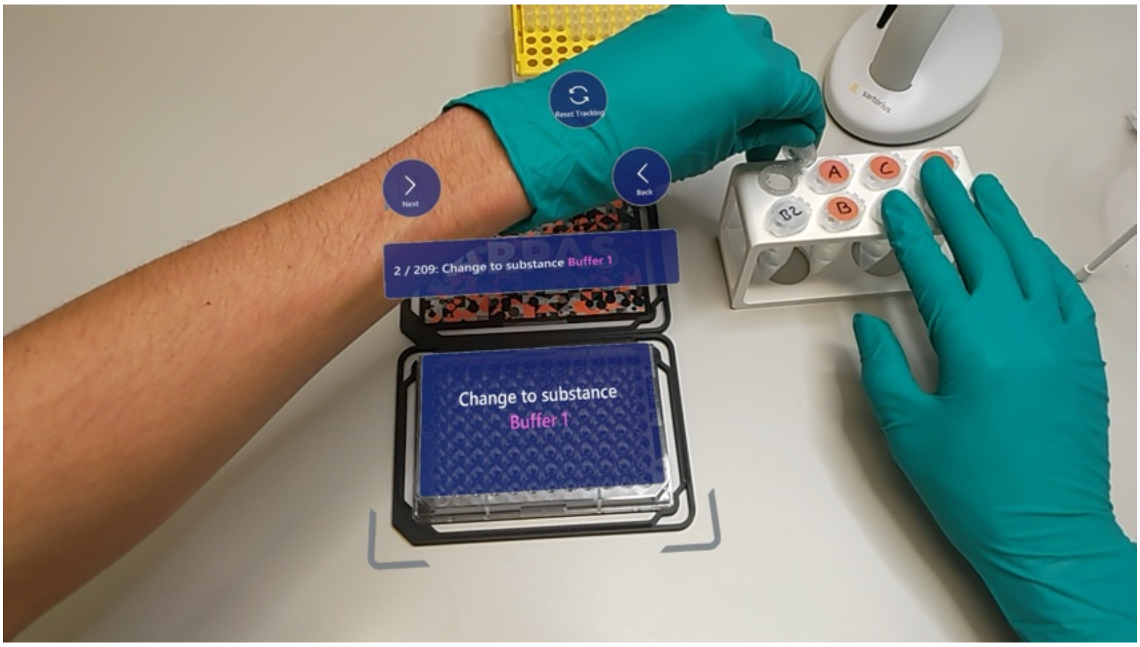
(b) Top View: Capsule, cone, and ball indicators fully occupy the well space, while disc and pin are solid and centered with minimal well space usage; ring and half-cylinder are hollow, situated along the well's perimeter



(c) 3D View: The leftmost disc-style stands out for its small footprint, located at the bottom of the well, ensuring clear well observation and minimal interference with pipette tip.

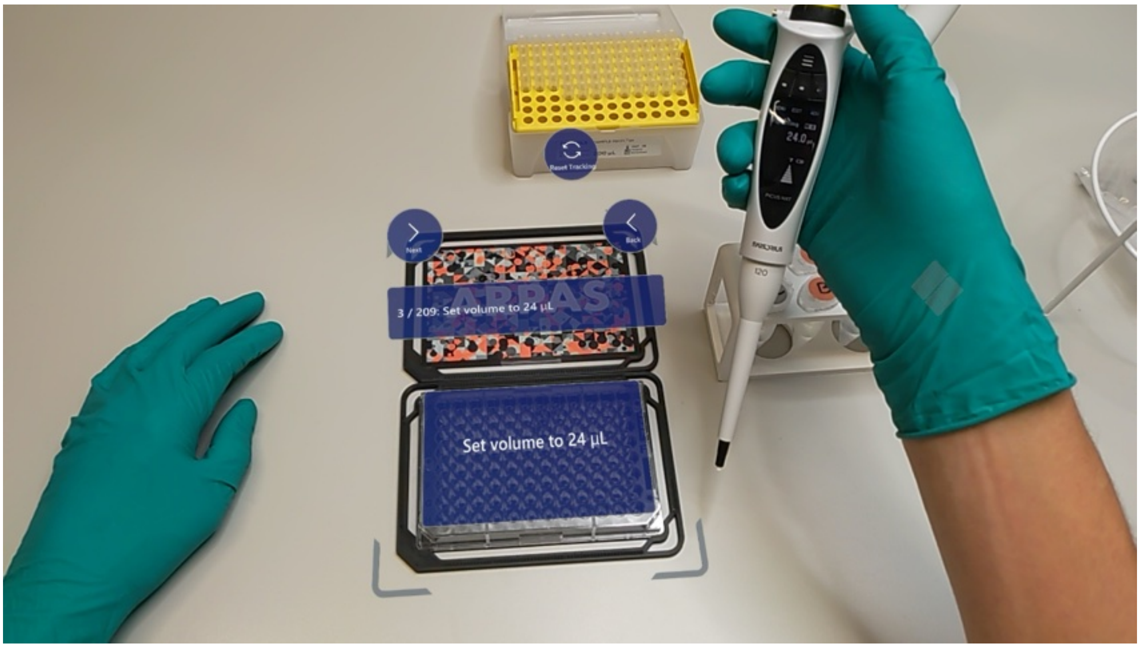
Figure 3.2. Rendered images comparing different well indicator styles in top, side and 3D perspective. Left to right: disc-, ring-, capsule-, cone-, pin–, ball–, and half-cylinder-style. Disc-style was chosen for implementation in ARPAS v2.

(a) Screencapture of the UI in the first-step state of the loaded protocol, "Protocol Start", showing protocol name above the microplate and in step counter.

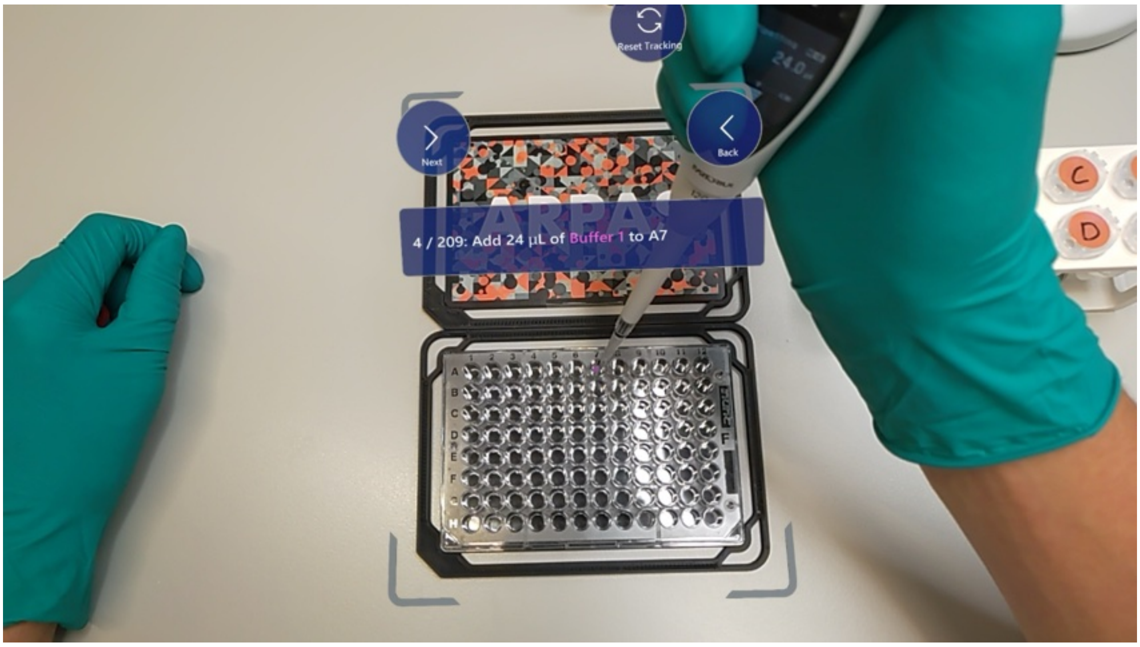


(b) Screencapture of the UI in the second-step state, "Substance Change", showing the currently selected substance above the microplate and in the step counter.

Figure 3.3. Screencaptures from HoloLens 2 with updated ARPAS v2 UI - part I.



(a) Screencapture of the UI in the third-step state, "Volume Change", showing the currently selected volume to be set in the electronic pipette.



(b) Screencapture of the UI in the fourth-step state, "Dispensing", highlighting the target well for dispensing the set volume of the selected substance with a disc-style indicator at the bottom center of the physical well on the microplate.

Figure 3.4. Screencaptures from HoloLens 2 with updated ARPAS v2 UI - part II.

This modification ensures clear visibility of the pipette tip during liquid transfer, enhancing accuracy and user experience.

These advancements in ARPAS v2 demonstrate a commitment to improving AR-assisted pipetting tasks by addressing previous limitations and harnessing the strengths of existing software solutions. This iterative approach underscores the potential for ongoing development and integration in the realm of laboratory AR technology, paving the way for more refined and efficient systems in the future.

4 Methods

This chapter outlines the research methodology, beginning with an explanation of the study's hypotheses and design. It then details the participant selection process and the sample characteristics. The Materials and Apparatus section describes the lab consumables, fluorescent analysis tools, and the software and hardware used in the study, followed by a comprehensive outline of the experimental procedure. This includes the tasks performed, the pre- and post-task questionnaires, debriefing, and the structure of the qualitative interviews aimed at collecting in-depth feedback and insights.

The key measurement metrics, task execution time, error count, and subjective workload, are covered in section 4.6. System usability and user experience assessments are also included, through the System Usability Scale (SUS) and AttrakDiff questionnaire. Finally, the chapter concludes with a description of the statistical methods employed in data analysis, setting the stage for the results presented in chapter 5.

4.1 Hypotheses

This study evaluated three manual pipetting assistance methods: *Tablet*, *AR*, and *Paper*. The *Paper* method served as a baseline for comparison. Both the *Tablet* and *AR* methods are designed to assist in accurately locating target wells on a microplate, aiming to minimise pipetting errors, reduce time, and alleviate subjective workload. Performance impacts of these methods were compared using three primary metrics: task execution time, error count, and subjective workload. For each of these metrics, a specific hypothesis was formulated to investigate the potential differences among the methods.

- **Task Execution Time Hypotheses:**

  - H_{0_Time}: There is no significant difference in the mean task execution time across the three assistance methods (*Tablet*, *AR*, and *Paper*).
  - H_{1_Time}: There is a significant difference in the mean task execution time between at least two of the assistance methods (*Tablet*, *AR*, and *Paper*).

- **Error Count Hypotheses:**
  - H_{0_Error}: There is no significant difference in the number of errors across the three assistance methods (*Tablet*, *AR*, and *Paper*).
  - H_{1_Error}: There is a significant difference in the number of errors between at least two of the assistance methods (*Tablet*, *AR*, and *Paper*).

- **Subjective Workload Hypotheses:**
  - H_{0_Load}: There is no significant difference in subjective workload across the three assistance methods (*Tablet*, *AR*, and *Paper*).
  - H_{1_Load}: There is a significant difference in subjective workload between at least two of the assistance methods (*Tablet*, *AR*, and *Paper*).

4.2 Study Design

In order to asses the hypotheses outlined in the previous chapter, a between-subjects design was chosen to examine the impact of different assistance methods on the performance of laboratory pipetting tasks. The study engaged 48 pharmaceutical technicians, equally grouped into three categories based on the assistance method used: a paper protocol (Figure 4.1a, Figure 4.1d), Pipette Show tablet app (Figure 4.1b, Figure 4.1e), and ARPAS v2 on the HoloLens 2 (Figure 4.1c, Figure 4.1f), with the detailed protocol outlined in subsection 4.4.5. Participants, all Boehringer-Ingelheim employees with their demographics and professional backgrounds provided in section 4.3, were recruited through company internal mailing lists. These communications included detailed information about the study and a link to the Doodle booking platform, where they voluntarily scheduled their 60-minute experiment slot.

To ensure uniformity across sessions, participants were guide by a standardised script detailed in section 4.5. Each session started with a greeting and the presentation of an information sheet, which explained the study and affirmed the voluntary nature of participation. Following this, participants signed consent forms, provided and previously approved by the University of Würzburg's ethics commission.

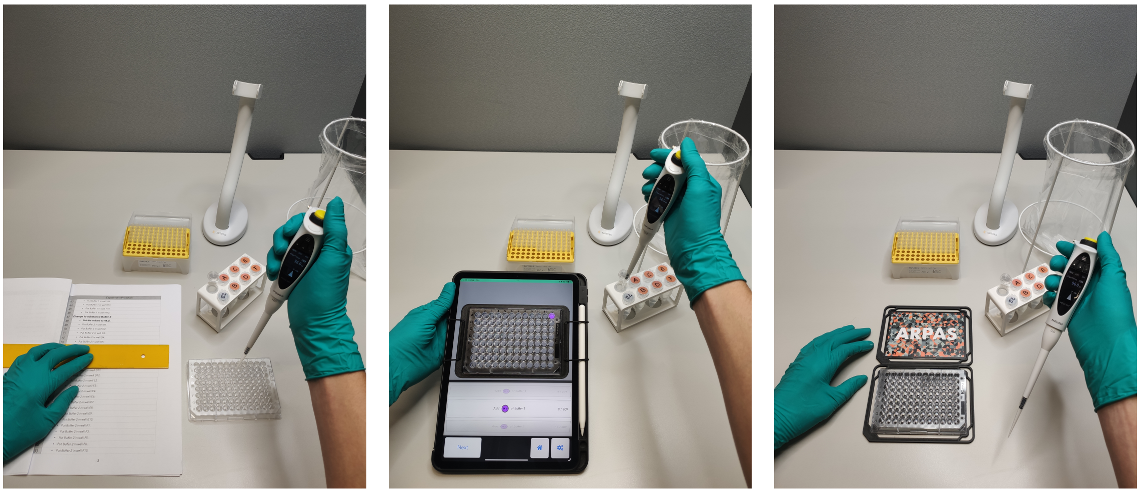
Before the experiment, participants completed a pre-task questionnaire to gather demographic data and assess lab and technical experience (subsection 4.5.1). They were introduced to the lab equipment and setup (subsection 4.4.3 and subsection 4.4.7), followed by method-specific training for their assigned assistance tool, ensuring familiarity with the iPad for the *Tablet* group and HoloLens 2 for the *AR* group.

During the task, participants followed a uniform protocol (subsection 4.4.5) and were instructed to prioritise accuracy over speed, aligning with standard lab practices and controlling for possible speed-accuracy trade-offs (Wickelgren, 1977). Post-task, they completed questionnaires on workload and user experience (subsection 4.5.3), participated in a debriefing about the study's objectives and all assistance methods, and engaged in a semi-structured interview about lab routines and error-prone processes (subsection 4.5.4)

A between-subjects design was chosen over a within-subjects design to minimise the time commitment for the life science professionals involved. While this approach restricted participants from directly comparing each method, it facilitated a straightforward comparison of task performance across different conditions, as all participants followed the same pipetting protocol. The rationale for this design choice and its limitations are further explored in section 6.2.

After participants left the room, the microplate used by the participants was loaded into a Tecan Safire 2 plate reader and an fluorescent analysis of the plate was conducted. Results of the readout were matched to a calibration readout initially done before the start of the study. Significant deviations between the calibration and participants were manually examined and if found to be an error, error count was noted in the experiment protocol for analysis. Detailed information on the use and setup of microplate fluorescent analysis can be found in subsection 4.4.6.

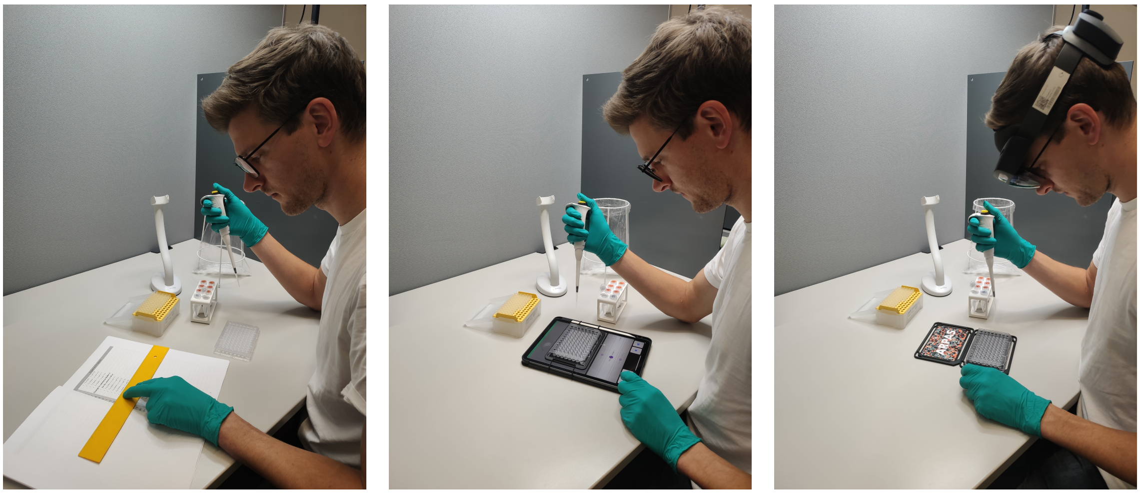
This structured approach was designed to assess the effectiveness of the assistance methods in a controlled, unbiased environment, allowing for a clear evaluation of their benefits and limitations within a laboratory setting. The subsequent sections will detail the sample, materials, procedure, specific metrics and analysis methods used to measure the outcomes of this study design.



(a) First-person perspective on the *Paper* method.

(b) First-person perspective on the *Tablet* method.

(c) First-person perspective on the *AR* method (not through HoloLens 2).



(d) Third-person perspective on the *Paper* method.

(e) Third-person perspective on the *Tablet* method.

(f) Third-person perspective on the *AR* method.

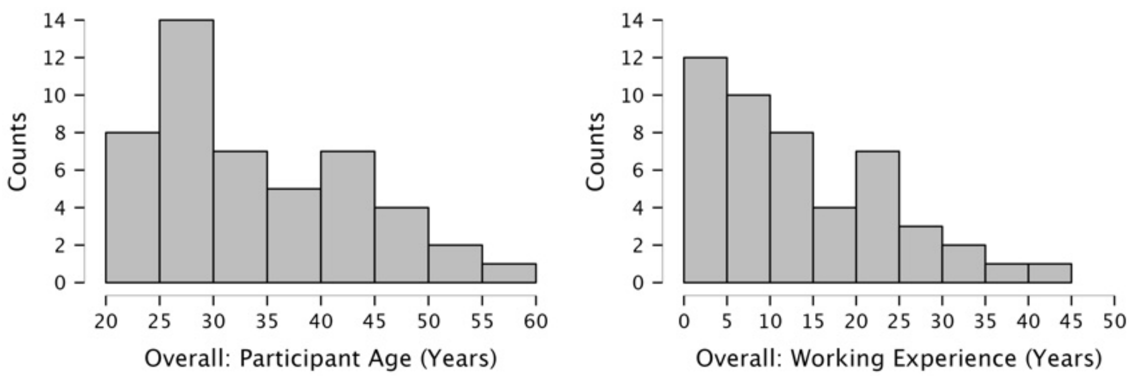
Figure 4.1. Comparison of first and third person views onto the workplace using different methods: *Paper*, *Tablet*, *AR*.

4.3 Subjects and Sample Description

The study included 48 pharmaceutical technicians (29 females, 19 males) from Boehringer-Ingelheim, aged 20 to 57 years (median 32, Figure 4.2a) with 3 to 41 years of work experience (median 12, Figure 4.2b). Figure 4.3 displays plots of participants' age and working experience, categorised by the assistance method used. A visual comparison of these plots indicates a balanced distribution of age and working experience within each group, consistent with the overall distribution of these metrics across all participants.

Figure 4.4 presents participants' familiarity with mobile devices and AR/VR technology, essential for contextualising results in the Discussion (chapter 6). While all participants in the *Tablet* group reported experience with smartphones or tablets, only five in the *AR* group had prior AR/VR experience, predominantly limited to a single demo use. This contrast underscores varying levels of technological familiarity among participants, particularly relevant given the study's focus on both mobile and AR-based assistance methods.

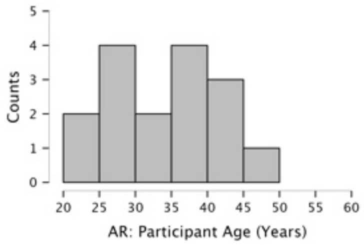
Participants were required to have experience with single-channel pipettes and normal or corrected-to-normal vision. All participants met this criterion, with varying degrees of experience in using both mechanical and electronic single-channel pipettes. Those without any pipetting experience or with non-corrected vision were excluded.



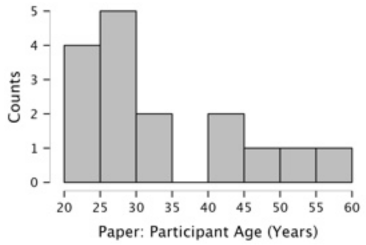
(a) Distribution plot of participant age in 5-year wide bins (median is 32).

(b) Distribution plot of participants working experience in 5-year wide bins (median is 12).

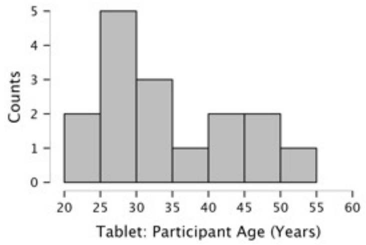
Figure 4.2. Distribution plots of participants age and working experience.



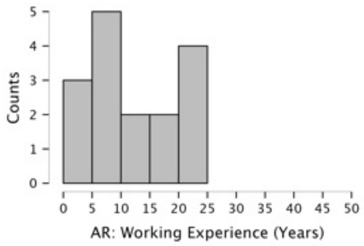
(a) Distribution plot of participant age in the *AR* group (median is 35).



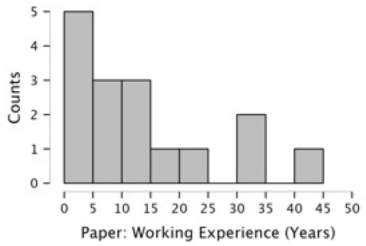
(b) Distribution plot of participants age in the *Paper* group (median is 30).



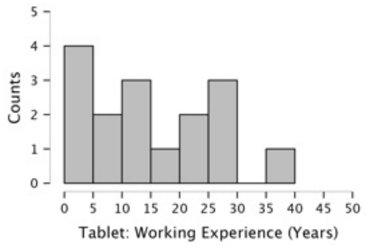
(c) Distribution plot of participants age in the *Tablet* group (median is 33).



(d) Distribution plot of participants working experience in the *AR* group (median is 10.5).



(e) Distribution plot of participants working experience in the *Paper* group (median is 10.5).



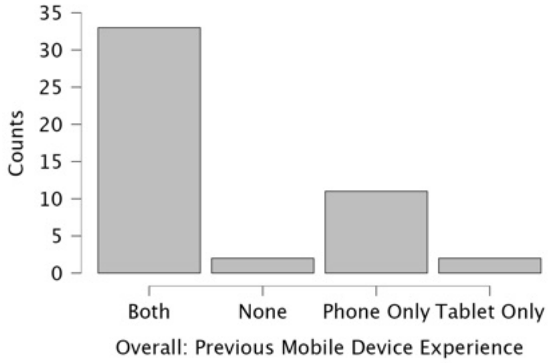
(f) Distribution plot of participants working experience age in the *Tablet* group (median is 13.5).

Figure 4.3. Distribution plots and median values of participants age and working experience split by condition.

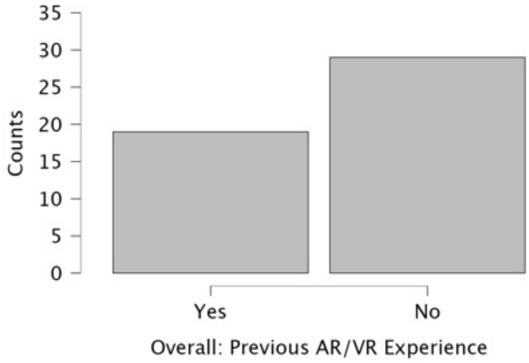
Voluntarily recruited via internal mailing lists, participants presumably had diverse educational backgrounds ranging from vocational training to advanced degrees due to the nature of their professional roles at the organisation. A thorough introduction to the specific electronic pipette and microplate used, detailed in subsection 4.4.3, was provided, with practice sessions ensuring participant comfort.

Participants were offered a chance to win a 50€ Amazon Gift Card, regardless of experiment completion. By self-report, participants were primarily motivated by curiosity and a desire to contribute to the improvement of laboratory processes. Participants were randomly assigned to one of three experimental conditions to mitigate selection bias.

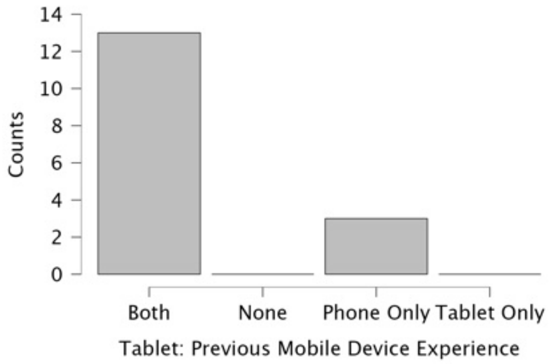
While this experiment was tailored for those with significant pipetting experience in laboratory settings, such as found in large life science companies, similar criteria could apply to personnel in bio-chemical research institutions and universities.



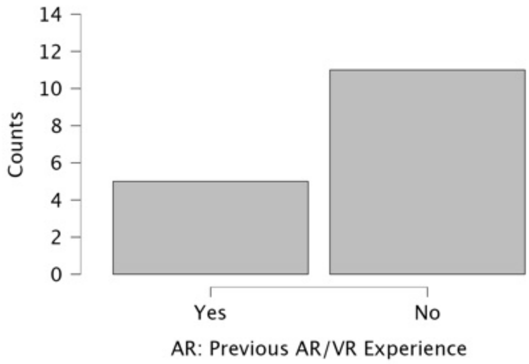
(a) Distribution plot of participants previous Smartphone and Tablet experience.



(b) Distribution plot of participants previous AR/VR experience.



(c) Distribution plot of previous mobile device experience in the *Tablet* group.



(d) Distribution plot of previous AR/VR experience in the *AR* group.

Figure 4.4. Distribution plots of participants previous experience with mobile devices and AR/VR technology.

4.4 Materials and Apparatus

This section is organised into focused subsections that detail the specific software and hardware setups, lab consumables employed, substances and methodology used for fluorescent analysis, structure of the experimental pipetting task and the workbench setup.

4.4.1 Software

This study employed two distinct software systems to represent the *Tablet* and *AR* conditions: Pipette Show, detailed in section 2.4, served as the representative for the *Tablet* condition, while ARPAS v2, as described in chapter 3, represented the *AR* condition. Both systems used the same protocol, generated using the Pipette Show Build module, with comprehensive details available in subsection 4.4.5.

4.4.2 Hardware

Microsoft HoloLens 2

Palumbo, 2022 extensively reviewed the application of the HoloLens 2 in various medical contexts, particularly in surgical navigation, underscoring its superiority to the first-generation HoloLens, both Magic Leap versions, and Google Glass 2. Its enhanced input capabilities, including head, hand, voice, and gaze tracking, were also noted.

Although Condino et al., 2020 identified depth-of-field limitations in the first-generation HoloLens for close-range manual tasks, relevant to this study's experiment, Rieder et al., 2021 observed significant improvements in close-range accuracy and precision in the second generation, aligning with this study's requirements.

Choosing the right HMD involves balancing costs and benefits. The HoloLens 2 was selected primarily because it was readily available at the HCI chair as a development device, having been used in the preceding project, ARPAS v1 (see section 2.5), ensuring technological consistency and leveraging existing development frameworks. The device's capabilities, particularly effective hand-tracking, spatial awareness, OST visor with an adequate Field of View (FoV), and high accuracy and precision, coupled with ease of development, further justified its selection for this study.

Apple iPad Pro

A 2018 iPad Pro 11" was used to display the Show module of Pipette Show, responsible for presenting the pipetting instructions. A simple yet effective setup involves a 3D printed frame, designed to hold the micro well plate, securely attached to the device using elastic cords. This setup ensures precise alignment of the micro well plate with the well illumination graphics in the Show module.

Pipette Show offers its own 3D printable model of the plate holder, available in the project's GitHub repository ("GitHub GBA," 2023). In this experiment, a slightly modified version of the original frame was used, featuring a reduced footprint, anti-slip material on the bottom, and improved anchor points for the elastic cords (Figure 2.1). These modifications, while minor, were made for enhanced practicality and aesthetics without affecting the frame's functional effectiveness.The iPad, equipped with the experiment protocol detailed in subsection 4.4.5, was thus prepared for participant use in the experiment.

4.4.3 Electronic Pipette and Microplate

The Sartorius Picus NxT 5 - 120µL electronic pipette (Figure 4.5a) was selected for its superior functionality over traditional mechanical pipettes. Mechanical pipettes require manual operation of a spring-controlled mechanism for liquid aspiration and dispensing, a skill that demands considerable training and experience. In contrast, the Picus NxT employs an electronic motor, reducing user strain and improving operational consistency and precision. It features the capability to store custom volume presets. This ensures consistent volume adjustments at a similar speed for all participants, enhancing the experiment's efficiency. The four volume presets - 96, 72, 48, and 24µL - are easily accessible via the pipette's preset button (Figure 4.5c). The aspirating and dispensing speeds were uniformly set to level 7 out of 9 for this study, mitigating variance in participants' prior pipetting experience and standardising it for this experiment. Aspirating and dispensing speeds are dependent on the viscosity of the liquid in use - with higher viscosity liquid requiring lower speeds to avoid air gaps in the tip. The liquid used in this experiment is a low viscosity, water-based Phosphate-Buffered Saline (PBS) solution, therefore the speed level 7 was deemed appropriate (Figure 4.5b). The use of Sartorius Optifit Tips 0.5 - 200µL, recommended for this pipette model and sold by Sartorius, further ensures accuracy.

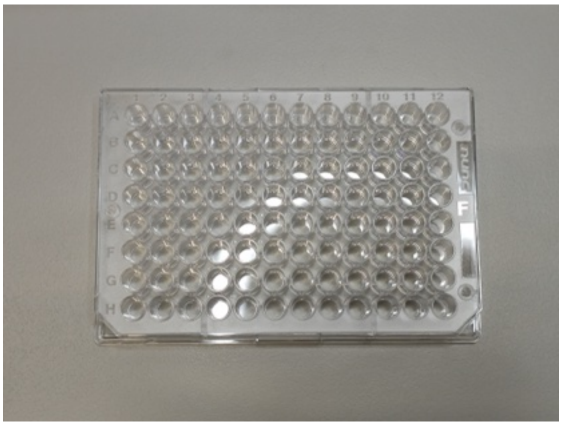
(a) Picus Nxt 5-120µL with Optifit Tip Box.

(b) Pipette screen showing selected volume (24µL) and pipetting speed (7).

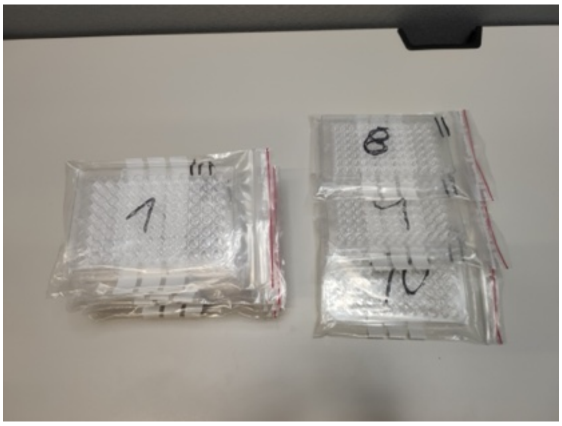
(c) Pipette screen showing the four volume presets accessible via the three-striped button.

Figure 4.5. Sartorius Picus NxT 5 - 120µL electronic pipette.

Microplates are a staple in both automated and manual life science workflows, with their dimensions standardised by the American National Standards Institute (ANSI) and the Society for Laboratory Automation and Screening (SLAS) ("ANSI/SLAS Microplate Standards," 2023). For this experiment, the flat-bottom, transparent 96-well microplate variant was selected (Figure 4.6a). These plates are not only cost-effective and compatible with fluorescent plate readers but also benefit from standardised dimensions, ensuring compatibility and interchangeability. The chosen design is particularly suited for illumination from below, a feature essential in the *Tablet* condition of the study. While various configurations of 96-well microplates exist — including different colors (white, black, transparent), bottom types (optical or non-optical), and shapes (flat, round, or V-shaped) — the transparent, flat-bottom type was found to be the most appropriate for the experimental needs. Adhering to ANSI and SLAS standards allows for flexibility in sourcing plates from multiple manufacturers without compromising consistency. To optimise resource use, a rotating stack of ten microplates was employed, guaranteeing a fresh plate for each participant. These plates were washed and dried daily, then marked to track usage (Figure 4.6b). This ensured uniform wear and prevented residue buildup, maintaining consistent plate quality throughout the experiment.



(a) Top view of the 8x12 alpha-numerical well grid.



(b) Packaged and labeled plates with usage markings.

Figure 4.6. Transparent, flat bottom, 96-well microplate used in the experiment.

This approach to selecting lab consumables ensures a balance of cost-effectiveness, functionality, and compatibility with the experimental setup, crucial for achieving reliable and reproducible results in the study.

4.4.4 Use of Uranine as Fluorscent Dye

This section discusses the use of uranine as a fluorescent tracer dye, and the creation of different uranine solutions crucial for the pipetting task's fluorescent analysis in this experiment.

Uranine, a fluorescein derivative chemically known as fluorescein disodium salt ($H_{10}Na_2O_5$, CAS 518-47-8), is a widely-used fluorophore in various scientific applications due to its high solubility and pH-dependent fluorescent intensity (Hammer et al., 2005; Martin & Lindqvist, 1975). These properties make it ideal for use with PBS, a water-based solvent with a stable pH of 7.4 ("Thermofisher PBS Product Catalog," 2023). This stability ensures that each solution's fluorescence can be consistently measured, a key factor for the experiment's objectives to reliably identify participant‘s pipetting errors in the trial.

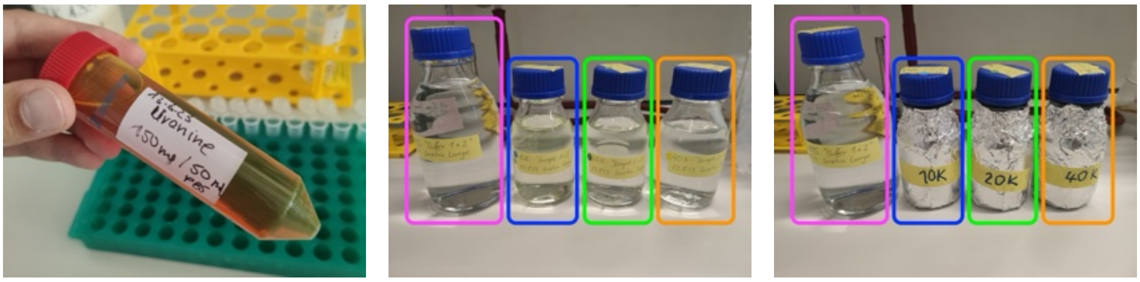
The experiment began with the preparation of a uranine stock solution (Figure 4.7a), created by dissolving 150mg of uranine in 50ml of PBS, resulting in a 7.97 mM concentration,. This concentration was chosen to ensure a robust base for multiple trials. From this stock solution, three dilutions were developed – 1/10000, 1/20000, and 1/40000 or 797nM, 398nM and 199nM – to meet two critical criteria: transparency and measurable fluorescence intensity. Transparency was essential to ensure that participants could not visually distinguish between the substances, while the fluorescence intensity needed to be detectable and result in distinct profiles across the dilutions. These dilutions were selected after thorough experimentation, as they produced a range of fluorescent intensity readings from near-maximum to distinctly above background noise levels when analysed by a plate reader. Figure 4.7b shows created solutions in 500ml and 250ml glass bottles. The required volume of each solution was calculated based on the expected maximum number of participants plus a safety margin.

The creation of these four substances – PBS with no fluorescence and the three uranine solutions with distinct fluorescent profiles – provided the necessary materials for constructing the pipetting protocol described in detail in subsection 4.4.5. This protocol incorporated multiple substance changes and included dilution steps, significantly increasing its complexity compared to the ARPAS v1 evaluation protocol. The transparent nature of all solutions, combined with their distinct fluorescence readings, allowed for an intricate and challenging task design, aligning with the study's goal of evaluating pipetting accuracy and efficiency under various conditions.

Furthermore, considerations were made for photobleaching, a phenomenon where fluorophores lose their ability to fluoresce due to light exposure (Diaspro et al., 2006). To mitigate this, the solutions were wrapped in aluminium foil and stored under conditions that minimised light exposure, preserving their fluorescent properties for consistent measurements throughout the experiment (Figure 4.7c).

4.4.5 Laboratory Experiment Protocol

This section details the experimental pipetting protocol, created using the Pipette Show Build module, which contains the steps participants followed during the task execution phase of the experiment.



(a) Plastic tube containing the 7.97mM uranine stock solution (150mg uranine in 50ml PBS).

(b) Magenta: 500ml pure PBS, Blue: 1/10K, Green: 1/20K, Orange: 1/40K dilution of the uranine stock solution, 250ml each.

(c) Bottles containing the uranine dilution are wrapped in aluminium foil to prevent photo bleaching from ambient light.

Figure 4.7. Uranine stock solution and prepared experiment substances in laboratory glass bottles.

Insights from APRAS v1's evaluation, as discussed in subsection 2.5.3, indicated the need for a more extensive protocol. The initial 30-step protocol, with liquid transfers performed on a 96-well microplate, was completed in under 10 minutes, proving ARPAS v1's basic functionality but highlighting the need for a more challenging protocol to assess the assistance systems thoroughly.

To address this, the new protocol for this experiment was designed with a greater number of steps in a complex layout. This aims to test the limits of participant concentration and to accentuate the potential benefits of the assistance systems over an extended period of time. By increasing the task's complexity, the study intends to provide a clearer distinction in performance outcomes, measuring both task execution time and error count.

The revised protocol now consists of inter-container transfers to better simulate typical laboratory processes, where substances are often transferred from separate containers (e.g. micro tubes) to a microplate. This adjustment brings the experiment's design closer to real-world lab operations.

To facilitate inter-container transfers in the experiment, four distinct liquid solutions were used as described in subsection 4.4.4. Of each substance two, 5 mL, micro tubes were filled. To the naked eye, the resulting eight tubes contained transparent liquid, indistinguishable from another. This design choice, leveraging the challenge of working with colorless liquids, made it difficult for participants to visually track the liquid's placement and amount on the microplate. The eight micro tubes were labeled with a unique identifier and put into a tube rack, seen in Figure 4.8.

Although there are really only four different solutions, participants were informed that each tube contained a distinct liquid solution. This approach not only obscured the actual number of solutions to the participant but also added complexity through frequent micro tube changes and the possibility of mixing liquids from a wider range of containers. With the liquids prepared as material, the individual pipetting steps of the protocol are planned.

Figure 4.8. Rack for 5 mL micro tubes, labeled B1, B2, A-F. Magenta: pure PBS, Blue: 1/10K, Green: 1/20K, Orange: 1/40K uranine dilution.

In this experiment, a single pipetting step consists of aspirating a predefined volume of a liquid out of a labeled micro tube and dispensing it into a specific well on the microplate, identified by an alphanumeric index (A1-H12, Figure 4.6a). To maintain a controlled complexity level, the study focuses solely on the liquid transfer process, excluding additional real-life pipetting actions such as pre-wetting the tip or mixing / stirring the liquid in the target well. This decision was made to minimise variability that could arise from participants' differing pipetting experiences and to concentrate on measurable aspects like task execution time and error count.

Consistent volume in each well is crucial for reliable results in the fluorescent analysis, as described in subsection 4.4.6. The analysis, calibrated for a fixed volume in all wells across the microplate, allows for precise identification of pipetting errors. To achieve this consistency, each well was required to contain exactly 96µL of liquid.

The choice of 96µL was informed by several factors. Most 96-well microplates can hold 120 to 200µL per well, making 96µL a practical volume for various plate types. The volume range of the used electronic pipette is 5 to 120µL (subsection 4.4.3); thus, a 96µL target volume uses 80% of its capacity, minimising volume variability. Additionally, 96µL can be evenly divided into four fractions - 96, 72, 48, and 24µL - allowing for diverse transfer techniques in the protocol.

The protocol incorporates four types of transfer techniques to vary complexity:

1. Transfers:
   a) Single-step: transferring 96µL from the substance tube to the target well.
   b) Two-step: dividing the target volume between two transfers from two different substance tubes to the well containing the same substance, either 72 + 24µL or 48 + 48µL.
2. Dilutions:
   a) Two-step: involving two single transfers of each a fluorescent sample and a diluent; 72µL of fluorescent sample and 24µL of PBS.
   b) Three-step: combining a single transfer of the fluorescent sample with two transfers of the diluent; 24µL of fluorescent sample and 48 + 24 µL of PBS.

A 'two-step transfer' involves filling a well in two steps with different volumes from micro tubes labeled as separate liquids but containing the same solution. Combining 72µL and 24µL from two tubes with the same substance (e.g. B1+B2, A+B, C+D, E+F, Figure 4.8) yields the expected 96µL well volume. This approach effectively doubles the steps needed to fill one well. The 'two-step dilution' process combines 72µL of a fluorescent dye solution with 24µL of PBS, thereby diluting the uranine concentration by a factor of 0.75. A more complex variant, the 'three-step dilution', involves dividing the diluent volume into two separate transfers (e.g., 24µL uranine solution + 48µL + 24µL of PBS from differently labeled tubes). The fluorescent concentration is thereby diluted with factor of 0.25.

All transfer techniques aim to fill each well with the same amount of volume, but the latter steps increase the protocol's length and complexity by requiring additional actions to achieve this goal.

While the minimum number of steps to fill all wells on a microplate with a single substance is 96, the introduction of these varied transfer techniques in combination with eight micro tubes, inflated the protocol to 208 steps. Steps counting substance and volume changes as well as the liquid transfers themselves. The schematic plate layout showing the substance composition of each well can be seen in Figure 4.9a, with a segmentation view differentiating the different transfer techniques in Figure 4.9b.

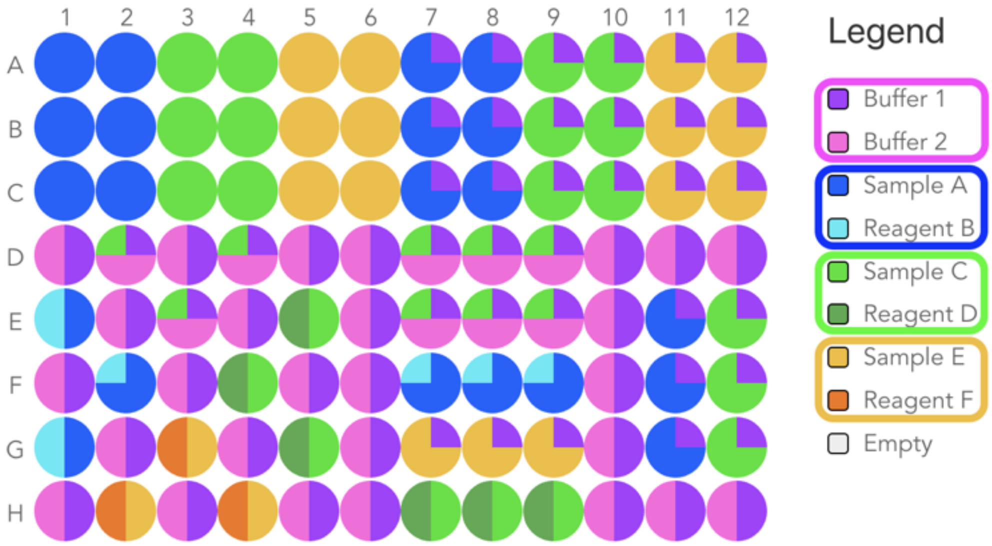
This extended procedure, amounting to approximately 30 minutes of manual pipetting in a self-test, was designed to evaluate the effectiveness of the different assistance methods - *Paper*, *Tablet* and *AR*. This duration was considered optimal for observing potential concentration degradation over time and its influence on task execution time and number of pipetting errors.

4.4.6 Fluorescent Analysis

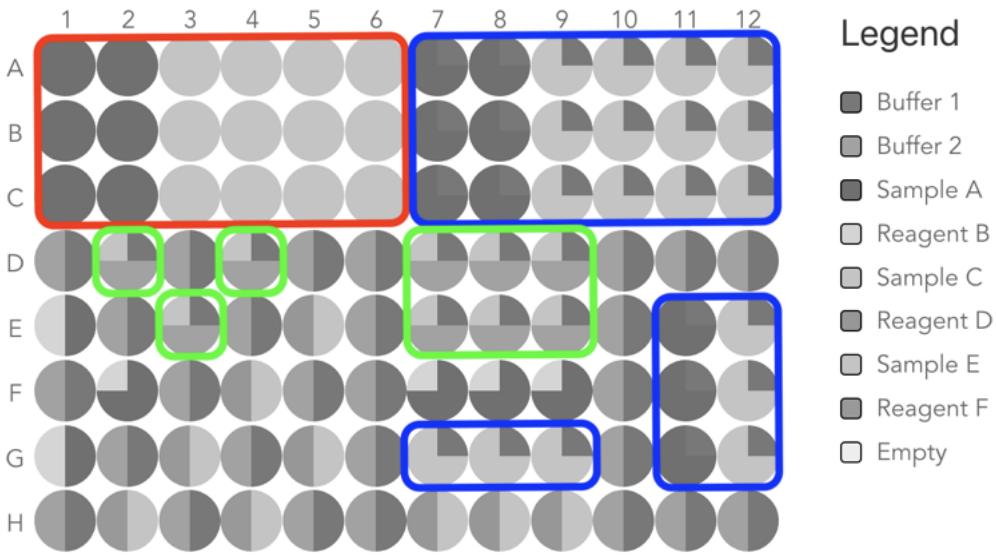
This section focuses on the Tecan Safire 2 microplate reader's use and setup, key for assessing participants' performance in the experiment. It explains the principles of plate readers and fluorescence intensity measurement, followed by an exemplary microplate analysis. This includes the comparison of calibration plate and participant trial results, illustrating the method for quantifying pipetting errors, which ties in with the previously described experimental protocol for a thorough evaluation of participant trials.

Fluorescence intensity measurement is based on the principle of excitation and emission. Fluorophores, like uranine used in this experiment, absorb light at a specific wavelength and emit light at a different, typically longer wavelength (Figure 4.10b). A microplate reader operates by directing a beam of light at a sample, which then absorbs and re-emits light (Figure 4.10a). The device measures the intensity of the emitted light, effectively quantifying the fluorescent dye concentration in each well (BMGLabtech, 2023).

The design of the pipetting protocol aligns with the microplate reader's functionality, which enables precise measurements of liquid in each well on the microplate. If no errors were made, each well should contain 96µL of either pure PBS or one of

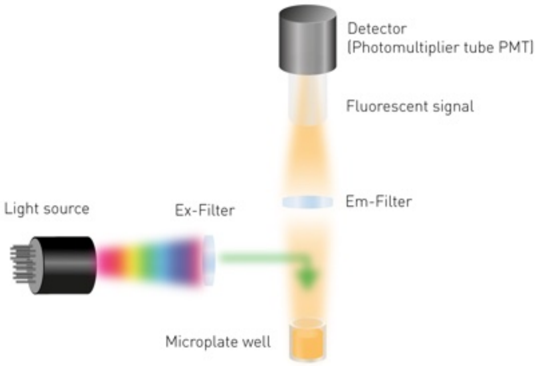


(a) Plate layout with color coded substances and their composition per well. Composition is based on 96µL and its four fractures - 96, 72, 48, 24µL. All eight substances used in the layout are listed in the legend on the right. Colored borders indicate which individually labeled liquids are derived from the same substance. Magenta: PBS, Blue: 1/10K, Green: 1/20K, Orange: 1/40K uranine dilution.

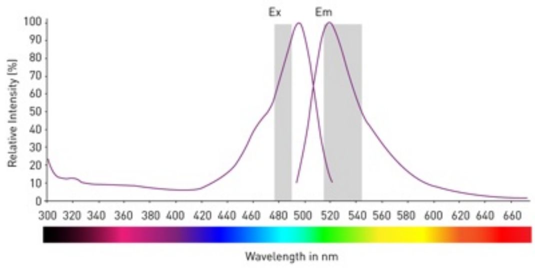


(b) Desaturated plate layout with colored borders grouping blocks of different transfer techniques. Red: single-step transfer (no dilution), Blue: two-step dilution (0.75 dilution factor), Green: three-step dilution (0.25 dilution factor). All remaining wells are two-step transfers adding the same substance but from differently labeled tubes (no dilution).

Figure 4.9. Top: Schematic layout of experiment protocol detailing substance composition in each microplate well. Bottom: Plate layout overlaid with segments differentiating transfer techniques.

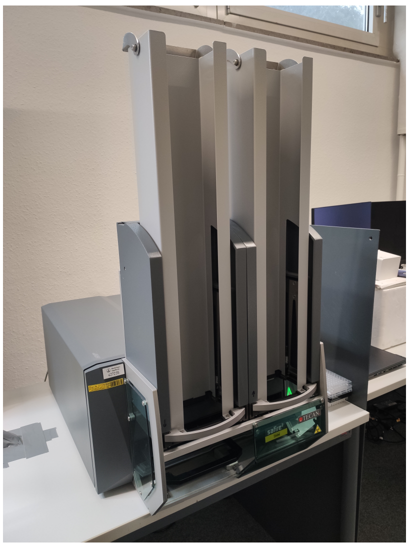


(a) Simplified design of a fluorescence intensity detection system in a microplate reader, featuring a light source, filters, the read well, and a detector for emitted light intensity.



(b) Schematic diagram displaying excitation and emission filters for fluorescein, overlaid on corresponding spectral graphs.

Figure 4.10. Schematic of plate reader's internal construction and spectral properties of excitation and emission filters. (BMGLabtech, 2023)



(a) Full view of the Safire 2 microplate reader.

(b) Extended plate carrier loaded with a microplate ready to be measured.

Figure 4.11. Tecan Safire 2 microplate reader used for fluorescence intensity measurement in this experiment.

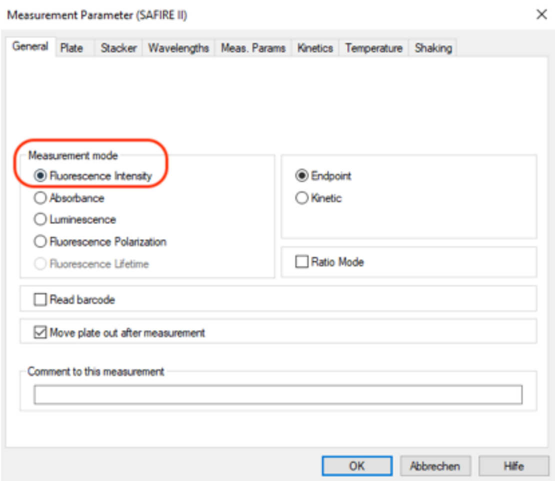
the different uranine solutions designated for that specific well. By analyzing the fluorescent intensity in each well, the microplate reader can effectively detect and quantify any discrepancies in the composition of that well.

Based on the pipetting protocol, the liquid composition of each well or a group of wells is known and can be linked to a certain measurement of fluorescence intensity in that well. The output of fluorescence intensity measurement by a plate reader is quantified as Relative Fluorescence Unit (RFU). As the name suggests, RFU values are relative and subjective to the plate readers configuration and setup.

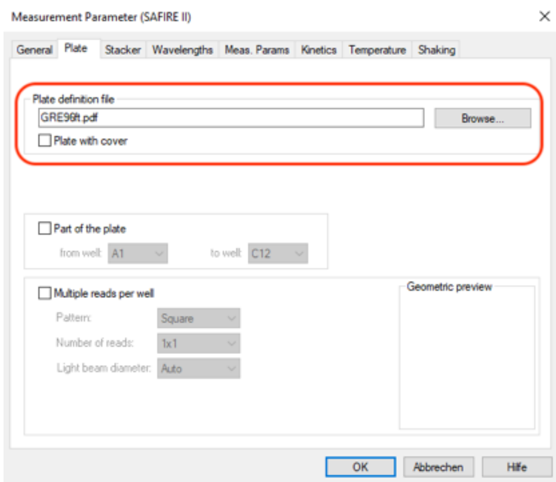
For comparable results of all participants plate measurements, the plate reader needs to be configured exactly the same throughout the experiment. The plate reader is configured via a software interface and the setup is saved as a configuration file. An overview of the relevant configuration parameters is shown in Figure 4.12. First, *Fluorescence Intensity* is set as the readers mode (Figure 4.12a), then the plate definition file for the transparent, flat bottom 96-well microplate is selected from the plate reader library is selected (Figure 4.12b). Most plate readers come with a *Fluorescein* wavelength configuration which is chosen for the use or uranine in this experiment (Figure 4.12c). The excitation and emission wavelengths are optimally set for the spectral characteristics of uranine (483 nm and 525 nm respectively, Figure 4.10b). This ensures optimal yield of the fluorescence intensity measurement. Lastly, *Gain* (signal amplification) is set based on the highest uranine concentration measured and *Z-Position* (focal point of the laser module) is set based on the expected 96µL well volume throughout the plate (Figure 4.12d). Both *Gain* and *Z-Position* values are device dependent and are listed as exemplary values for completeness and explanation.

With the given configuration of the plate reader, a calibration plate was created precisely following the protocol. The results of the calibration plate act as reference - the source of truth - to which participants individual plate measurements are compared in order to identify pipetting errors. The average of three measurements of this calibration plate is used to calculate the calibration values per well. After that, the average of all wells with the same compositions was calculated to form the sector-averaged calibration measurement seen in Figure 4.13a.

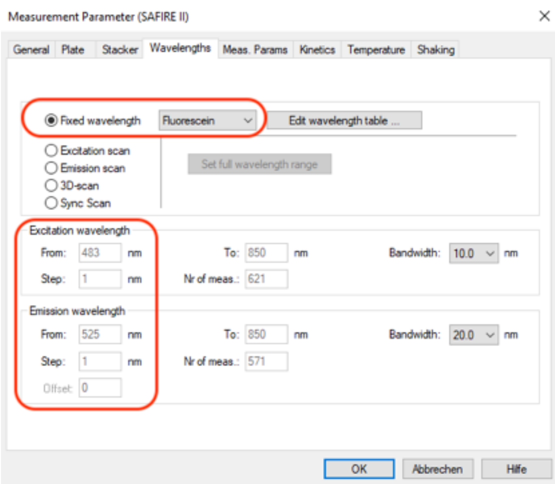
Through comparison of the participant's plate measurement (Figure 4.13b) to the sector-averaged calibration measurement, pipetting errors are easily identified by comparing the RFU values per well. To aid in the identification of pipetting errors,



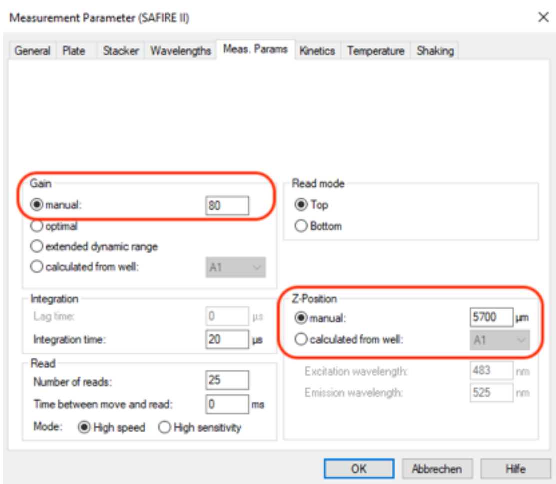
(a) General settings: Select *Fluorescence Intensity* mode.



(b) Plate settings: Plate definition files for all common microplates come with the device; GRE96tf: GRE (manufacturer), 96 (plate format) t (transparent) f (flat bottom).



(c) Wavelength settings: *Fluorescein* preset configures the device automatically for the anlysis of uranine. Excitation wavelength: 483 nm, emission wavelength: 525 nm



(d) Measurement parameters: *Gain* is set manually based on the highest uranine concentration, *Z-Position* is also set manually on 96µL well volume.

Figure 4.12. Software configuration of the Tecan Safire II plate reader used.

an Excel formula is created which show the percentile deviation between the calibration and the participants trial measurements. The general threshold for labeling possible errors was set to the deviation of above 15% to the calibration measurement (Figure 4.13c). Automatically labeling the wells of possible mistakes within the Excel sheet gave a sound basis for visual investigation of the physical plate.

Through visual investigation of the plate, possible error can be verified as participant errors or dismissed as non-relevant or uncontrolled deviations. Reasons for dismissing deviation above 15% as non-relevant include contaminated wells (fine debris can scatter light), non-planar liquid surface due to disturbed surface tension (for optimal measurements the surface tension needs to be plane for the laser to pass through). As this experiment is mainly focused on identifying pipetting errors that are due to misplacement (wrong well location), double executing or missing a protocol step entirely, the automatic labeling of deviations paired with the visual investigation of well volume is deemed precise enough to determine pipetting errors.

As described in subsection 4.4.5, the nature of the volume fractures control the minimal volume deviation per well to be 24µL. Either missing or added to much, both are visible to the eye when visually inspecting wells labeled as possible errors. All higher fractures - 48, 72 and 96µL - are even more pronounced.

When an error is confirmed via visual inspection, based on the well composition and the nature of the error (lower or higher volume then expected) the concrete protocol step in which the error occured can be identified. For example, looking at plate measurement in Figure 4.13b, three wells are marked as possible errors. E1 and F4 having half of the expected RFU value, H8 showing almost double the expected RFU value. Through inspection of the well composition in Figure 4.9a it can be seen that all wells are 'two-step transfer' type wells using 48µL of the same uranine concetration from different micro tubes to fill the well. E1 and F4 having lower RFU readings suggest the participant missed one of the transfers while H8 having higher RFU values suggests the participant double executed one of the transfers. By analysing the protocol's step instructions, it can be determined at which step the participants failed, giving insights to when and where the error happened.

The use of the Tecan Safire 2 plate reader with the described configuration in combination with the uranine solutions and well compositions dictated by the experiment protocol enable the analysis pipeline of this study to reliably measure where pipetting errors occur during the participants trial. This method allows for an objective assessment of the pipetting task's execution, linking the number of pipetting errors to the efficiency and precision of each assistance method used during the task. The microplate reader's ability to analyze multiple samples rapidly makes it an ideal tool for evaluating the complex plate layout designed for this experiment.

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	54302	54302	28037	28037	13922	13922	40833	40833	21129	21129	10476	10476
B	54302	54302	28037	28037	13922	13922	40833	40833	21129	21129	10476	10476
C	54302	54302	28037	28037	13922	13922	40833	40833	21129	21129	10476	10476
D	57	7177	57	7177	57	57	7177	7177	7177	57	57	57
E	54302	57	7177	57	28037	57	7177	7177	7177	57	40833	21129
F	57	54302	57	28037	57	57	54302	54302	54302	57	40833	21129
G	54302	57	13922	57	28037	57	10476	10476	10476	57	40833	21129
H	57	13922	57	13922	57	57	28037	28037	28037	57	57	57

(a) Sector-averaged fluorescent unit measurements from the initial calibration plate used to compare participants readings against, the 'source of truth'.

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	52195	51237	26660	26065	13308	13455	37765	39261	20144	20306	10076	10099
B	51674	51757	27007	26943	13725	13750	39080	39201	20192	20438	10724	10217
C	51611	51621	26489	26578	13734	13581	39964	39022	20453	20119	10151	9760
D	57	7011	55	6968	53	54	6980	7058	7014	55	54	57
E	19243	58	6873	55	27060	55	6876	6934	7014	55	38952	20596
F	57	51609	54	9870	52	56	51280	52142	52500	54	38516	19795
G	51735	55	13350	55	26970	52	10402	10112	10101	53	39456	20151
H	54	13376	52	13406	55	53	26712	40811	27204	50	52	53

(b) Exemplary post trial reading of the participants plate. Marked in red are significant value deviations that indicate errors made during the task.

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	3%	5%	4%	7%	4%	3%	7%	3%	4%	4%	3%	4%
B	4%	4%	3%	4%	1%	1%	4%	4%	4%	3%	2%	2%
C	4%	4%	5%	5%	1%	2%	2%	4%	3%	4%	3%	6%
D	2%	2%	7%	2%	9%	5%	2%	1%	2%	5%	7%	2%
E	65%	4%	4%	4%	3%	4%	4%	3%	2%	7%	4%	2%
F	2%	5%	7%	64%	9%	5%	5%	3%	3%	5%	5%	6%
G	4%	2%	4%	5%	3%	9%	0%	3%	3%	9%	3%	4%
H	7%	3%	11%	3%	4%	9%	4%	46%	2%	14%	12%	9%

(c) Excel formula calculating the percentile deviation between calibration (4.13a) and trial run (4.13b).

Figure 4.13. Cropped screenshots of Tecan Safire 2 plate measurement results in Microsoft Excel.

4.4.7 Workbench Setup

As shown in Figure 4.14, the workbench included a laptop for participants to complete pre- and post-task questionnaires. The task area, located to the right of the laptop, featured controlled lighting provided by an LED video light to ensure even illumination (Figure 4.14b). During the pipetting task, the area was equipped with only the necessary items: an electronic pipette, a 96-well microplate, a bin for used pipette tips, and the assigned assistance method (printed paper protocol, iPad, or HoloLens 2). The method-specific organisation of the task area is shown in Figure 4.1.

Participants were seated in adjustable office chairs, allowing them to set their preferred seating position. The experiment supervisor was present in the room during the questionnaire completion and task execution but remained out of the participants' sight to maintain an uninterrupted environment.

(a) Workbench with laptop and monitors for pre- and post-task questionnaires.

(b) Task execution area setup to control ambient lighting.

Figure 4.14. Workbench setup with laptop for questionnaires on the left and space for pipetting on the right.

4.5 Procedure

The procedure of the experiment encompassed several stages, beginning with pre-task questionnaires to gather participants' demographic data and prior experiences regarding pipetting, mobile devices, and AR/VR technologies. Following the task execution, where participants employed one of the assistance methods, post-task evaluations were conducted using the TLX, SUS, and AttrakDiff questionnaire to assess subjective workload, usability, and user experience, respectively. This section concludes with details of the debriefing process and the semi-structured qualitative interviews conducted to gather in-depth participant feedback. An overview of the experiments procedure is depicted in flow-chart form in Figure 4.15.

4.5.1 Pre-Task Questionnaires

The pre-task questionnaire was designed to capture a comprehensive profile of each participant. The questions were grouped into several categories to streamline the process and ensure a thorough understanding of each participant's background and experience:

1. General and Demographic Information
    - Gender (male, female, diverse)
    - Age
    - Years of working experience
    - Handedness (left, right, ambidextrous)
    - Colorblindness (yes or no, if yes specify type)
2. Pipetting Experience
    - Type of pipettes used (single-channel, multi-channel; electronic, mechanical)
    - Frequency of pipetting activity (days per week)
    - Number of pipetting sessions per day
    - Average session length
    - Perceived complexity of typical pipetting tasks (ranking based on provdied examples)
    - Frequency of using 96-well microplates
3. Experience with Mobile Devices
    - Daily usage of smartphones and tablets (hours per day)
    - Self-assessed competency with smartphones and tablets (5-point Likert item, very insecure to very secure)
4. Experience with HMDs
    - Previous experience with HMDs
    - (If so) Weekly usage of HMDs

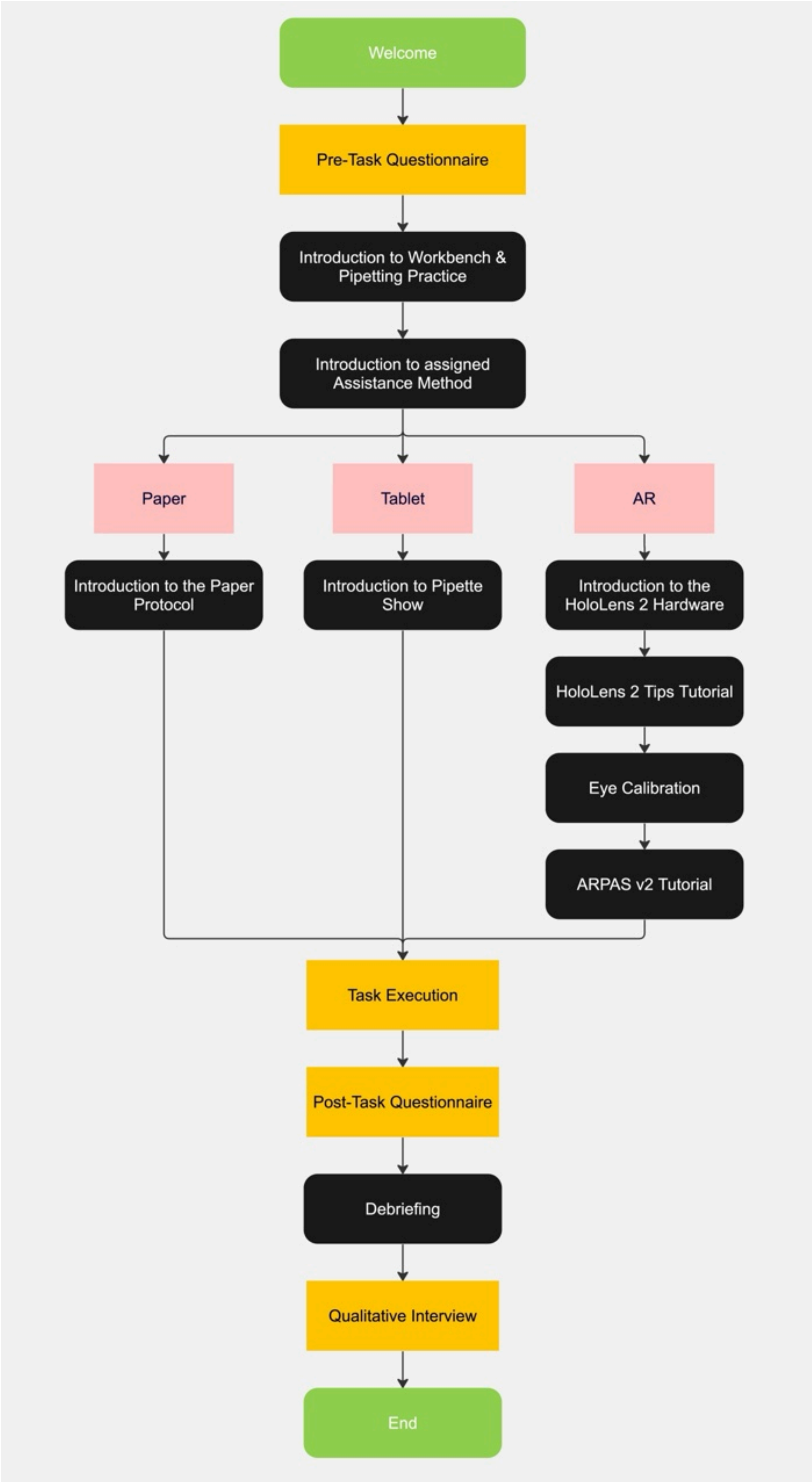


Figure 4.15. Graph visualisation of the experiments procedure flow.

- Context of HMD usage (entertainment, gaming, etc.)
- Self-rated competency with HMDs (5-point Likert item, very insecure to very secure)

The questionnaire was crafted to not only collect basic demographic data but also to gauge the participants' familiarity and comfort level with the tools relevant to the experiment. This information was critical for contextualising the study's findings and for understanding the potential impact of each participant's background on their performance and preference for different pipetting assistance methods.

4.5.2 Task Execution

Participants began with an orientation on the electronic pipette, learning essential functions such as liquid aspiration and dispensation, tip ejection, and volume preset adjustment (Figure 4.5).

They were then introduced to the microtube rack with various liquids for the task (Figure 4.8), labeled with abbreviations indicating different substances. Participants were instructed to handle one tube at a time from the rack to prevent contamination and to change pipette tips when switching liquids.

Following a practice session to ensure comfort with the equipment, participants were briefed on their specific assistance method. The *Paper* method involved reviewing printed instructions for the pipetting task (Figure 4.1a & Figure 4.1d), while *Tablet* users explored the Pipette Show application's interface (Figure 4.1b & Figure 4.1e). Those using the *AR* method via HoloLens 2 received comprehensive training, including a Microsoft Tips App tutorial and a demonstration of the ARPAS v2 application, concluding with participants indicating their readiness.

Participants were instructed to prioritise accuracy over speed in reflection of standard laboratory practice. This approach ensured that all participants had a consistent understanding of the task's requirements and controlling possible speed-accuracy trade-offs. Participants were responsible for initiating and signaling the completion of their trial autonomously. The primary task was to accurately follow the pre-defined pipetting protocol detailed in subsection 4.4.5, using the provided liquids (subsection 4.4.4) and electronic pipette to transfer substances onto the microplate (subsection 4.4.3) as specified in the protocol steps.

4.5.3 Post-Task Questionnaires

The post-task questionnaire, administered immediately after the experimental task, was structured to evaluate participants' reactions to the experiment and the assistance methods used. It included a comparison of task complexity, where participants rated the complexity of the experimental task against their regular work related tasks as a 5-point Likert item, ranging from 'a lot easier' to 'a lot more complex'. The unweighted TLX was incorporated to assess the perceived mental load across various dimensions. Additionally, the SUS evaluated the usability of the specific assistance method used during the task. Lastly, the AttrakDiff questionnaire measured the pragmatic and hedonic qualities, as well as the overall attractiveness, of the assistance method. This immediate post-task feedback is vital for gauging participants' experiences and assessing the effectiveness and user satisfaction of each assistance method, contributing significantly to the overall evaluation of the experiment.

4.5.4 Debriefing and Qualitative Interview

Upon completion of the post-task questionnaires, participants were thanked for their contributions and briefed on the study's goals. The primary objectives discussed were the analysis of task performance, specifically execution time and error count, across the different assistance methods used.

Participants were then briefly introduced to the assistance methods they had not interacted with during the experiment, providing them with a comprehensive understanding of all methods evaluated in the study.

The session proceeded to a semi-structured interview, which explored their specific work experiences and impressions of the method they used. The experiment supervisor documented participants' responses in a condensed bullet-point format, capturing the essence of their feedback. Questions initiating the interview included:

1. "What are your positive and negative impressions of the assistance method you used?"
2. "Can you describe your learning curve in using this method?"
3. "Which pipetting aids do you employ in your day-to-day work to ensure accurate pipetting?"

4. "Which challenges do you face while pipetting?"
5. "What are the primary causes of pipetting errors in your experience?"
6. "Do you have any general ideas for optimising the methods you used or the manual pipetting procedure?"

With the interviews concluded, the experiment was brought to a close, and participants were free to leave the room. This final interaction provided valuable qualitative data to complement the quantitative findings of the study. Both quantitative and qualitative results are presented in chapter 5, distilled insights from the qualitative interview are used to contextualise finding in chapter 6.

4.6 Measurements

This section details the measurements aligned with the research focus and hypotheses, including task execution time, error count, subjective workload, SUS, AttrakDiff, and qualitative interviews. Each subsection emphasises the importance of the specific measurement within the laboratory setting and in the context of this research.

4.6.1 Task Execution Time

Task Execution Time quantifies the duration (minutes, seconds) participants spent completing the experimental task and is used to compare the efficiency of the assistance methods. This metric was recorded from the moment participants verbally indicated the start of their task to their verbal confirmation of completion. The timing was manually captured by the experiment supervisor using the stopwatch function on a smartphone. While this method may not offer the precision of automated software timing, it was selected for its simplicity and to maintain a consistent approach across all assistance methods, including the paper-based method which does not allow for integrated timing. Given the expected variation in task execution time spanning several minutes, the potential for minor discrepancies introduced by manual timing is deemed acceptable and unlikely to affect the comparative analysis of the data significantly.

4.6.2 Error Count

Error Count serves as a key indicator of the assistance methods' accuracy and reliability, which is critical since pipetting inaccuracies can compromise laboratory experiments. Defined as the occurrence of non-conformative actions during task execution, errors are quantified as whole numbers (0, 1, 2, etc.). Specific errors tracked in the study were liquid misplacement into wrong wells, skipped steps, and unnecessary repetitions. As detailed in subsection 4.4.6, these were identified by inconsistencies in fluorescence intensity measured by the microplate reader and recorded in Excel sheets for each participant. Misplacement errors, indicated by differing fluorescence intensities in two wells, were counted as a single error, as they stemmed from one incorrect transfer action. The experiment supervisor performed a visual inspection of the microplates to verify flagged errors, ensuring accurate error attribution. Verified errors were then recorded in the participants' respective data sets.

4.6.3 Subjective Workload

The TLX was employed to assess the subjective workload imposed on participants by the different assistance methods. This multidimensional tool is known for its reliability and ease of use, making it a staple in user research for capturing subjective workload assessments (Hart, 2006; Hart & Staveland, 1988). The decision to measure subjective workload complements the task- and process-centric metrics of task execution time and error count, offering a user-centered perspective on the experiment. It provides insights into whether variations in assistance methods affect the participants' subjective workload, even if objective performance metrics like task execution time and error count are similar.

Administered immediately after task completion, the TLX ensured that participants' responses were reflective of their immediate experience, minimising the influence of memory decay or post-rationalisation. Each participant's subjective ratings across the six subscales of the TLX — mental demand, physical demand, temporal demand, performance, effort, and frustration — were recorded, providing a nuanced understanding of the cognitive impact of each pipetting assistance method used in the study.

4.6.4 System Usability Scale

The SUS was integrated into the study to offer a reliable estimate of the overall usability of the assistance methods employed (Brooke, 1995). Participants were asked to focus their evaluation on the method of protocol information delivery they used. This was to ensure consistency in their understanding of the term 'system' in the questionnaire, which might not naturally apply to the paper protocol.

The SUS, consisting of a 10-item questionnaire with a 5-point Likert scale, provided a global view of subjective assessments of usability. It captured aspects of usability such as effectiveness, efficiency, and satisfaction (Brooke, 1995). Given immediately after the task execution, the SUS scores from participants offered an immediate and direct reflection of their experience with the assistance method in question. This data was crucial for comparing the perceived ease of use between the innovative AR and tablet interfaces and the traditional paper protocol, contributing to a comprehensive evaluation of each method's user-friendliness.

4.6.5 Attrak Diff

The AttrakDiff questionnaire further assessed the user experience of the different assistance methods. It measured the methods' usability (pragmatic attributes), user engagement (hedonic attributes), and overall appeal (attractiveness). Evaluation through the AttrakDiff questionnaire captures subjective feedback on the design and interaction quality of the *Paper*, *Tablet*, and *AR* method, beyond conventional performance metrics.

Post-task administration of AttrakDiff provided immediate participant perceptions, revealing how the methods facilitated task completion and personal satisfaction. The insights obtained offered a nuanced understanding of each method's user experience, essential for evaluating their overall desirability in a lab environment.

4.6.6 Qualitative User Interviews

To complement the quantitative data, semi-structured qualitative interviews were conducted. Focused note-taking during interviews was employed to record significant aspects of participants' responses. This approach helped to capture insights in areas where quantitative measures might fall short. The qualitative user interviews focused on participants' subjective experiences with the assistance methods and their daily lab work. Aimed at complementing quantitative data, the interviews were designed to gather insights on user perceptions, learning experiences, and everyday lab challenges, thereby identifying areas for method improvement and revealing subtleties not captured by numerical data. The interview questions are detailed in subsection 4.5.4.

4.7 Data Analysis

This section outlines the statistical methodologies applied to analyse the experimental data. Descriptive statistics were used to summarise participant demographics and baseline characteristics, with distribution and box plots generated using JASP, an open-source statistical software, to analyse and visually represent these data.

For performance metrics — task execution time, error count, and subjective workload — inferential statistics were employed. One-way Analysis of Variance (ANOVA) was the primary method used to compare mean differences across the three groups (*Paper*, *Tablet*, and *AR*), aligning with the study's between-subjects design. Before conducting ANOVA, assumption checks were performed. The Shapiro-Wilk test assessed the normality of data distributions, while Levene's test checked for homogeneity of variances across groups.

In instances of assumption violations, alternative methods were adopted. Welch's ANOVA was used when homogeneity of variance was not met, and the Kruskal-Wallis H test was employed for significant deviations from normality. Post-hoc tests, including standard tests following ANOVA, Games-Howell post Welch's ANOVA, and Dunn's test with Bonferroni correction post Kruskal-Wallis, were conducted to identify specific differences.

Individual measurements that could be considered outliers were retained to maintain the integrity of real-world performance variations. Complete data sets negated the need for imputation methods.

The analysis of the qualitative interview data involved thematic grouping to uncover patterns in responses, frequency counting to determine the prevalence of certain views or practices, and trend identification, such as preferences for specific lab tools.

These techniques provided a deeper understanding of participants' experiences and strategies working in complex laboratory environments.

By combining these qualitative methods with the statistical analysis, the study achieves a comprehensive analysis that aligns with its objectives and hypotheses. The results from this integrated approach are detailed in chapter 5, laying the groundwork for an in-depth discussion in chapter 6, where the findings are contextualised and interpreted within the scope of the research.

5 Results

This section states the outcomes of the comprehensive evaluation, as outlined in section 4.7, and correlates them with the hypotheses detailed in section 4.1.

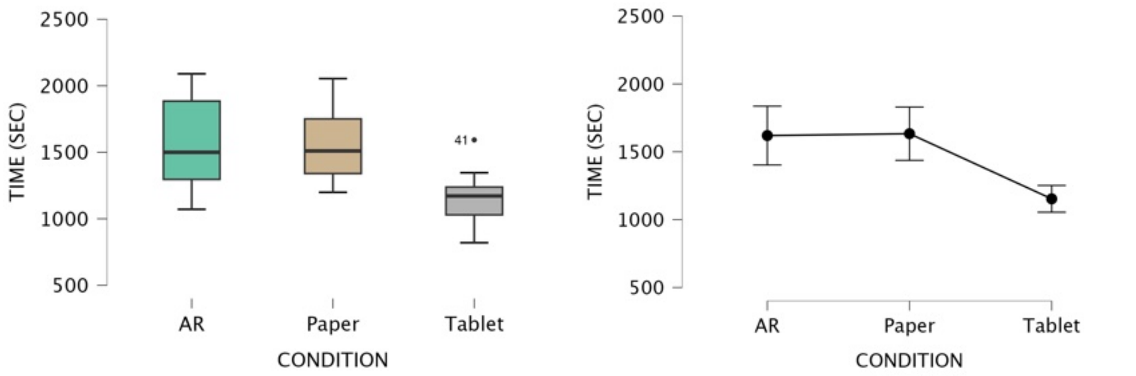
A parametric analysis demonstrated that task execution times significantly varied across methods, with the *Tablet* method proving to be the fastest. Similarly, error count analysis via non-parametric methods showed that *Tablet* and *AR* methods significantly reduced errors compared to the *Paper* method. Participants reported the lowest subjective workload with the *Tablet* method. Usability assessments via the SUS revealed higher ratings for the *Tablet* method compared to *AR*, though this analysis was not hypothesis-driven.

Descriptive analysis of AttrakDiff scores, which assess pragmatic and hedonic qualities as well as overall attractiveness, showed that the scores for pragmatic qualities were in line with the SUS results. The qualitative interviews were thematically analyzed, with frequencies highlighting key areas for potential enhancement. All results are interpreted and discussed in-depth in chapter 6.

5.1 Task Execution Time

Descriptive statistics, shown in Table 5.1, indicate that for each group, the mean execution times were 26:59 min (1619.50 s) for *AR*, 27:13 min (1633.18 s) for *Paper*, and 19:13min (1153.18 s) for *Tablet*. Variability in task execution times, as measured by the standard deviation, was greatest for the *AR* group and least for the *Tablet* group, with values of 406.4 s for *AR*, 368.5 s for *Paper*, and 185.0 s for *Tablet*.

A visual inspection of the box plots in Figure 5.1a supports these findings, with the *Tablet* condition showing a tighter distribution of times indicating a tendency towards faster task completion.

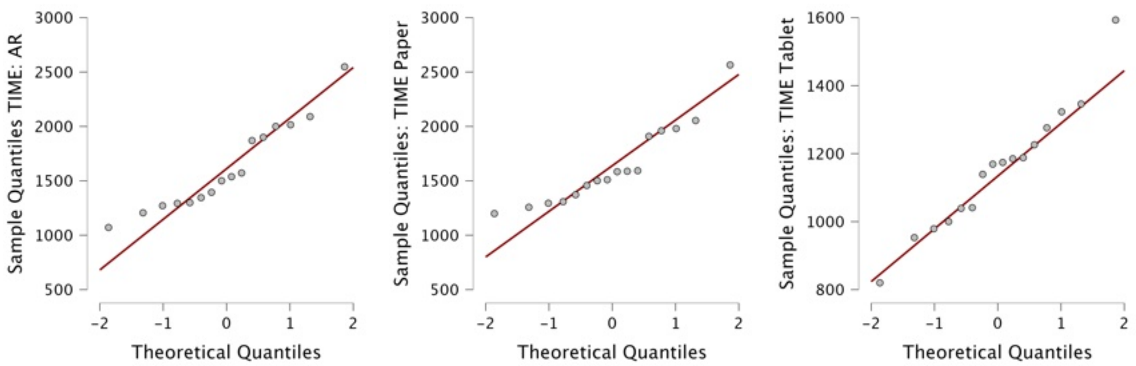


(a) Boxplot comparing task execution times illustrating the mean values and variability within each group.

(b) ANOVA plot of mean task execution times with error bars representing 95% confidence intervals, demonstrating the significant differences between methods.

Figure 5.1. Combined display of boxplot and ANOVA plot for task execution times across *AR*, *Paper*, and *Tablet* conditions, showing distribution, mean values, and 95% confidence interval.

Continuing with the differential statistical analysis, Shapiro-Wilk tests confirmed that the data for task execution time adhered to a normal distribution for all conditions (*AR*: p = .197, *Paper*: p = .066, *Tablet*: p = .806), satisfying one of the necessary assumptions for ANOVA. This is further visually supported when looking at the Quantile-Quantile plots (Q-Q plots) in Figure 5.2 showing the data closely following the theoretical distribution.



(a) Distribution of task execution data in the *AR* condition.

(b) Distribution of task execution data in the *Paper* condition.

(c) Distribution of task execution data in the *Tablet* condition.

Figure 5.2. A comparison of Q-Q plots for the *AR*, *Paper*, and *Tablet* conditions for task execution time, each demonstrating the adherence of sample quantiles to the expected theoretical quantiles under normal distribution.

However, Levene's test indicated a violation of the homogeneity of variances ($F(2, 45) =$

Table 5.1. Descriptive Statistics of Task Execution Time Results

	TIME (SEC)		
	AR	Paper	Tablet
Samples	16	16	16
Median	1519.0	1547.0	1171.5
Mean	1619.5	1633.18	1153.18
95% CI Mean Upper	1836.08	1829.56	1251.77
95% CI Mean Lower	1402.91	1436.80	1054.60
Std. Deviation	406.45	368.54	185.01
Shapiro-Wilk	0.924	0.895	0.968
P-value of Shapiro-Wilk	0.197	0.066	0.806
Minimum	1071.0	1199.0	820.0
Maximum	2548.0	2565.0	1593.0

$4.991, p = .011$), necessitating the use of Welch's ANOVA for a more robust comparison between groups. The Welch's ANOVA, shown in Table 5.2, revealed a significant effect of the used method on task execution time ($F = 16.109, p < .001$), with an observed large effect size of $\eta^2 = .322$, indicating that approximately 32.2% of the variance in task execution time was associated with the method of assistance used.

Table 5.2. ANOVA of Task Execution Time Results

Homogeneity Correction	Cases	df	F	p	η^2
Welch	CONDITION	2.000	16.109	$< .001^{***}$	0.322
	Residuals	26.322			

Note. Type III Sum of Squares
** p < .01, *** p < .001

This finding leads to the rejection of the null hypothesis H_{0_Time} stated in section 4.1, as there is evidence of a significant difference in mean task execution times between at least two assistance methods. This is visually represented in the accompanying ANOVA plot shown in Figure 5.1b. The ANOVA plot illustrates the mean task execution times for each condition along with their 95% confidence intervals, highlighting the differences between groups. The plot suggests that *Tablet* users had a faster task completion time compared to the other methods.

Based on these significant differences, post-hoc comparisons were conducted using the Games-Howell test to pinpoint the direction. The Games-Howell post-hoc test is a suitable choice for further analysis as it does not assume equal variances among groups, aligning with the Welch's ANOVA approach. This test, as shown in Table 5.3, revealed that the *Tablet* condition resulted in significantly faster task completion times compared to both the *Paper* ($p < .001$) and *AR* conditions ($p = 0.001$). No significant difference was found between the *AR* and *Paper* conditions ($p = 0.995$).

Table 5.3. Games-Howell Post Hoc Comparisons of Conditions regarding Task Execution Time

Comparison	Mean Difference	SE	t	df	p_{tukey}
AR - Paper	−13.688	137.165	−0.100	29.717	0.995
AR - Tablet	466.313	111.646	4.177	20.960	0.001**
Paper - Tablet	480.000	103.094	4.656	22.109	< .001***

** p < .01, *** p < .001
Note. Results based on uncorrected means.

Given the results, the alternative hypothesis H_{1_Time} is supported, and we can conclude that there is a statistically significant difference in the mean task execution times across the tested conditions.

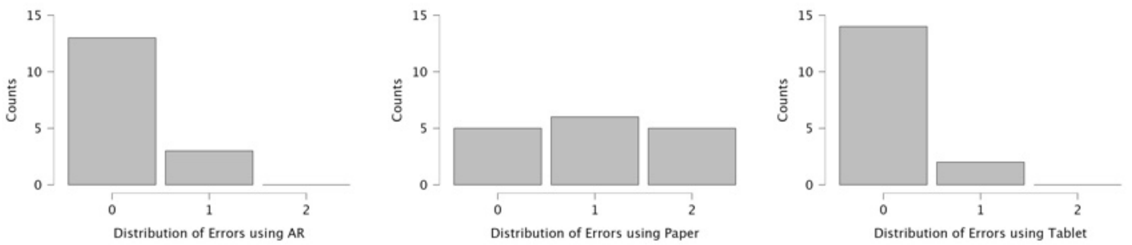
5.2 Error Count

Analysing the error count data for the *AR*, *Paper*, and *Tablet* method shown in Table 5.4, reveals significant differences. Each group, with 16 valid samples, exhibited distinct error count medians: 0 for *AR* and *Tablet*, and 1 for *Paper*. This suggests a higher precision in the *AR* and *Tablet* methods. The Shapiro-Wilk test confirmed non-normal distribution for all conditions, which is characteristic of count data with a discrete scale. The Kruskal-Wallis H test, suitable for such data, indicated significant differences across the conditions ($H = 15.375, df = 2, p < .001$), justifying the use of Dunn's post hoc test with Bonferroni correction for pairwise comparisons. Dunn's test uncovered significant differences between the *Paper* and both *AR* ($p = .004$) and *Tablet* ($p < .001$) groups, indicating fewer errors in the *AR* and *Tablet* conditions compared to *Paper* (Table 5.5). Figure 5.3 visually presents the differences in error counts based on the applied assistance method. These results lead to the rejection of the null hypothesis H_{0_Error}.

The estimated effect size η^2_H (Figure 5.4) based on the Kruskal-Wallis H statistic is .297, which - although notably large - should be interpreted cautiously due to

Table 5.4. Descriptive Statistics of Error Count Results

	ERROR COUNT		
	AR	Paper	Tablet
Samples	16	16	16
Median	0	1	0
Mean	0.188	1.000	0.125
95% CI Mean Upper	0.375	1.375	0.313
95% CI Mean Lower	0.000	0.625	0.000
Std. Deviation	0.403	0.816	0.342
Shapiro-Wilk	0.484	0.812	0.398
P-value of Shapiro-Wilk	$< .001$	0.004	$< .001$
Minimum	0	0	0
Maximum	1	2	1
Sum	3	16	2



(a) Distribution plot for error count using the *AR* method.
(b) Distribution plot for error count using the *Paper* method.
(c) Distribution plot for error count using the *Tablet* method.

Figure 5.3. A comparison of distributions plots for the *AR*, *Paper*, and *Tablet* conditions error count data.

the somewhat reduced power available to the Kruskal-Wallis H test compared to its parametric counterpart (Cohen, 2008).

These results indicate that the method of assistance significantly impacts error count, with both the *Tablet* and *AR* method showing a distinct advantage. These findings, aligned with the H_{1_Error} stated in section 4.1, suggest a meaningful influence of assistance methods on performance accuracy.

5.3 Subjective Workload

The unweighted (raw) TLX scores were collected to evaluate the subjective workload experienced by participants using each of the three assistance methods: *Paper*,

Table 5.5. Dunn's Post Hoc Comparisons of Conditions regarding Error Count

Comparison	z	W_i	W_j	p	p_{bonf}
AR - Paper	−3.221	20.531	33.781	0.001**	0.004**
AR - Tablet	0.327	20.531	19.188	0.744	1.000
Paper - Tablet	3.547	33.781	19.188	< .001***	0.001**

** p < .01, *** p < .001

$$\eta_H^2 = \frac{H - k + 1}{N - k}$$

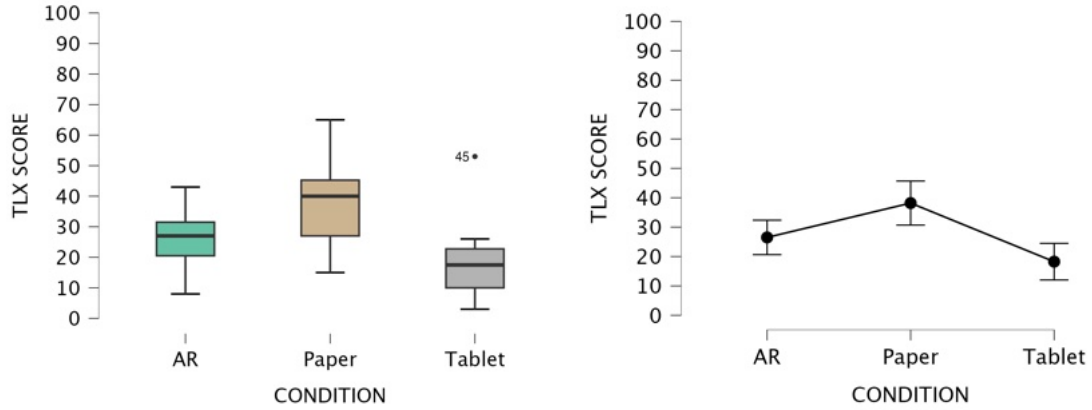
Figure 5.4. Formula for estimated overall effect size η_H^2 based on the Kruskal-Wallis H statistic (Cohen, 2008).

Tablet and *AR*. The descriptive statistics for the TLX scores are presented in Table 5.6, with the mean TLX score being the lowest for the *Tablet* group (18.25), indicating a perceived lower workload. The *AR* group reported a mean TLX score of 26.50, and the *Paper* group's mean score was higher at 38.18.

Table 5.6. Descriptive Statistics of raw TLX Scores

	TLX		
	AR	Paper	Tablet
Samples	16	16	16
Median	27.000	40.000	17.500
Mean	26.500	38.188	18.250
95% CI Mean Upper	32.357	45.681	24.475
95% CI Mean Lower	20.643	30.694	12.025
Std. Deviation	10.991	14.063	11.682
Shapiro-Wilk	0.946	0.959	0.856
P-value of Shapiro-Wilk	0.436	0.647	0.017
Minimum	8.000	15.000	3.000
Maximum	43.000	65.000	53.000

The boxplot in Figure 5.5a visualises the distribution of TLX scores for each condition, clearly showing a spread of scores with an outlier present in the *Tablet* condition. This outlier, participant ID 45, has a significant impact on the distribution and suggests the presence of a participant who either experienced an exceptionally high subjective workload compared to others in the same group or possibly misunderstood the TLX questionnaire ranking scale.

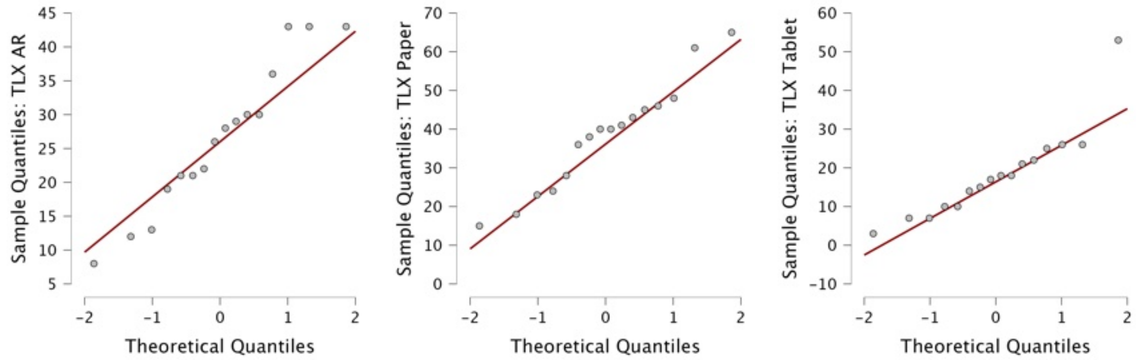


(a) Boxplot comparing raw TLX results illustrating the mean values and variability within each group.

(b) ANOVA plot of mean raw TLX results with error bars representing 95% confidence intervals, demonstrating the significant differences between methods.

Figure 5.5. Combined display of boxplot and ANOVA plot for raw TLX results across *AR*, *Paper*, and *Tablet* conditions, showing distribution, mean values, and 95% confidence interval.

Q-Q plots depicted in Figure 5.6 further illustrate the distribution of scores within each group. While the *AR* and *Paper* groups' scores closely followed the line of theoretical quantiles, indicating a normal distribution, the *Tablet* group's scores deviated from normality, largely due to the aforementioned outlier.



(a) Distribution of raw TLX data in the *AR* condition.

(b) Distribution of raw TLX data in the *Paper* condition.

(c) Distribution of raw TLX data in the *Tablet* condition.

Figure 5.6. A comparison of Q-Q plots for the *AR*, *Paper*, and *Tablet* conditions for raw TLX results.

The assumption check for equality of variances, conducted via Levene's test, was not significant ($F = 0.509, p = 0.605$), indicating that the variances across groups were statistically similar and homogeneity of variance was not violated.

Given the skewness caused by the outlier in the *Tablet* group, a Kruskal-Wallis H test, a non-parametric alternative to ANOVA, was performed. This test is robust to non-normal distributions and, while given in this case, does not require the assumption of homogeneity of variances. The Kruskal-Wallis H test indicated a significant difference in TLX scores across conditions ($H = 15.152, p < .001, \eta^2_H = .292$), suggesting that the method of assistance used has a significant effect on subjective workload. With these results clearly indicating evidence of significant differences in subjective workload between at least two assistance methods, the null hypothesis H_{0_Load} is rejected.

Following the Kruskal-Wallis H test, Dunn's Post Hoc test with Bonferroni correction was used to conduct pairwise comparisons between conditions. This test is appropriate when the sample size is small and the data distribution is not normal. Dunn's test revealed significant differences between the *Paper* and *Tablet* conditions, with the *Tablet* method associated with a lower subjective workload ($p < .001$, Table 5.7).

Table 5.7. Dunn's Post Hoc Comparisons of Conditions regarding raw TLX Scores

Comparison	z	W_i	W_j	p	p_{bonf}
AR - Paper	−1.908	24.625	34.063	0.056	0.169
AR - Tablet	1.984	24.625	14.813	0.047*	0.142
Paper - Tablet	3.892	34.063	14.813	< .001***	< .001***

* p < .05, *** p < .001

In addition to the non-parametric analysis, a one-way ANOVA was conducted to examine the potential for a parametric analysis, assuming normal distribution and larger sample sizes in future studies. The ANOVA results, shown in Table 5.8, corroborated the findings from the Kruskal-Wallis H test, with significant differences in TLX scores across conditions ($F = 10.587, p < .001$). The effect size η^2 was 0.320, suggesting that approximately 32% of the variance in TLX scores could be explained by the assistance method used. Although the effect size indicated by the ANOVA is substantial and is aligned with the estimated effect size η^2_H, it warrants cautious interpretation due to the analysis being conducted in the context of a violated assumption of normal distribution. The ANOVA plot, visually presenting the differences in experienced subjective workload, can be found in Figure 5.5b.

Tukey's Post Hoc test was then performed for pairwise comparisons. This test in-

Table 5.8. ANOVA of raw TLX Scores

Cases	df	F	p	η^2
CONDITION	2	10.587	< .001	0.320
Residuals	45			

Note. Type III Sum of Squares

Table 5.9. Tukey Post Hoc Comparisons of Conditions regarding raw TLX Scores

		Mean Difference	SE	t	p_{tukey}
AR	Paper	−11.687	4.354	−2.684	0.027*
	Tablet	8.250	4.354	1.895	0.152
Paper	Tablet	19.937	4.354	4.579	< .001***

Note. P-value adjusted for comparing a family of 3
* p < .05, *** p < .001

dicated that participants using the *Paper* method reported a significantly higher subjective workload compared to those using the *AR* method ($p = 0.027$) and the *Tablet* method ($p < .001$). The Tukey's Post Hoc test uncovered significant differences not observed in the non-parametric Dunn's test, specifically between the *Paper* and *AR* methods, suggesting the *Paper* method resulted in a higher subjective workload than *AR*, a finding unique to the parametric analysis.

The integration of both non-parametric and parametric analyses provides a comprehensive understanding of the subjective workload data. These findings strongly support the hypothesis H_{1_Error}, indicating that the assistance method significantly impacts subjective workload during pipetting tasks. The non-parametric analysis distinctly highlighted the *Tablet* method as reducing subjective workload compared to *Paper*, while the parametric analysis, despite its limitations due to the outlier, suggests that *AR* may also offer subjective workload advantages over *Paper*.

5.4 System Usability Scale

The SUS provides a quantitative measure of the perceived usability of the assistance methods used during the experiment. The mean SUS scores for each condition were 79.063 for *AR*, 85.00 for *Paper*, and the highest at 90.938 for *Tablet*, suggesting a trend towards the *Tablet* method being perceived as more usable. The descriptive statistics of the SUS data are detailed in Table 5.10.

Table 5.10. Descriptive Statistics of SUS Results

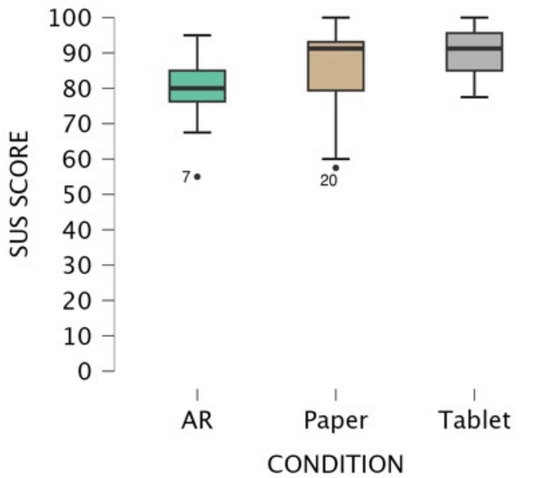
	SUS SCORE		
	AR	Paper	Tablet
Samples	16	16	16
Median	80.000	91.250	91.250
Mean	79.063	85.000	90.938
95% CI Mean Upper	84.344	92.032	94.638
95% CI Mean Lower	73.781	77.968	87.237
Std. Deviation	9.911	13.197	6.945
Shapiro-Wilk	0.945	0.857	0.943
P-value of Shapiro-Wilk	0.409	0.017	0.383
Minimum	55.000	57.500	77.500
Maximum	95.000	100.000	100.000

The boxplots (Figure 5.7a) illustrate the spread of SUS scores, revealing a broad range for the *Paper* condition. Both *Paper* and *AR* include a notable low outlier that suggests a particularly low perceived usability in one instance each. In contrast, the *Tablet* condition displays a compact interquartile range with consistently higher median and mean scores, indicating a uniform and favorable usability perception among participants.

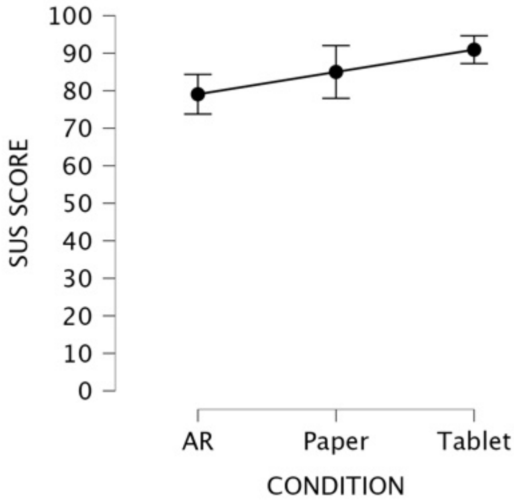
Given the violation of the normal distribution assumption for the SUS scores, as indicated by the Shapiro-Wilk test for the *Paper* group ($p = 0.017$), a non-parametric approach was adopted for further analysis. Q-Q plots showing the distribution of SUS scores for each method are provided in Figure 5.8. The Kruskal-Wallis H test was performed to determine if there were statistically significant differences in SUS scores across the three conditions. The results of the Kruskal-Wallis H test were significant ($H(2) = 10.719, p = 0.005$, $\eta^2_H = .193$), suggesting that at least one of the groups differed significantly in terms of perceived usability.

Post-hoc analysis using Dunn's test with Bonferroni correction (Table 5.11) provided pairwise comparisons. This analysis indicated significant differences between the *AR* and *Tablet* conditions ($p = 0.004$) with no significant difference noted comparing the *Paper* method to both *AR* and *Tablet* methods. These results underscore the higher usability rating for the *Tablet* compared to the *AR* method.

The findings from the SUS scores analysis suggest that, among the three conditions tested, participants rated the *Tablet* method as having a higher usability. The results, while requiring cautious interpretation due to the non-parametric nature of

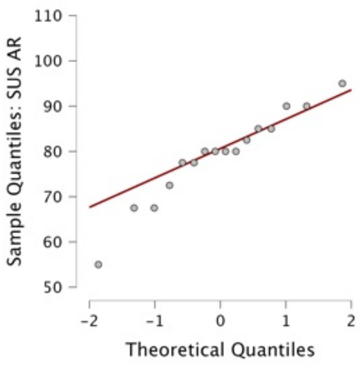


(a) Boxplot comparing SUS results illustrating the mean values and variability within each group.

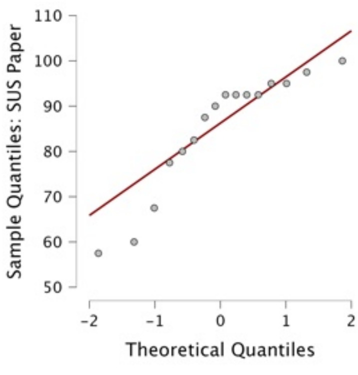


(b) ANOVA plot of mean SUS results with error bars representing 95% confidence intervals, demonstrating the significant differences between methods.

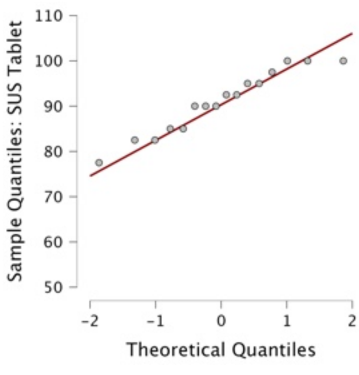
Figure 5.7. Combined display of boxplot and ANOVA plot for SUS results across *AR*, *Paper*, and *Tablet* conditions, showing distribution, mean values, and 95% confidence interval.



(a) Distribution of SUS scores in the *AR* method.



(b) Distribution of SUS scores in the *Paper* method.



(c) Distribution of SUS scores in the *Tablet* method.

Figure 5.8. A comparison of Q-Q plots for the *AR*, *Paper*, and *Tablet* conditions for SUS results.

Table 5.11. Dunn's Post Hoc Comparisons of Conditions regarding SUS Results

Comparison	z	W_i	W_j	p	p_{bonf}
AR - Paper	−2.029	15.844	25.844	0.042*	0.127
AR - Tablet	−3.240	15.844	31.813	0.001**	0.004**
Paper - Tablet	−1.211	25.844	31.813	0.226	0.678

* $p < .05$, ** $p < .01$

the analysis, offer an additional perspective on the performance of the assistance methods beyond error count and task execution time. It is notable that all ratings tightly group between 79 and 90 points, well below the cutoff value of 68, which is considered average (Brooke, 1995). Looking at the adjective scale for SUS scores developed by Bangor et al., 2009 all methods rank between 'Excellent' and 'Best Imaginable'.

5.5 AttrakDiff

The AttrakDiff questionnaire assessed participants' perceptions of the three pipetting assistance methods. Descriptive statistics for the pragmatic quality (PQ), hedonic quality (HQ), and attractiveness (ATT) scores are contained in Table 5.12. The overall HQ rating is the average of its two equal subcomponents: stimulation (HQ-S) and identification (HQ-I).

For PQ, the *Tablet* condition reported the highest mean score ($M = 5.446$), followed by the *Paper* ($M = 5.348$) and the *AR* ($M = 5.159$) conditions. Notably, the *Paper* condition displayed a higher standard deviation, indicating greater variability in PQ scores. The Shapiro-Wilk test identified significant deviations from normality for the *Tablet* condition's scores across all AttrakDiff dimensions.

In terms of HQ, the *AR* condition led with a mean score ($M = 5.648$) that suggests a positive experience in stimulation and identification potential. This was in contrast to the lower mean scores of the *Paper* ($M = 3.183$) and *Tablet* ($M = 4.942$) conditions.

ATT scores were also reflective of these trends, with the *AR* condition being rated the most appealing ($M = 6.017$). The *Tablet* condition was rated slightly lower

Table 5.12. Descriptive Statistics of AttrakDiff Results

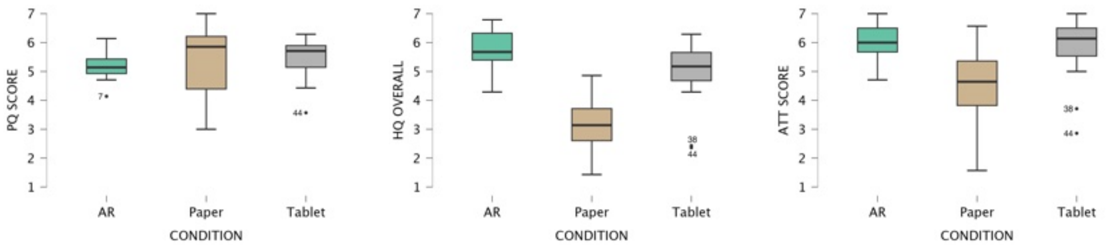
		Samples	Mean	STD
PQ	AR	16	5.159	0.473
	Paper	16	5.348	1.232
	Tablet	16	5.446	0.740
HQ	AR	16	5.648	0.758
	Paper	16	3.183	0.885
	Tablet	16	4.942	1.140
HQ-I	AR	16	5.304	0.908
	Paper	16	4.214	1.261
	Tablet	16	4.982	1.244
HQ-S	AR	16	5.991	0.726
	Paper	16	2.151	0.919
	Tablet	16	4.902	1.123
ATT	AR	16	6.017	0.624
	Paper	16	4.509	1.337
	Tablet	16	5.821	1.160

($M = 5.821$), with the *Paper* condition having the lowest mean attractiveness score ($M = 4.509$).

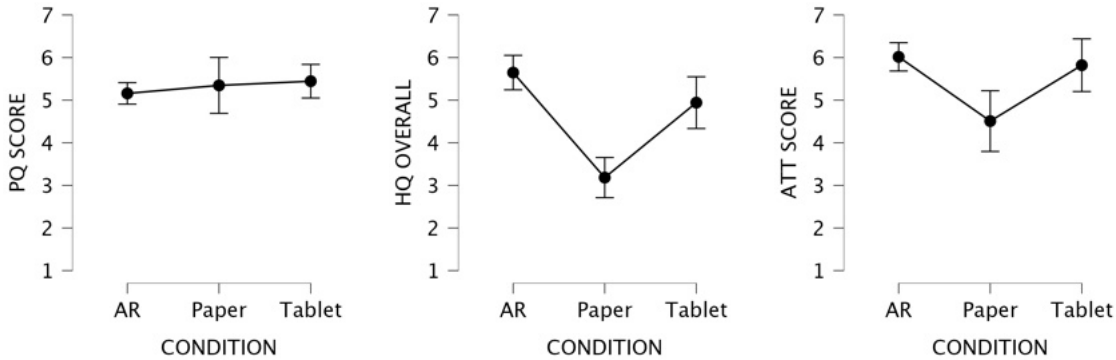
Boxplots for PQ, HQ, and ATT scores illustrate the data distribution and outliers for each condition and are accompanied by the mean scores for each assistance method, depicted alongside 95% confidence intervals (Figure 5.9). For PQ scores (Figure 5.9a), the *AR* condition exhibits less variability. The HQ scores (Figure 5.9b) reveal a higher median for the *AR* condition, and similarly, the ATT scores (Figure 5.9c) underscore the *AR* condition's elevated appeal, with noteworthy outliers in the *Tablet* condition.

Within the HQ dimension of the AttrakDiff questionnaire, the identification (HQ-I) and stimulation (HQ-S) components highlighted differing perceptions among the assistance methods. The HQ-I scores, reflecting the system's identity affirmation, were highest for the *AR* condition ($M = 5.304$), indicating a strong user connection. Similarly, the HQ-S scores, which evaluate the system's ability to engage the user, also favored the *AR* method ($M = 5.991$), but with the *Paper* condition registering significantly lower scores ($M = 2.151$), suggesting a less engaging experience.

The box plots for these components (Figure 5.10a, Figure 5.10b) further affirm these

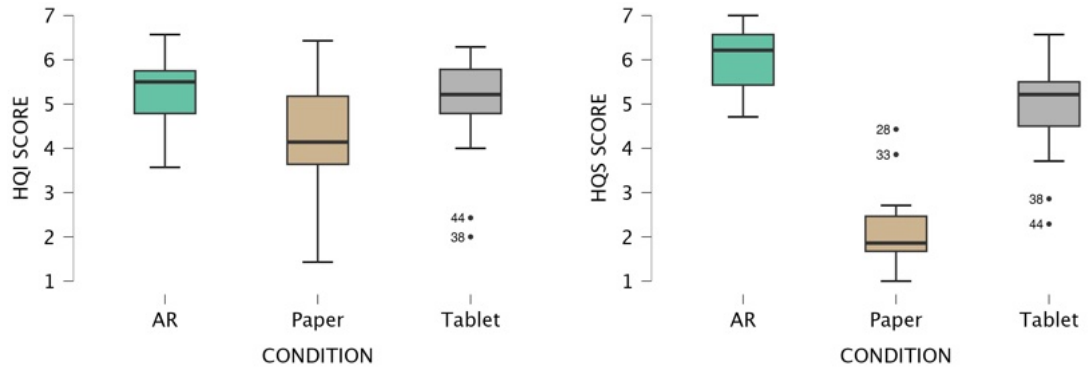


(a) Pragmatic Quality Boxplot **(b)** Hedonic Quality (I+S) Boxplot **(c)** Attractiveness Boxplot



(d) Pragmatic Quality Mean Scores with 95% confidence intervals **(e)** Hedonic Quality (I+S) Mean Scores with 95% confidence intervals **(f)** Attractiveness Mean Scores with 95% confidence intervals

Figure 5.9. Boxplots and mean scores with 95% confidence intervals of PQ, HQ(I+S) and ATT results.



(a) Hedonic Quality - Identity **(b)** Hedonic Quality - Stimulation

Figure 5.10. Comparison of the sub-scales within hedonic quality: stimulation and identity.

trends. Notably, the *Paper* condition's HQ-I scores exhibit a wide interquartile range ($STD = 1.261$), signaling a diverse range of user experiences concerning identification with the method. Meanwhile, the interquartile range for HQ-S scores across all methods is narrower, indicating a more uniform perception of stimulation among participants.

The AttrakDiff evaluations, inherently subjective due to the between-subjects design, reflect participants' experiences with their respective pipetting assistance methods. These self-assessments stand in contrast to the objective metrics such as task execution time and error count, which allow for a direct inter-condition comparison. Hence, while the AttrakDiff scores yield valuable insights into individual user experience, they should not be construed as a comprehensive comparative analysis.

5.6 Qualitative User Interviews

In-depth qualitative interviews with participants revealed prominent themes regarding the pipetting assistance methods used, as detailed in Table 5.13 and Table 5.14. For the *AR* group, the most noted positive aspect was the mental relief provided by the system (11 mentions), while the primary concern involved the physical strain associated with wearing the HMD (11 mentions). *Paper* users found the straightforward protocol easy to follow (10 mentions) but also noted that longer and more complex protocols could lead to mistakes (6 mentions). *Tablet* users frequently highlighted the ease of learning and routine establishment with the method (12 mentions) but also mentioned a tendency to drift into automatism (8 mentions), which could potentially lead to errors.

Participants commonly employed personal aids, such as colored markings on microplates (25 mentions) and visual layout diagrams (10 mentions). They reported three primary challenges: the physical demands of operating the pipette (20 mentions), localising well positions (18 mentions) and sustaining concentration over long periods of time (17 mentions). To improve efficiency, several participants, particularly those using the *AR* and *Tablet* method, recommended the adoption of smart pipettes (17 mentions). These devices could potentially streamline the workflow by controlling the pipetting protocol and automating volume adjustments

This qualitative data, representative of the participants' subjective experiences and perspectives, provides a complementary layer of understanding to the quantitative results presented in this chapter and suggests areas for future improvements in pipetting assistance technology. All results are contextualised and discussed within chapter 6.

Table 5.13. Summary of Participant Feedback: Positive and Negative, by Method

AR Users	Paper Users	Tablet Users
Positive Remarks		
Provided mental relief (11) With little practice easy to operate (8) Information proximity to microplate (3)	Linear structure protocol easy to understand (10)	Easy to learn, quick routine (12) Provided mental relief (10) Good and bright visualisation (6)
Negative Remarks		
Physical strain of HMD (11) Jittery tracking is unsettling (8) Limited FoV (8) No haptic feedback (6) Handtracking sometimes unreliable (4)	Long and complex protocols are error prone (6) Switch between paper and microplate tiring (3)	Drift off into automatism (8) Only limited space for more plates (2)

Table 5.14. Key Themes from Participant Responses by Interview Question

"Which pipetting aids do you employ personally?"
Colored markings on the microplate (permanent marker) (25)
Visual representation of the plate layout (printed or digital) (10)
Workplace organisation (micro tube rack, pipette tip box) (7)
Colored background and controlled lighting to increase contrast (3)
"Which challenges do you face while pipetting?"
Pipette operation (physical strain, small volumes, challenging liquids) (20)
Finding and keeping track of well positions (18)
Upholding concentration (monotonous or complex protocols, disruptions) (17)
Workplace organisation and sample preparation (5)
"What are primary causes of pipetting errors?"
Monotony (lost position, forgot tip change, redoing steps) (14)
Disruptions by the workplace or colleagues (13)
Concentration degradation in complex protocols (12)
"Any suggestions for improvement?"
AR & Tablet:
Use of smart pipettes for protocol control and automatic volume change (17)
Improved interaction (audio and haptic feedback for button presses) (8)
Visualisation of step history (3)
AR: Smaller, more comfortable and better HMDs (FoV, resolution, weight) (5)
Tablet: Inclined surface for better visibility (3)

6 Discussion

The Discussion chapter delves into the interpretation of the results from chapter 5, relating them to the initial hypotheses (section 4.1) and the preliminary findings of ARPAS v1 (subsection 2.5.3). It situates these findings within the broader context of existing research, evaluating the study's strengths and limitations, and concludes with recommendations for future research.

This study aimed to determine the effectiveness of AR as an assistive technology in manual pipetting tasks in laboratory settings, building upon ARPAS v1's demonstration of AR's feasibility in this domain. To offer a comprehensive evaluation, the study included Pipette Show, representing a screen-based 2D UI assistance, for comparison against the conventional paper-based method and an innovative 3D UI.

The evaluation centered on three primary metrics: task execution time, error count, and subjective workload. Task execution time and error count are critical in the context of efficiency and accuracy in industrial and research-oriented laboratories. Meanwhile, subjective workload assessment offers valuable insights into the user experience, a key consideration in HCI research. The aim was to not only enhance efficiency but also reduce user strain through an innovative assistive technology.

6.1 Interpretation of Findings

In this between-subject study of 48 life science professionals, three distinct pipetting assistance methods — *Paper*, *Tablet*, and *AR* — were evaluated. Key findings include the *Tablet* method's significant time efficiency, completing the approximately 30-minute protocol in less than 20 minutes ($Mean_{Tablet} = 19:13min$), considerably faster than the other methods ($Mean_{Paper} = 27:13min$, $Mean_{AR} = 26:59min$). The *Tablet* and *AR* methods markedly reduced errors with mean scores of 0.125 and 0.188, respectively, compared to the *Paper* method's 1.000. This improvement is also evident in the total error count: *Tablet* and *AR* groups had only two and three errors, significantly fewer than the *Paper* group's 16. Participants in the *Tablet* and *AR* group also reported lower TLX scores, suggesting reduced subjective workload.

These outcomes led to the rejection of all null hypotheses - H_{0_Time}, H_{0_Error} and H_{0_Load} formulated in section 4.1 - demonstrating differences in method effectiveness. Furthermore, these findings align with the error reduction trends observed in the smaller-scale ARPAS v1 evaluation, where the *AR* and *Paper* method were compared in a within-subject study. The next step involves interpreting these results in the context of the participant sample, work environment, and relevant literature, to discern practical and theoretical implications for real-world applications and future research.

Influence of Technological Background and Proficiency

To interpret task execution times for the *Tablet* and *AR* methods, understanding participants' technical background and proficiency is crucial. The prevalent use of mobile devices like smartphones and tablets by the participants implies general familiarity with direct manipulation of 2D UIs, characterised by touch input and tactile feedback. Since 2014, yearly worldwide smartphone sales have surpassed one billion units ("Smartphone sales worldwide 2007-2022, Statista," 2023) leading to an estimated 6.4 billion smartphones users worldwide ("Mobile network subscriptions worldwide 2016-2022, Statista," 2023), around 80% of the worlds population. Decades of HCI research and evolving design guidelines for popular mobile operation systems like Android ("Material Design," 2023) and iOS ("Apple Human Interface Guidelines," 2023) have likely contributed to high proficiency in 2D UI operation, confirmed by participants' self-assessed confidence with mobile devices (section 4.3).

In contrast, HMD technology for AR and VR, despite its growing applications in various industries and the consumer market, remains less spread in the general population. The volume of VR unit sales was less than 2% compared to smartphone sales in the same year ("VR headsets volume 2018-2022, Statista," 2023). Most participants experienced AR for the first time in this study, indicated by the polling of previous AR/VR experiences. The participants' lack of prior exposure to AR technology likely contributed to the absence of significant differences in task execution times between the *AR* and *Paper* methods. While they were adept at using a paper protocol for pipetting, the initial adjustment period to the 3D UI and HMD interactions may have temporarily offset any efficiency gains, suggesting that familiarity with the system could lead to improved performance over time.

Participants adeptly navigated the Pipette Show's UI displayed on the iPad, aligning with their mobile device experience. The ease of use, as reported in qualitative feedback (section 5.6), highlights efficient interaction facilitated by fixed button placement and distinct display boundaries. Conversely, the unfamiliar 3D UIs in HMDs posed challenges, potentially impacting *AR* group performance despite a dedicated training session before the trial. Werrlich et al., 2018 observed a similar initial learning curve of participants taking part in an AR assembly training task using a HMD. Future research could explore how extended AR training with ARPAS v2, or a smiliar system, affects pipetting task efficiency and user confidence. Increased technical proficiency may reveal a more pronounced effect of AR assistance for pipetting tasks.

The accuracy analysis of the *Tablet* and *AR* methods demonstrates that visual assistance significantly reduces errors in complex pipetting tasks. Both methods, yielding comparable error counts, proved markedly superior to the traditional *Paper* method. Users from both groups emphasised the effectiveness of well illumination on the microplate, aiding in both well location and tracking during task execution. This approach resonated with common lab practices stated by the participants, such as marking microplates for visual reference, underscoring its practicality.

Moreover, the notably lower subjective workload ratings for both *Tablet* and *AR* users further validate the advantages of these assistance methods. Despite the *Tablet* method's current edge due to user familiarity and ease of use, the *AR* method showcases potential for similar performance levels with more extensive training and acclimatisation. A future long-term study could investigate subjective workload ratings in real, standardised laboratory processes, using both AR and conventional methods. This research would provide valuable insights into the subjective workload of exerted by AR method, particularly when the novelty factor of the technology diminishes, offering a deeper understanding of AR's practical impact in laboratory settings.

Potential Benefits of AR Beyond Pipetting Assistance

AR technology can extend its utility in the laboratory setting beyond the demonstrated benefits in pipetting scenarios. Modern VST HMDs like the Meta Quest 3 (“Meta Quest 3,” 2023), with their expansive FoV (110 degrees horizontally, 96 degrees vertically) and high pixel density, offer a vast virtual screen space that far surpasses the physical limitations of tablets and laptops. With the eLabBench, Tabard et al., 2012 demonstrated the benefits of a large interactive screens for accessing and modifying experimental data and the tracking of lab utensils like tube racks. Modern VST HMDs could offer similar, or even enhanced, functionality. Capable of 2D and 3D data display, extended tracking of lab hardware through marker or object-based recognition, with the added benefit of mobility for use throughout the laboratory.

The spatial awareness capabilities of HMDs open new possibilities for room-scale interactions. These go beyond traditional individual workbench applications, allowing for novel AR interactions across larger laboratory spaces. Advancing adoption of the Internet of Things within the laboratory, to monitor and automate experimental processes (Parks et al., 2022; Poongothai et al., 2018), provides important information that could effectively be displayed in AR. These connected lab devices, potentially linked via MQTT or web-enabled APIs (Blanco-Novoa et al., 2020; Perkel, 2017), could have their statuses displayed on a Heads Up Display-like interface or in close proximity to the actual devices, identified through marker-based tracking. This approach could revolutionise interactions in the lab, streamlining processes and enhancing efficiency.

Particularly in hazardous lab environments, where safety and contamination are paramount, AR on HMDs could be used to overcome shortcomings of handheld devices, like tablets. Participants working with fume-producing substances in laminar flow cabinets envisioned the advantage of HMDs, which, unlike tablets, reduce contamination risks through touch-less interactions like hand-tracking and voice commands. This application of AR could enhance safety and increase efficiency when handling dangerous materials.

While current HMDs face challenges like limited FoV, battery life, and comfort, technological advancements suggest these issues will be addressed in the future. The potential applications of mobile AR on HMDs are extensive, extending well beyond the effective pipetting assistance demonstrated in this thesis. As the lab environment evolves with more connected devices, improved user proficiency, and novel use-cases being discovered, AR technology is poised to transform laboratory work, particularly in areas where traditional touch-based devices are less suitable.

6.2 Strengths, Limitations and Implications for Future Work

The study stands out for its comprehensive design, which included Pipette Show as an exemplar of screen-based pipetting assistance software akin to commercially available products. This inclusion facilitated rich comparisons among *AR*, *Tablet*, and *Paper* methods, effectively challenging the AR system's capabilities. The expert sample of laboratory professionals enriched the study with significant insights, particularly through qualitative interviews, enhancing the weight and relevance of the results. The real-world aligned pipetting task, involving complex inter-container transfers, highlighted practical implications for task efficiency and accuracy. Moreover, the employment of a precise fluorescent analysis pipeline marked a notable advancement in evaluating individual performance metrics in pipetting tasks, thereby contributing valuable methodologies to HCI research in the life science domain.

The study's methodological limitations were primarily rooted in its between-subject design, which featured only a single testing round. This format restricted insights into how familiarity and usability of each assistance method might evolve over time. For a nuanced understanding of task performance influenced by prolonged use, a future study could employ a repeated measures design, exploring how users adapt to these technologies and the subsequent effects on performance. Alternatively, a within-subject design over an extended period could provide a comprehensive evaluation of usability, as participants would have the opportunity to compare all methods after sufficient adaptation time. While the current sample size was sufficient, a larger participant pool could enhance the robustness of statistical findings. Additionally, the diverse expertise of participants, although beneficial for breadth, presented challenges in standardising the experimental task. Collaborations with research institutions for future studies could help in designing tasks that are more representative of specific laboratory practices, albeit potentially reducing the generalisability of results.

The use of the HoloLens 2 in this study, while generally effective, faced limitations, particularly in terms of FoV, resolution, and focus distances, as highlighted in participant interviews. Since its 2019 release, HMD technology has advanced considerably, with devices like the Meta Quest 3 and the anticipated Apple Vision Pro ("Apple Vision Pro," 2023) presenting opportunities for future evaluations. These modern VST devices could potentially overcome the limitations of the HoloLens 2, offering a trade-off analysis between smaller waveguide FoV and native passthrough versus larger FoV and higher resolution with video passthrough. Similarly to the evaluation of depth perception in OST and VST HMDs made by Adams et al., 2022, new HMD hardware should undergo a careful evaluation process with regard to the specific use case.

Future work could involve comparative studies of current-generation HMDs, assessing how different interaction modalities impact task performance and usability. This could address some participants' difficulties with hand-tracking interaction with holographic buttons. For instance, a study comparing hand-tracking, voice control, and hardware controllers using the same HMD platform and pipetting protocol could reveal whether alternate controls enable *AR* users to match the speed of *Tablet* users, especially considering the participants' challenges of actuating holographic buttons without haptic feedback.

Additionally, integrating smart Bluetooth pipettes, such as the Sartorius Picus 2 ("Sartorius Picus 2," 2023), into the *AR* and *Tablet* systems could offer insights into the effects of automated volume changes on task performance and subjective workload. A comparison between non-connected electronic pipettes and smart pipettes with automated volume adjustments might indicate whether the latter enhances speed and accuracy, albeit potentially reducing user engagement. This touches on the broader theme of technical trust and the balance between human oversight and system automation, raising the question of the optimal level of assistance. A future research idea is to assess possible automation-induced complacency, a phenomenon where users are over-trusting and over-reliant on an automated system neglecting appropriate monitoring (Parasuraman et al., 1993). This could be investigated by intentionally creating mismatches between textual instructions and visualisations in the *AR* or *Tablet* system, to see if users continue to actively monitor the system or become complacent, simply following the instructions without due diligence.

The current iteration of ARPAS v2 presents a limitation in its tracking capabilities, specifically in its inability to track the user's pipette alongside the microplate. Previous research by Hile et al., 2004 demonstrated pipette tip tracking using colored markers and a camera array for computer vision analysis, effectively turning the pipette into an interactive pointer within the application. However, this method required holding the pipette in an unnatural position for effective tracking, compromising the ergonomics of liquid handling.

Future improvements could explore a hybrid approach combining object-based with marker-based tracking, like the Vuforia Engine's Image and Model Target features, on modern HMDs. This blend could potentially offer a more ergonomic solution, integrating the pipette seamlessly as a TUI. Further enhancing this concept with the integration of smart Bluetooth pipettes could elevate the usability and functionality of AR pipetting assistant software. Such advancements would open new avenues for how AR can interact with physical tools in the laboratory, enhancing the user experience and operational efficiency in ways yet to be fully realised.

This discussion has underscored the potential of AR assistance in laboratory settings, especially in improving the precision and efficiency of complex tasks. By addressing the capabilities and limitations of AR technology, this thesis lays the foundation for future advancements. The forthcoming chapter will summarise these key findings, highlighting the impact and future implications of AR in laboratory workflows, and marking the culmination of this research into the benefits AR task assistance in laboratory environments.

7 Conclusion

This thesis evaluated the effectiveness of AR in manual pipetting tasks within laboratory environments, comparing AR assistance on the Microsoft HoloLens 2 against the tablet-based assistance, Pipette Show, and a traditional paper protocol. The study successfully demonstrated measurable differences in task execution time, error count, and subjective workload across these methods, leading to the rejection of all null hypotheses related to these key metrics.

The *Tablet* method emerged as the most efficient in terms of speed. Additionally, both the *Tablet* and *ar* methods significantly reduced errors, with the *Tablet* method also lowering subjective workload compared to the *Paper* method. These results corroborate previous findings in AR task assistance and suggest that current AR HMD technology is a viable solution for this specific context. The practical implications are clear: current AR HMD platforms are sufficiently capable of enhancing real-world pipetting scenarios, indicating broader potential for application in varied laboratory workflows. Therefore, the initial research question, "Can an AR-based assistance system be effectively used in manual liquid handling scenarios by expert personnel?" can be answered affirmatively.

The main limitations of the study include its sample size, between-subject design and single point of measurement. The single point of measurement didn't allow participants, most of whom were first-time AR users, to familiarise themselves with HMD technology. This is a significant issue, especially given that all participants were already familiar with mobile devices. Although HMD training was provided, the design did not fully account for the learning curve associated with new technology. Future research could benefit from larger, longer, and more focused studies on specific laboratory workflows or a within-subject design to gain deeper insights into the usability and efficacy of different assistance methods.

Further developments could explore the integration of connected lab devices like smart pipettes as TUI, tracking of additional work objects, and multiple input modalities to improve the functionality and user experience of the AR assistance system. Evaluating newer VST HMDs could also provide insights into the evolving landscape of AR technology in laboratory settings.

In conclusion, this research transcends the realm of a mere proof of concept, firmly establishing itself as a significant contributor to the practical application of AR in laboratory settings. By meticulously evaluating AR assistance for pipetting tasks, the study not only demonstrates the feasibility of such technology but also reveals its profound potential to revolutionise laboratory workflows. The findings go beyond theoretical validation; they provide a solid foundation for the practical implementation of AR in enhancing task efficiency and accuracy in real-world laboratory environments.

The successful application of AR in this context underscores a shift in how manual laboratory tasks can be approached, integrating cutting-edge technology to meet the demands of precision, speed, and user ergonomics. This study is a stepping stone along the way for AR to become an integral part of daily laboratory procedures, potentially transforming the landscape of research and experimentation. Thus, while rooted in the principles of proof of concept, the implications of this work extend into the tangible realms of laboratory innovation, setting a benchmark for future studies and technological developments in this field.

8 Disclaimer

This thesis was written with the help of DeepL ("DeepL Translator," 2023), Grammarly ("Grammarly," 2023), and ChatGPT 4.0 ("ChatGPT," 2023) to optimise the body of text. ChatGPT was used to improve readability and English language quality with prompts similar to the following example:

```
"[Initial draft of text, usually 3 - 5 sentences]
---
Optimise this text draft for readability and flow,
while being concise, maintaining chain of thought
and all relevant information."
```

The output was used as a basis for further improvements made by myself to optimally represent ideas, personal thoughts, and conclusions relevant to this thesis.

Bibliography

Adams, H., Stefanucci, J., Creem-Regehr, S., & Bodenheimer, B. (2022). Depth perception in augmented reality: The effects of display, shadow, and position. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 792–801 (cit. on p. 84).

Ali, N., Ullah, S., & Khan, D. (2022). Minimization of students' cognitive load in a virtual chemistry laboratory via contents optimization and arrow-textual aids. *Education and Information Technologies*, *27*(6), 7629–7652 (cit. on p. 8).

An, J., Poly, L.-P., & Holme, T. A. (2019). Usability testing and the development of an augmented reality application for laboratory learning. *Journal of Chemical Education*, *97*(1), 97–105 (cit. on p. 8).

Ansi/slas microplate standards. (2023, December 17). Retrieved December 17, 2023, from https://www.slas.org/education/ansi-slas-microplate-standards/. (Cit. on p. 38)

Apple human interface guidelines. (2023, December 19). Retrieved December 19, 2023, from https://developer.apple.com/design/human-interface-guidelines/. (Cit. on p. 80)

Apple vision pro. (2023, December 18). Retrieved December 18, 2023, from https://www.apple.com/apple-vision-pro/. (Cit. on p. 84)

Arnstein, L., Hung, C.-Y., Franza, R., Zhou, Q. H., Borriello, G., Consolvo, S., & Su, J. (2002). Labscape: A smart environment for the cell biology laboratory. *IEEE Pervasive Computing*, *1*(3), 13–21 (cit. on pp. 6 sq.).

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114–123 (cit. on p. 72).

Barrow, J., Forker, C., Sands, A., O'Hare, D., & Hurst, W. (2019). Augmented reality for enhancing life science education (cit. on p. 8).

Baumeister, J., Ssin, S. Y., ElSayed, N. A. M., Dorrian, J., Webb, D. P., Walsh, J. A., Simon, T. M., Irlitti, A., Smith, R. T., Kohler, M., & Thomas, B. H. (2017). Cognitive cost of using augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, *23*(11), 2378–2388 (cit. on p. 8).

Biocenter, jmu. (2023, April 16). Retrieved April 16, 2023, from https://www.biozentrum.uni-wuerzburg.de/en/about-the-biocenter/the-biocenter-today/. (Cit. on p. 11)

Bird, C. L., Willoughby, C., & Frey, J. G. (2013). Laboratory notebooks in the digital era: The role of elns in record keeping for chemistry and other sciences. *Chemical Society Reviews*, *42*(20), 8157–8175 (cit. on p. 2).

Blanco-Novoa, Ó., Fraga-Lamas, P., A. Vilar-Montesinos, M., & Fernández-Caramés, T. M. (2020). Creating the internet of augmented things: An open-source framework to make iot devices and augmented and mixed reality systems talk to each other. *Sensors*, *20*(11), 3328 (cit. on p. 82).

BMGLabtech. (2023, November 15). *Fluorescence intensity measurements*. Retrieved November 15, 2023, from https://www.bmglabtech.com/en/fluorescence-intensity/. (Cit. on pp. 44, 46)

Borriello, G. (2006). The invisible assistant: One lab's experiment with ubiquitous computing. *Queue*, *4*(6), 44–49 (cit. on p. 8).

Brooke, J. (1995). Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, *189* (cit. on pp. 58, 72).

Büttner, S., Funk, M., Sand, O., & Röcker, C. (2016). Using head-mounted displays and in-situ projection for assistive systems: A comparison. *Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments*, 1–8 (cit. on p. 8).

Chatgpt. (2023, December 20). Retrieved December 20, 2023, from https://chat.openai.com/. (Cit. on p. 89)

Cohen, B. H. (2008). *Explaining psychological statistics*. John Wiley & Sons. (Cit. on pp. 65 sq.).

Condino, S., Carbone, M., Piazza, R., Ferrari, M., & Ferrari, V. (2020). Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. *IEEE Transactions on Biomedical Engineering*, *67*(2), 411–419 (cit. on p. 36).

Da Rin, G. (2009). Pre-analytical workstations: A tool for reducing laboratory errors [Errors in Laboratory Medicine and Patient Safety]. *Clinica Chimica Acta*, *404*(1), 68–74 (cit. on pp. 1, 5).

Deepl translator. (2023, December 20). Retrieved December 20, 2023, from https://www.deepl.com/translator. (Cit. on p. 89)

Diaspro, A., Chirico, G., Usai, C., Ramoino, P., & Dobrucki, J. (2006). Photobleaching. In J. B. Pawley (Ed.), *Handbook of biological confocal microscopy* (pp. 690–702). Springer US. (Cit. on p. 40).

Dirnagl, U., & Przesdzing, I. (2016). A pocket guide to electronic laboratory notebooks in the academic life sciences. *F1000Research*, *5* (cit. on p. 6).

Echtler, F., Häussler, M., & Klinker, G. (2010). Biotisch: The interactive molecular biology lab bench. In *Chi'10 extended abstracts on human factors in computing systems* (pp. 3439–3444). (Cit. on p. 7).

EppendorfAG. (2015, June 19). *Pipetting technique userguide*. Retrieved December 12, 2023, from https://www.eppendorf.com/uploads/media/USERGUIDE_20_GB_Final.pdf. (Cit. on p. 6)

Falk, J., Mendler, M., & Kabisch, J. (2022). Pipette show: An open source web application to support pipetting into microplates. *ACS Synthetic Biology*, *11*(2), 996–999 (cit. on pp. 5, 9, 19).

Funk, M., Kosch, T., & Schmidt, A. (2016). Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 934–939 (cit. on p. 8).

Gerlach, B., Untucht, C., & Stefan, A. (2020). Electronic lab notebooks and experimental design assistants. *Good Research Practice in Non-Clinical Pharmacology and Biomedicine*, 257–275 (cit. on p. 2).

Github gba. (2023, December 9). Retrieved December 9, 2023, from https://github.com/Global-Biofoundries-Alliance/pipette-show. (Cit. on p. 37)

Grammarly. (2023, December 20). Retrieved December 20, 2023, from https://www.grammarly.com/. (Cit. on p. 89)

Guan, X. L., Chang, D. P. S., Mok, Z. X., & Lee, B. (2023). Assessing variations in manual pipetting: An under-investigated requirement of good laboratory practice. *Journal of Mass Spectrometry and Advances in the Clinical Lab*, *30*, 25–29 (cit. on p. 5).

Guerrero, S., Dujardin, G., Cabrera-Andrade, A., Paz-y-Miño, C., Indacochea, A., Inglés-Ferrándiz, M., Nadimpalli, H. P., Collu, N., Dublanche, Y., De Mingo, I., et al. (2016). Analysis and implementation of an electronic laboratory notebook in a biomedical research institute. *PLoS One*, *11*(8), e0160428 (cit. on p. 3).

Guerrero, S., López-Cortés, A., Garcıa-Cárdenas, J. M., Saa, P., Indacochea, A., Armendáriz-Castillo, I., Zambrano, A. K., Yumiceba, V., Pérez-Villa, A., Guevara-Ramırez, P., Moscoso-Zea, O., Paredes, J., Leone, P. E., & Paz-y-Miño, C. (2019). A quick guide for using microsoft onenote as an electronic laboratory notebook (F. Ouellette, Ed.). *PLOS Computational Biology*, *15*(5), e1006918 (cit. on p. 6).

Hammer, M., Schweitzer, D., Richter, S., & Königsdörffer, E. (2005). Sodium fluorescein as a retinal ph indicator? *Physiological measurement*, *26*(4), N9 (cit. on p. 39).

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 (cit. on p. 57).

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139–183). Elsevier. (Cit. on p. 57).

Heinrich, F., Schwenderling, L., Joeres, F., Lawonn, K., & Hansen, C. (2020). Comparison of augmented reality display techniques to support medical needle insertion. *IEEE Transactions on Visualization and Computer Graphics*, *26*(12), 3568–3575 (cit. on p. 8).

Hile, H., Kim, J., & Borriello, G. (2004). Microbiology tray and pipette tracking as a proactive tangible user interface. In *Lecture notes in computer science* (pp. 323–339). Springer Berlin Heidelberg. (Cit. on pp. 3, 7, 11 sq., 85).

Holland, I., & Davies, J. A. (2020). Automation in the life science research laboratory. *Frontiers in Bioengineering and Biotechnology*, *8*, 571777 (cit. on p. 1).

Hu, G., Chen, L., Okerlund, J., & Shaer, O. (2015). Exploring the use of google glass in wet laboratories. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2103–2108 (cit. on p. 3).

Iqbal, H., Tatti, F., & y Baena, F. R. (2021). Augmented reality in robotic assisted orthopaedic surgery: A pilot study. *Journal of Biomedical Informatics*, *120*, 103841 (cit. on p. 8).

Kanza, S., Willoughby, C., Gibbins, N., Whitby, R., Frey, J. G., Erjavec, J., Zupančič, K., Hren, M., & Kovač, K. (2017). Electronic lab notebooks: Can they replace paper? *Journal of cheminformatics*, *9*(1), 1–15 (cit. on p. 6).

Kapp, K., Sivén, M., Laurén, P., Virtanen, S., Katajavuori, N., & Södervik, I. (2022). Design and usability testing of an augmented reality (ar) environment in pharmacy education—presenting a pilot study on comparison between ar smart glasses and a mobile device in a laboratory course. *Education Sciences*, *12*(12), 854 (cit. on p. 8).

Kim, S., Cranor, B. D., & Ryu, Y. S. (2009). Fatigue: Working under the influence. *Proceedings of the XXIst Annual International Occupational Ergonomics and Safety Conference*, *317322* (cit. on p. 1).

Kong, F., Yuan, L., Zheng, Y. F., & Chen, W. (2012). Automatic liquid handling for life science: A critical review of the current state of the art. *Journal of laboratory automation*, *17*(3), 169–185 (cit. on p. 1).

Lange, S., & Niebling, F. (2022). *Augmented reality pipetting assistance for life science laboratory applications* (HCI Projekt). Julius-Maximilians University Wuerzburg. (Cit. on pp. 3, 11).

Lippi, G., Lima-Oliveira, G., Brocco, G., Bassi, A., & Salvagno, G. L. (2017). Estimating the intra- and inter-individual imprecision of manual pipetting. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *55*(7) (cit. on p. 6).

Martin, M. M., & Lindqvist, L. (1975). The ph dependence of fluorescein fluorescence. *Journal of Luminescence*, *10*(6), 381–390 (cit. on p. 39).

Material design. (2023, December 19). Retrieved December 19, 2023, from https://m3.material.io/. (Cit. on p. 80)

Meta quest 3. (2023, December 18). Retrieved December 18, 2023, from https://www.meta.com/de/quest/quest-3/. (Cit. on p. 82)

Mobile network subscriptions worldwide 2016-2022, statista. (2023, December 19). Retrieved December 19, 2023, from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. (Cit. on p. 80)

Mrtk unity. (2023, December 15). Retrieved December 15, 2023, from https://github.com/microsoft/MixedRealityToolkit-Unity. (Cit. on p. 11)

Norman, D. A. (1991). Cognitive artifacts. *Designing interaction: Psychology at the human-computer interface*, *1*(1), 17–38 (cit. on p. 1).

Palumbo, A. (2022). Microsoft hololens 2 in medical and healthcare context: State of the art and future prospects. *Sensors*, *22*(20), 7709 (cit. on p. 36).

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology*, *3*(1), 1–23 (cit. on p. 84).

Parks, D. F., Voitiuk, K., Geng, J., Elliott, M. A., Keefe, M. G., Jung, E. A., Robbins, A., Baudin, P. V., Ly, V. T., Hawthorne, N., et al. (2022). Iot cloud laboratory: Internet of things architecture for cellular biology. *Internet of Things*, *20*, 100618 (cit. on p. 82).

Perkel, J. M. (2017). The internet of things comes to the lab. *Nature*, *542*(7639), 125–126 (cit. on p. 82).

Pipette show build. (2022, February 22). Retrieved December 15, 2023, from https://pipette-show.de/build. (Cit. on p. 10)

Platr pipetting aid. (2022, March 28). Retrieved December 4, 2023, from https://biosistemika.com/platr/. (Cit. on p. 9)

Plebani, M. (2006). Errors in clinical laboratories or errors in laboratory medicine? *Clinical Chemistry and Laboratory Medicine (CCLM)*, *44*(6) (cit. on p. 5).

Poongothai, M., Subramanian, P. M., & Rajeswari, A. (2018). Design and implementation of iot based smart laboratory. *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, 169–173 (cit. on p. 82).

Rieder, J. S., Van Tol, D. H., & Aschenbrenner, D. (2021). Effective close-range accuracy comparison of microsoft hololens generation one and two using vuforia imagetargets. *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 552–553 (cit. on p. 36).

Rubacha, M., Rattan, A. K., & Hosselet, S. C. (2011). A review of electronic laboratory notebooks available in the market today. *JALA: Journal of the Association for Laboratory Automation*, *16*(1), 90–98 (cit. on p. 6).

Sartorius picus 2. (2023, December 18). Retrieved December 18, 2023, from https://www.sartorius.com/en/products/pipetting/electronic-pipettes/picus-2. (Cit. on p. 84)

Schneikart, G., & Mayrhofer, W. (2022). Revolutionizing solutions of technological assistance for the integration of lab and office activities in biomedical research. *Journal of Industrial Information Integration*, *26*, 100333 (cit. on p. 6).

Scholl, P. M., Schultes, T., & Van Laerhoven, K. (2015). Rfid-based compound identification in wet laboratories with google glass. *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*, 1–5 (cit. on p. 8).

Scholl, P. M., Wille, M., & Van Laerhoven, K. (2015). Wearables in the wet lab: A laboratory system for capturing and guiding experiments. *Proceedings of* *the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 589–599 (cit. on pp. 3, 8).

Schreiter, J., Schott, D., Schwenderling, L., Hansen, C., Heinrich, F., & Joeres, F. (2022). Ar-supported supervision of conditional autonomous robots: Considerations for pedicle screw placement in the future. *Journal of Imaging*, *8*(10), 255 (cit. on p. 8).

Schwenderling, L., Hansen, C., & Heinrich, F. (2021). Ar visualization of automated access path planning for percutaneous interventions. *Current Directions in Biomedical Engineering*, *7*(1), 48–52 (cit. on p. 8).

Shaer, O., Kol, G., Strait, M., Fan, C., Grevet, C., & Elfenbein, S. (2010). G-nome surfer: A tabletop interface for collaborative exploration of genomic data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1427–1436 (cit. on p. 7).

Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). Enhancing genomic learning through tabletop interaction. *Proceedings of the SIGCHI conference on human factors in computing systems*, 2817–2826 (cit. on p. 7).

Smartphone sales worldwide 2007-2022, statista. (2023, December 19). Retrieved December 19, 2023, from https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/. (Cit. on p. 80)

Swangnetr, M., Kaber, D., Vorberg, E., & Thurow, H. F. K. (2018). Workload assessment for manual and automated processes in life sciences. *Advances in Physical Ergonomics and Human Factors: Part I*, 121 (cit. on pp. 1, 5).

Swangnetr, M., Zhu, B., Kaber, D., Thurow, K., Stoll, N., & Stoll, R. (2012). Cognitive workload analysis and development of a stress index for life science processes. In *Advances in cognitive engineering and neuroergonomics* (pp. 71–80). CRC Press. (Cit. on p. 5).

Tabard, A., Hincapié Ramos, J. D., & Bardram, J. (2012). The elabbench in the wild: Supporting exploration in a molecular biology lab. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3051–3060 (cit. on pp. 7, 82).

Tabard, A., Hincapié-Ramos, J.-D., Esbensen, M., & Bardram, J. E. (2011). The elabbench: An interactive tabletop system for the biology laboratory. *Proceedings of the ACM international conference on interactive tabletops and surfaces*, 202–211 (cit. on p. 7).

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 73–80 (cit. on p. 2).

Techera, U., Hallowell, M., Stambaugh, N., & Littlejohn, R. (2016). Causes and consequences of occupational fatigue. *Journal of occupational and environmental medicine*, *58*(10), 961–973 (cit. on p. 1).

Tegally, H., San, J. E., Giandhari, J., & de Oliveira, T. (2020). Unlocking the efficiency of genomics laboratories with robotic liquid-handling. *BMC genomics*, *21*, 1–15 (cit. on p. 1).

Thermofisher pbs product catalog. (2023, December 17). Retrieved December 17, 2023, from https://www.thermofisher.com/order/catalog/product/10010023. (Cit. on p. 39)

Towns, M., Harwood, C. J., Robertshaw, M. B., Fish, J., & O'Shea, K. (2015). The digital pipetting badge: A method to improve student hands-on laboratory skills. *Journal of Chemical Education*, *92*(12), 2038–2044 (cit. on p. 6).

Trackman connected. (2023, February 5). Retrieved February 5, 2023, from https://de.gilson.com/trackman-connected.html. (Cit. on p. 9)

Venkatesan, M., Mohan, H., Ryan, J. R., Schürch, C. M., Nolan, G. P., Frakes, D. H., & Coskun, A. F. (2021). Virtual and augmented reality for biomedical applications. *Cell Reports Medicine*, *2*(7), 100348 (cit. on p. 8).

Vr headsets volume 2018-2022, statista. (2023, December 19). Retrieved December 19, 2023, from https://www.statista.com/forecasts/1331896/vr-headset-sales-volume-worldwide. (Cit. on p. 80)

Vuforia engine. (2023, December 15). Retrieved December 15, 2023, from https://developer.vuforia.com/library/. (Cit. on p. 11)

Vuforia image targets. (2023, December 8). Retrieved December 8, 2023, from https://developer.vuforia.com/library/objects/image-targets. (Cit. on p. 20)

Vuforia model targets. (2023, December 7). Retrieved December 7, 2023, from https://developer.vuforia.com/library/objects/model-targets. (Cit. on p. 12)

Werrlich, S., Daniel, A., Ginger, A., Nguyen, P.-A., & Notni, G. (2018). Comparing hmd-based and paper-based training. *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 134–142 (cit. on p. 81).

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, *41*(1), 67–85 (cit. on p. 31).

Yamamoto, T., Yamamoto, T., Yoshida, K., Fuchinoue, S., Hayama, Y., Hidano, A., & Tsutsui, T. (2014). Effectiveness of a short training session for improving pipetting accuracy. *Accreditation and Quality Assurance*, *19*(6), 459–463 (cit. on p. 6).

Yung, M., Manji, R., & Wells, R. P. (2017). Exploring the relationship of task performance and physical and cognitive fatigue during a daylong light precision task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(7), 1029–1047 (cit. on p. 1).

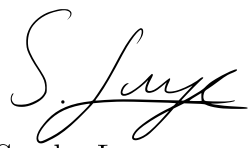
Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht.

Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Würzburg, December 20, 2023



Sascha Lange

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Versicherung zur selbständigen Leistungserbringung

Titel der Masterarbeit

Evaluating the Effect of AR-Assistance on Task Performance in Manual Liquid Handling Scenarios:
A Comparative Study

Thema bereitgestellt von (Titel, Vorname, Nachname, Lehrstuhl):

Dr. Martin Fischbach, Informatik IX, Universität Würzburg

Eingereicht durch (Vorname, Nachname, Matrikel):

Sascha Lange, 2609220

Ich versichere, dass ich die vorstehende schriftliche Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die benutzte Literatur sowie sonstige Hilfsquellen sind vollständig angegeben. Wörtlich oder dem Sinne nach dem Schrifttum oder dem Internet entnommene Stellen sind unter Angabe der Quelle kenntlich gemacht.

Weitere Personen waren an der geistigen Leistung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines Ghostwriters oder einer Ghostwriting-Agentur in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar Geld oder geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Arbeit stehen.

X Mit dem Prüfungsleiter bzw. der Prüfungsleiterin wurde abgestimmt, dass für die Erstellung der vorgelegten schriftlichen Arbeit Chatbots (insbesondere ChatGPT) bzw. allgemein solche Programme, die anstelle meiner Person die Aufgabenstellung der Prüfung bzw. Teile derselben bearbeiten könnten, entsprechend den Vorgaben der Prüfungsleiterin bzw. des Prüfungsleiters eingesetzt wurden. Die mittels Chatbots erstellten Passagen sind als solche gekennzeichnet.

Der Durchführung einer elektronischen Plagiatsprüfung stimme ich hiermit zu. Die eingereichte elektronische Fassung der Arbeit ist vollständig. Mir ist bewusst, dass nachträgliche Ergänzungen ausgeschlossen sind.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung zur Versicherung der selbstständigen Leistungserbringung rechtliche Folgen haben kann.

Würzburg, 20.12.2023

Ort, Datum, Unterschrift