Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Iterative Enhancement of AR-based Task Assistance System for Construction Workers: Insights from Comparative User Study and Thematic Interview Analysis

Xiang, Yuan[1], Qipei Mei[2], and Xinming Li[3]*

[1] Ph.D. Student, Department of Mechanical Engineering, University of Alberta, Canada
[2] Assistant Professor, Department of Civil & Environmental Engineering, University of Alberta, Canada
[3] Associate Professor, Department of Mechanical Engineering, University of Alberta, Canada

ABSTRACT: Construction workers face challenges in efficiently utilizing augmented reality (AR) task assistance systems due to issues such as design complexity, information overload, and suboptimal system usability. Despite AR's demonstrated capacity to enhance productivity and safety, many existing systems lack systematic refinement processes grounded in industrial design theory and often fail to incorporate user-centered principles suitable for real-world construction contexts. This study aims to iteratively enhance and evaluate an AR-based task assistance system tailored for construction workers, by applying industrial design theory and a user-centered approach to improve system usability, interface behavior, and user satisfaction. The study begins by illustrating the system's core functionalities, design parameters, and the prototype development process. The study involved a comparative experiment with six participants, each completing two test sessions of using System A (baseline) and System B (iterated version) with a two-week interval to reduce learning effects. Each round testing was followed by a structured evaluation involving task performance analysis, System Usability Scale (SUS) assessments, and a semi-structured interview feedback. The qualitative data were analyzed using NVivo 14 software and revealed six critical areas for system optimization, including better interface alignment with worker behavior, simplifying task steps, clarifying instructional content, incorporating legends, enhancing object search support, and integrating basic voice control. Statistical results indicate significant improvements in task completion time, error rates, and SUS scores between initial and refined system versions, demonstrating the effectiveness of iterative enhancements in improving productivity, reducing operational errors, and increasing usability in real-world construction scenarios. The findings contribute an empirical, user-verified framework for AR system development in construction, bridging industrial design and human factors engineering. This work provides a solid foundation for scalable, user-aligned AR deployment in industrial settings.

1. INTRODUCTION

Augmented Reality (AR) is increasingly recognized as a transformative technology that enhances work efficiency, quality, and safety by superimposing digital information onto real-world views without interfering to physical environment (Tan et al. 2022). In the context of the construction industry, AR's advanced functionalities and its seamless integration with Industry 4.0 technologies offer diverse applications across various phases of construction projects, including architectural design, on-site operations, construction management and post-construction maintenance (Hajirasouli et al. 2022). From the perspective of worker-oriented application, AR serves as an egocentric, close-range augmented tool that provides real-time detailed task guidance, timely safety warnings, quality control and inspection assistance, remote collaboration annotations, and serve as mutual bridge of human-robot interactions (Hajirasouli et al. 2022).

Additionally, it facilitates bidirectional communication in human-robot interactions, supporting autonomous and semi-autonomous construction processes (Maio et al. 2024). Among these applications, hands-free AR-based task assistance systems have gained significant adoption and commercialization, particularly in training novice operators and enhancing task execution in industrial settings (Eswaran and Raju Bahubalendruni 2023). Construction workers can access a seamless blend of digital and physical world, enabling them to execute tasks in a more accurate and well-managed way. It significantly enhances workers' perception and performance, thereby potentially revolutionizing traditional tasks completion ways (Moencks et al. 2022). This capability not only optimizes workflow but also reduces the likelihood of errors and accidents, contributing to a safer and more productive working environment (Di Pasquale et al. 2022). Meanwhile, these applications are not only limited to the construction sector, but also extend to industries such as manufacturing, aerospace, and healthcare, offering valuable references to empower workers and augmented their capabilities (Yang et al. 2020).

Previous research has pointed out several critical limitations of AR systems from the view of human factor and cognitive science, including issues related to system usability, design complexity, information overload, and use satisfaction(Wolf 2022). For example, existing AR systems often struggle to deliver precise real-time guidance in complex construction environments (Chen et al. 2024). Moreover, their user interfaces tend to be overly complex and overloaded with information, which impedes rapid adoption among workers (Stefanidi et al. 2022, Seeliger et al. 2022). Despite the great potential and advantages of AR and integrated multi-technologies in worker-related applications, there is a significant research gap in systematically applying industrial design theory to iteratively improve these systems (Quandt and Freitag 2021). Most existing AR systems lack the necessary user-centric refinements that make them truly effective in real-world construction environments. Furthermore, while scholars have conducted numerous comparative studies between AR tools and traditional paper-based instructions, these studies primarily focus on their relative effectiveness without considering the initial design flaws of the AR systems themselves. This lack of self-verification leads to an underestimation of the true potential of AR applications (Lima and Hwang 2024). If the initial AR system has significant design deficiencies, directly comparing it with paper-based instructions might unfairly highlight its shortcomings and leave a gap in the academic maturity of comparative studies.

To address these challenges and fulfill the above gaps, it is essential to adopt a two-pronged approach. Firstly, this research will evaluate and iteratively enhance AR-based task assistance systems designed for construction workers, so that continuously converge the design intent or expected experience of AR systems with the actual needs of construction workers. This research focuses on continuously improving system usability, enhancing user-center interface and context design, and increasing user satisfaction through systematic application of industrial design, ergonomics factors, and cognitive science. Secondly, this research also optimizes the AR-based task assistance system to a robust state before conducting comparative studies with traditional methods. By first addressing and mitigating design flaws, the research can set a valuable benchmark and more accurately evaluate the incremental benefits of AR systems. Specifically, this study answered question regarding to "How can a user-centered iterative process be designed to ensure continuous improvement of designed AR task assistance system and improve its usability and user satisfaction?". The innovation of this study lies in continuously enhancing the performance and user experience of the AR task assistance system, by presenting a process of iterative improvement and rigorous system self-verification. This research adopted a comprehensive evaluation methodology, including comparative user studies, task performance data analysis, System Usability Scale (SUS) feedback, and thematic analysis of semi-structured interviews. These methods not only provide empirical foundations for the design and optimization of AR systems but also demonstrate its adaptability and potential applications across various scenarios within or out construction industry.

The background and motivation of this research has been elaborated in this section and the remaining structure of this paper is as follows. Section 2 of Methodology first describes the key functionalities, design parameters, and prototype development process and then explains the experimental design, data collection methods, and how to analyze data and understand analysis metrics. Section 3 of Results first identifies the most urgent improvement needs raised by user interactions and their feedback, and then presents results from the comparative user study. Section 4 mainly discusses the actual effects of various improvement measures and user feedback, and comprehensively analyze the improvement of system performance and user experience from multi-dimensional data. Section 4 summarizes research findings, discuss research innovation and contribution, and highlights current limitation and future direction in this research field.

2. METHODOLOGY

2.1 System Overview

This study employed the Microsoft Trimble XR10, an augmented reality (AR) helmet integrated with HoloLens 2, which combines a certified construction hard hat with the HoloLens 2 to facilitate convenient, wireless operation in real construction sites. The prototype system was developed using the Unity engine (version 2021.3.20f1 LTS), programmed in C#, and utilized the Mixed Reality ToolKit (MRTK 2.8.3) package (lolambean 2022) for asset creation and interactive UI design.

Overall, the AR-based task assistance system developed in this study is to provide real-time contextual task guidance for construction workers and to enhance their efficiency and accuracy. It is fundamentally an open-ended platform which can be customized to assist with a wide range of tasks, depending on the needs of both end users and system designers and the specific requirements of the tasks at hand. Furthermore, it is designed to be adaptable to various construction scenarios, accommodating both indoor and outdoor environments. However, its performance may be affected by variations in lighting conditions, which represent a critical parameter for the visualization of AR holograms.

Generally, it encompasses following layers: Perception layer includes environmental sensors and user motion capture devices to collect real-time data. Data processing layer processes collected data locally, including generation of real-time task instructions and object detection. Interaction layer engages with users by displaying task steps and overlaying instant information. Communication layer facilitates data transmission between system and remote server or other devices.

The development process involved several critical steps. Initially, researchers identified workers' needs by referencing existing studies and analyzing usage patterns documented in prior research. Task procedures, sequences of activities, required tools, and official manuals were incorporated as system inputs to define system functionalities and performance requirements. Then an initial system prototype was developed using Unity engine, focusing on implementing core functional modules. These modules included user login and registration module, holograms and interface guidance for new users module, user interaction module, debugging and system control module, eye tracking data and user activity logging module, and object detection module (optional).

2.2 Experiment Design

There were six participants involved in this study, each with varying levels of familiarity with augmented reality (AR) and virtual reality (VR), as well as different technical expertise and understanding of construction processes. This diverse background was to ensure the representativeness of the findings. The participants came from multiple professions and occupations, covering different task types and work environments and also including both novice and experienced construction workers. All participants signed an informed consent form before the experiment began and understood the purpose and procedures of the study.

As shown in Figure 1, the experiment was conducted in a controlled environment of Occupational Ergonomics Research Lab at University of Alberta, that simulated a real construction site. This research selected Metaltech Multipurpose 4-in-1 6 ft. Baker Scaffold as the specific user case for experiment, which included various typical construction elements, such as scaffolding and miter saw stand, construction tools, and building materials. This task scenario was chosen to reflect typical construction activities and ensure ecological validity. The experimental procedure consisted of two phases, spanning a total duration of two weeks. In the first phase, all participants signed informed consent forms and received experimental introductions. Quick start guides in both interacting with AR holograms and system user interface is introduced to users and served as a training section to familiarize themselves with the AR headset and the user interface. Then, in the test phase 1, participants used System A (AR task-assisted prototype before iteration) to complete the predefined task of assembling a miter saw stand. Upon task completion, they were asked to fill out the System Usability Scale (SUS) questionnaire and participate in semi-structured interviews. Two weeks later in the test phase 2, participants performed the same tasks using System B (AR task-assisted prototype after iteration). Following this second trial, they completed the SUS questionnaire again and had the option to participate in an additional semi-structured interview. The comparative analysis of task performance and SUS scores enabled the identification of key areas for system iteration and the evaluation of improvements in AR system usability.

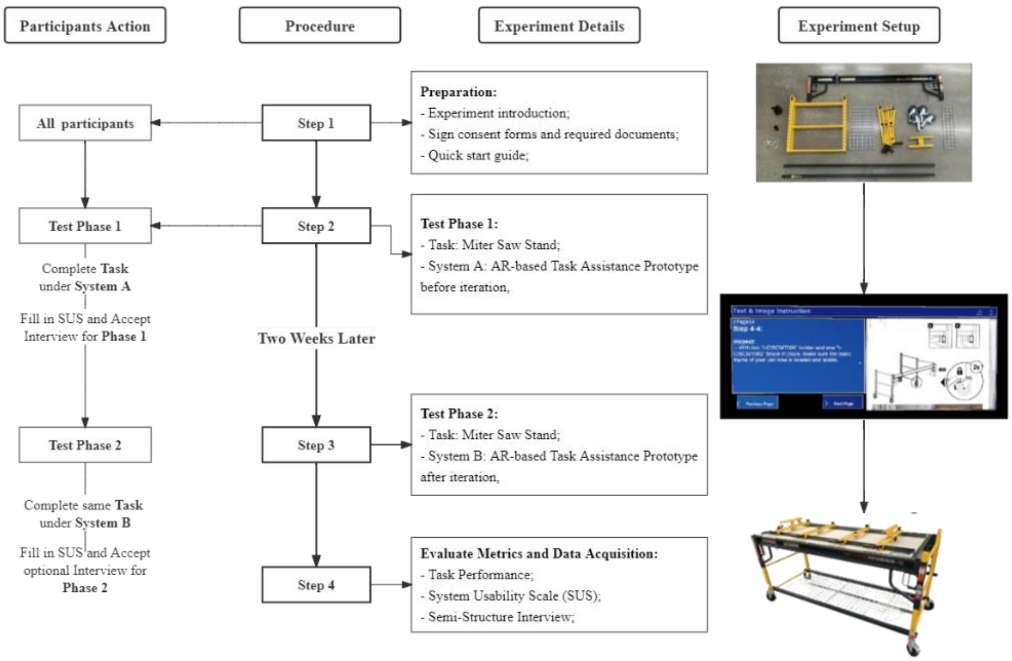


Figure 1 Experimental Process

2.3 Data Collection and Analysis

2.3.1 Task Performance

This study collects task performance data through a combination of automatic system recording and manual observation, focusing on two key metrics: task completion time and error rate. Task completion time was derived from timestamped user activity logs, recorded at each step from task initiation to completion within the HoloLens local storage. The accuracy rate was quantified by counting the number of instances where participants followed the provided instructions during task execution.

2.3.2 System Usability Scale (SUS)

Created by John Brooke in 1986, System Usability Scale (SUS) is originally used as a quick and straightforward tool for usability assessment and has since become a widely adopted questionnaire for evaluating system usability (Escalada-Hernandez et al. 2024). To gather a subjective self-assessment of system usability, participants in this research completed SUS after testing both AR systems before and after iteration. SUS consists of ten statements rated on a five-point Likert scale, ranging from “Strongly Disagree” to “Strongly Agree”. Each item evaluates aspects such as ease of use, complexity, and system consistency. The overall SUS score is computed by converting individual responses into a single composite score ranging from 0 to 100. Though there is a SUS score available for comparison, a more useful and understandable interpretation method is to convert it into a percentile ranking. The original SUS score is a relative indicator, while the percentile ranking can more intuitively understand the position of the system’s usability among all the evaluated systems, thereby more accurately evaluating its strengths and weaknesses.

2.3.3 Semi-Structured Interviews

After completing the tasks, participants were interviewed to capture their insights and experiences regarding the AR system. The interviews focused on areas of overall satisfaction, improvement suggestions and expectations, and interview conclusion. The interview format is a semi-structured interview and structured outline is presented in Table 1.

Table 1. Semi-Structure Interview Questions

Focus Area	Sample Questions and Probes
Overall Satisfaction	How satisfied are you with the system overall? Please rate on a scale of 1-10. How do you think the system performs in terms of improving your work efficiency and performance?

	Which features do you think are missing or could be further enhanced?
Improvement Suggestions and Expectations	Which steps do you think can be simplified or optimized?
	Do you think there needs to be more guidance or prompts? Which specific guidance or prompts would be helpful?
	Have you encountered any system crashes or errors that affected your operation?
	Do you think the system needs to be more user-friendly? In which specific aspects?
Interview Conclusion	Do you have any other comments or suggestions that we haven't covered?

To evaluate the impact of the iterative AR system improvements on task performance, two key performance metrics were analyzed: task duration (in seconds) and task accuracy (%). Paired t-tests were conducted for each measure with all variables met the normality assumption ($p > 0.05$). Effect sizes were calculated using Cohen's d, defined as the mean difference divided by the pooled standard deviation. A significance threshold of $p < 0.05$ was applied for all statistical comparisons, and adjusted p-values (Benjamini-Hochberg correction) were reported to control for multiple comparisons.

This study uses in-depth interview data of six participants as the information mining materials, and adopts the qualitative thematic analysis method to inductively extract the interviewees' expectations for the improvement of AR-based task assistance system from the bottom up (Byrne 2022). The subjects will answer the following questions and will either volunteer or be inspired by the researchers to engage in a discussion around the question. Interview is recorded after obtaining the consent of the interviewee and its duration is around 10-25 minutes.

When analyzing the interview data, there are three main steps. First of all, the voice interview materials are under text processing and anonymization coding. And then this research uses NVivo14 software to conduct thematic analysis and coding of the text materials. Finally, this research highly condensed users' suggestions for AR task assistance system improvements. Regarding to the second step, thematic analysis coding can be further divided into three steps: 1) Open coding of the interview data, listing its initial concepts and categories; 2) Axial Coding, summarizing and organizing the open coding, analyzing the correlation between different parts, and sorting out the main categories; 3) Selective coding, crossing the main category with other different conceptual categories.

In terms of qualitative exploratory analysis based on users' feedback, although some evaluation indicators such as standardization, representativeness and generalization in quantitative research are not recommended, its effectiveness and representativeness are still reflected. First of all, this paper uses two researchers to conduct simultaneous, anonymous, and alternating coding. Then they conducted in-depth discussions on the parts with large differences of opinion until the opinions are unified, thereby enhancing the credibility of the research. In addition, involved participants in this experiment reflect the commonalities and differences across subjects, further verifying the validity of the research. In addition, in order to verify the external validity and coding saturation of small sample studies, this study uses field notes of participants from the pilot experiments and for further supplementary verification. Since no more coding and new categories were generated through the content of the pilot experiments, it is verified that the coding has reached saturation.

3. RESULTS

3.1 Identifying Improvement Areas

After processing all transcribed interview texts and user interaction data, the data were imported into NVivo 14 software for sentence screening and thematic coding. Following conceptual refinement, the analysis ultimately identified six key themes: Overall Satisfaction, Functional Effectiveness, Step Optimization, User-Friendliness, Guidance and Prompts Improvement, and Device Issues. These six themes, along with 22 corresponding subcategories, present a comprehensive scoring framework that not only encapsulates a macro-level understanding of the AR system's usability and functionality but also provides multiple perspectives on system performance optimization. Table 2 summarizes the improvement suggestions of each user and their proportion data, showing the common problems and improvement needs encountered by different users during the use of the system. Specifically, "Missing Features – Searching for Objects" and "Steps Simplification" are the most frequently reported improvement suggestions, with an average proportion of 22.61% and 16.57% respectively, indicating that these functions require priority attention and improvement. In addition, users also have high demands for "Clear Instructions", "Enhanced Features - Adjust Follow Me Function", "Add Voice Command", and "Add a Legend" with an average proportion of 10.56%, 8.61%, 7.91%, and 6.60% respectively. Other suggested improvements received comparatively lower frequencies.

Table 2 Summary of Improvement Suggestions and Their Proportion (%)

System Improvement Suggestions	P1	P2	P3	P4	P5	P6	Avg.
1. Continuously Taking Photos	6.11	0	0	0	0	0	1.02
2. Popping Up HoloLens System Menu	0	20.26	3.38	0	0	0	3.94
3. Shaking UIs	0	0	0	11.25	0	0	1.88
4. Helmet Might Fall Off	0	0	4	0	0	4.81	1.47
5. Enhanced Features - Adjust Follow Me Function	25.76	0	3.85	0	15.91	6.15	8.61
6. Missing Features - Searching for Objects	0	40.53	21.23	37.30	20.02	16.58	22.61
7. Text- and Image-based Guidance	0	0	10.31	2.57	17.69	7.49	6.34
8. 3D Animation Models	0	0	0	0	0.44	0	0.07
9. Video Guidance	0	0	13.85	0	0.33	0	2.36
10. Clear Instructions	0	0	11.23	33.12	12.57	6.42	10.56
11. Steps Sequence Against Cognition	0	0	3.08	0	10.12	0	2.20
12. Steps Simplification	5.68	39.21	14.77	15.76	15.46	8.56	16.57
13. Add a Legend	29.69	0	6.15	0	0	3.74	6.60
14. Add Voice Command	10.04	0	0	0	7.45	29.95	7.91
15. Increase Responsiveness and Sensitivity	22.71	0	8.15	0	0	16.31	7.86

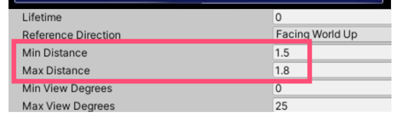
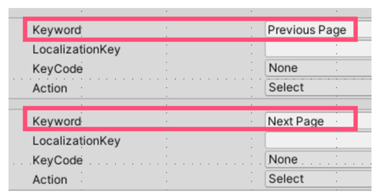
3.2 Implementing Iterative Improvements

Based on the identified areas, specific enhancements were implemented in the AR system, resulting in an updated version following the iterative process. These included refining instructions and simplifying steps, enhancing follow me functions, and integrating additional features such as searching for objects and adding a legend. Table 3 presented user feedback coding after using system A and how these issues were addressed in System B accordingly.

- The object searching functionality represents a significant adjustment and remains an active research topic within AR systems. While previous studies have explored this feature, ongoing research efforts will continue to refine its implementation (Yuan et al. 2023). In the current iteration, this function was enhanced by incorporating additional object images into users' field of view, thereby improving object recognition and retrieval. Furthermore, this feature has been designated as a future research direction for further development.
- The Follow Me function was adjusted to ensure an optimal following mode and distance, allowing instructional panels to dynamically track users within the physical environment while maintaining visibility and usability.
- Although the task steps and instructions were originally derived from the official scaffolding assembly manual provided by the manufacturer, certain descriptions and images were found to be unclear or ambiguous. To address these issues, both textual and visual instructions were revised based on user feedback, ensuring improved clarity and comprehension.
- Voice command functionality was identified as a critical enhancement to support hands-free interaction, particularly in construction environments where users frequently operate tools or wear gloves. This enhancement was primarily implemented to facilitate multimodal interaction, allowing users to operate the system via voice commands.
- The legend feature was introduced to provide explicit explanations for symbols used in image-based instructions. This addition not only clarifies the symbolic language but also guides users in interpreting instructional sequences, thereby improving overall system usability.

Table 3 Improvement areas and user feedback

Actionable Insights	User Feedback Coding After Using System A	Improvement Actions in System B

6. Missing Features – Searching for Objects	“But in a real construction site scenario, there will be many similar things, making it hard to choose. For example, a screw could have different sizes and uses. Wheels in scaffolding have different sizes, and although you know they are wheels, how do you know which one to use?”	
12. Steps Simplification	“And you have a single step or single sub-step for those repeat, for example, please assemble component A and in the next page, you will say please repeat them for the rest of component BCD. But actually at that time I already finished all of them, you can consider to delete the repeat section or merge the repeat section as a last command in each page.”	N/A
10. Clear Instructions	“The last step is the most unclear and very confusing. I still don’t know if I put them at the wrong place or not because there are only two things.”	
5. Enhanced Features – Adjust Follow Me Function	“When I use the follow me command, it was following me but it should have a bit of distance or something. It was coming to my face. It was a bit troublesome. I was supposed to pick a thing and it looks like someone is continuously at my face and it was a bit pressuring.”	
14. Add Voice Command	“It will be much helpful if you could integrate the voice command in the system. Because construction workers are using their both hands to conduct the construction tasks. With both hands are occupied, we hate to take anything to track those steps. But you wear the AR glass, you can see everything only by using your voice. That will be really cool.”	
13. Add a Legend	“It will be a better option to have something like a Legend. If a person who is not very knowledgeable in this field, just like labor workers, and they don’t know much, if they will be legend, always there, then it might help them.”	

3.3 System Usability Rates

According to Table 4, a comparative analysis of the System Usability Scale (SUS) scores between System A and System B reveals notable differences in usability performance. System A recorded an average SUS score of 57.08, which falls within the “OK” category, corresponding to a percentile rank of 21.95%, indicating that it falls below the industry benchmark. In contrast, System B achieved an average SUS score of 87.92, classified as “Best Imaginable,” with a percentile rank of 97.93%, significantly surpassing industry benchmarks. In terms of statistical distribution, System A exhibited a standard deviation of 20.21, suggesting considerable variability in user ratings. Conversely, System B demonstrated a substantially lower standard deviation of 6.41, indicating a higher level of consistency in user evaluations and a more reliable user experience. Regarding system acceptability and Net Promoter Score (NPS) ratings, System A was categorized as “Marginal” in acceptability and “Detractor” in NPS assessment, reflecting moderate user endorsement. Meanwhile, System B was rated as “Acceptable” in acceptability and classified as a “Promoter” in NPS, suggesting a substantial increase in user satisfaction and likelihood of recommendation. These findings underscore the substantial improvements in usability and user satisfaction achieved through the iterative enhancement of the AR system.

Table 4 Comparison Results of AR-based Task Assistance System Before and After Iteration

	System A	System B	Reference
SUS Study Score	57.08	87.92	
Standard Dev.	20.21	6.41	
Adjective	OK	Best Imaginable	(Bangor et al. 2009)

Grade	D	A	(Sauro and Lewis 2016)
Acceptability	Marginal	Acceptable	(Bangor et al. 2008)
Quartile	1st	4th	(Sauro and Lewis 2016)
NPS Scale	Detractor	Promoter	
Industry Benchmark	Below Average	Above Industry Standard	(Lewis and Sauro 2018)
Percentile	21.95	97.93	(Sauro and Lewis 2016)

3.4 Task Performance

The task completion time and accuracy were analyzed to compare the efficiency between System A and System B. As shown in Figure 2, the results revealed that System B significantly reduced task completion time compared to System A (t = -3.59, p = 0.016, Cohen's d = -1.47). The mean task duration for System A was 1647.87 ± 592.52 seconds, whereas for System B, it was 915.35 ± 418.17 seconds, indicating a substantial improvement in task efficiency. The observed large effect size (Cohen's d = -1.47) further supports the effectiveness of the optimized AR system in reducing task completion time. Meanwhile, the results demonstrated a statistically significant improvement in accuracy with System B (t = 5.44, p = 0.003, Cohen's d = 2.22). The mean accuracy for System A was 70.14 ± 6.93%, while for System B, it increased to 88.89 ± 8.19%, reflecting enhanced task performance with the refined AR system. The large effect size (Cohen's d = 2.22) suggests that the observed improvement in task accuracy was not only statistically significant but also practically meaningful.

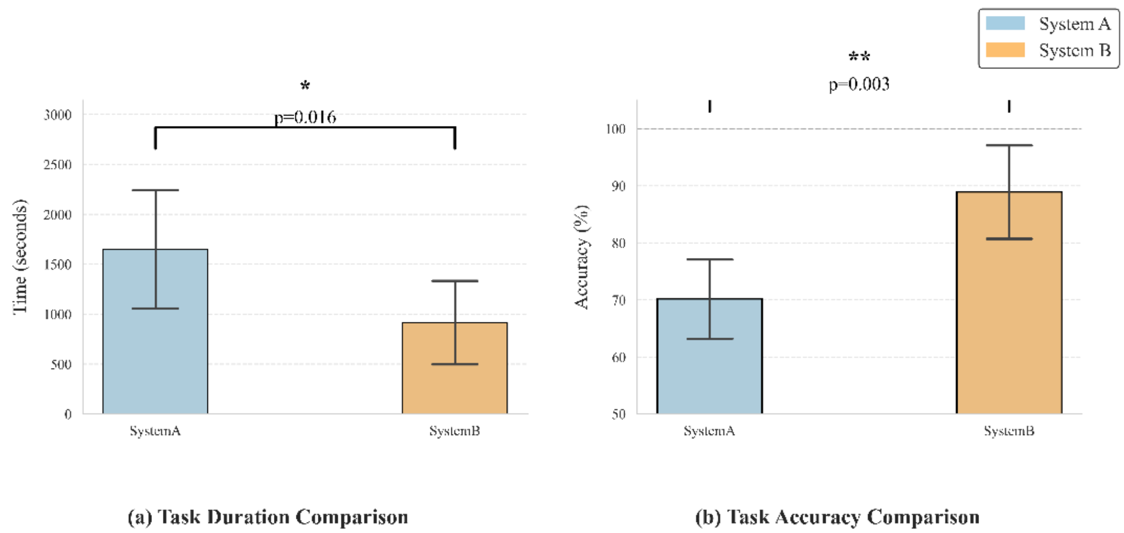


Figure 2 Comparison of task duration and accuracy

To further understand the nature of task errors, further analysis classified each error into one of five categories based on its cause and context: (1) search/navigation errors, (2) assembly/attachment errors, (3) instruction interpretation errors, (4) safety critical errors, and (5) AR interaction errors. This classification was inductively concluded based on a further review of user behavior, video-coded task breakdowns of experimental trials, and the root cause reported by users. As illustrated in Figure 3, System A revealed a significantly higher total number of errors across all categories, with particularly high frequencies in AR interaction errors (n = 84), instruction interpretation errors (n = 38), and assembly/attachment errors (n = 24). In contrast, System B showed marked reductions in every category, most notably in AR interaction errors (n = 18), reflecting improved interface responsiveness and interaction design. Statistical analysis using paired t-tests confirmed that these reductions were significant for assembly/attachment errors (p = 0.008), instruction interpretation errors (p = 0.001), and AR interaction errors (p < 0.001). Although search/navigation and safety-critical errors also declined in System B, these changes did not reach statistical significance (p > 0.05). These results suggest that iterative design changes, such as clearer visual cues and interface simplification, effectively addressed both cognitive and procedural challenges faced by users.

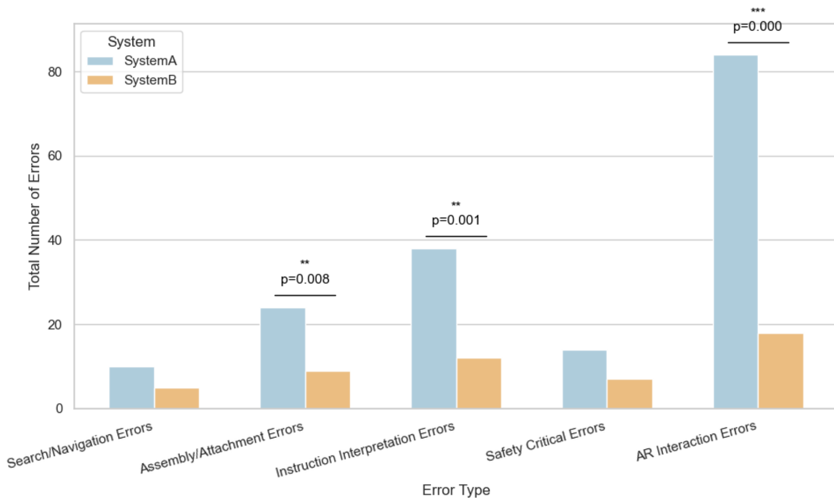


Figure 3 Comparison of Error Types Between System A and System B

4. DISCUSSION

This study first identified a series of user-driven improvement suggestions that reflect both the practical needs and interactive experiences of users when utilizing the AR task assistance system. As presented in Table 2, the feedback from users predominantly concentrated on three key aspects: functional enhancement, step optimization, and overall user-friendliness. Among these, object searching emerged as the most frequently suggested improvement, indicating a strong user demand for more precise and adaptive object detection functionalities. Users expressed a preference for system panels that could more effectively track their movements and provide enhanced object search capabilities. Secondly, a high demand for 'clear guidance' and 'simplified steps' was observed, highlighting the necessity of reducing operational complexity during task execution. Researchers noticed that all participants consistently skipped specific steps at the same points in the workflow, suggesting a misalignment between system instructions and user cognitive processes, necessitating further refinement.

The System Usability Scale (SUS) analysis further illustrated a significant improvement in system usability following the iterative refinements. Specifically, System B exhibited a SUS percentile rank of 97.93%, whereas System A only reached 21.95%, demonstrating that the enhanced system surpassed 97.93% of comparable systems in terms of user experience and usability. This substantial improvement underscores the effectiveness of iterative refinements in elevating user satisfaction and interaction quality. Additionally, the consistency of user ratings also improved markedly, with the standard deviation decreasing from 20.21 of System A to 6.41 of System B, suggesting a more stable and uniform user perception of the system's usability.

The findings of this study empirically validate the substantial impact of iterative AR system enhancements on both task efficiency and accuracy. The observed reduction in task completion time suggests that the optimized AR system provided more intuitive and streamlined task guidance, leading to improved workflow efficiency. These results align with previous studies, indicating that AR-based guidance systems minimize task complexity and accelerate execution by enhancing information accessibility and reducing mental processing demands. Furthermore, increasing task accuracy suggests that the improved system not only facilitated faster execution but also enhanced user comprehension and task precision. The higher accuracy rates indicate that the refined AR interface contributed to clearer task instructions and reduced procedural ambiguity, a critical factor in construction and industrial environments where errors can pose safety risks and operational inefficiencies. Future research should focus on evaluating long-term user adaptation, cognitive workload variations, and the scalability of AR-based task assistance across diverse construction environments to further substantiate its effectiveness in real-world applications.

Overall, user feedback provides valuable insights and directions for further optimization of the system. By prioritizing the improvement of high-frequency feedback functions and gradually solving other issues, AR-based task assistance system can better meet the needs of construction workers and improve their work efficiency and satisfaction. These improvements not only enhance overall performance of the system, but also provide empirical support and optimization paths for the application of AR technology in the construction industry. However, this study also has some limitations, including the small sample size and the limitation of the experimental environment. Future research should further validate these findings with a larger sample size and in a real construction environment to ensure the generalizability and practicality of the results. Users have a strong demand for improvement in the object search function. Although we have already started to integrate the YOLOv8 model with AR technology to improve the accuracy and efficiency of object detection and recognition, we can go deeper in the future (Castelo et al. 2024).

5. CONLUSION AND FUTURE WORK

This research proposed a structured framework for an AR-based task assistance system tailored for construction workers by integrating human interaction, the physical environment (interaction with tangible elements), the virtual environment (interaction with AR elements), and information management. The research findings indicate that, following multiple iterative modifications from the initial prototype, the system exhibited substantial improvements in usability and user satisfaction. In the first phase of the experiment, user feedback guided six critical system refinements: adjusting the 'Follow Me' function, simplifying task steps, clarifying instructions, integrating legends, enhancing object search capabilities, and implementing basic voice commands. Subsequently, comparative experiments on task performance and system usability empirically validated the superior performance of System B. The task execution time shorten by 44.45% with 26.73% increasing on task accuracy. The System Usability Scale (SUS) percentile rank of System B reached 97.93%, significantly surpassing the 21.95% of System A, underscoring the notable advancements achieved through iterative optimization.

These findings hold significant implications for the integration of AR and its associated technologies in the construction industry. From a prototyping perspective, this study demonstrates a systematic methodology that incorporates user-centered, demand-driven, and experience-based approaches into prototype development. By systematically integrating principles from industrial design, ergonomics, and cognitive science, this research establishes a structured and repeatable process for AR system refinement. From a methodological rigor perspective, this study ensured that the system underwent iterative optimization before being subjected to comparative analyses against traditional methods. This approach establishes a valuable benchmark for accurately assessing the incremental benefits of AR systems and provides empirical validation of their effectiveness in practical applications.

In summary, this research not only significantly enhances the usability and effectiveness of AR-based task assistance systems for construction workers but also proposes a robust and adaptable framework applicable to various worker-centered scenarios. Future research should prioritize expanding the sample size, conducting extended evaluations in real-world construction environments, and refining system functionalities based on continuous user feedback. Such efforts will further ensure the practical applicability and scalability of AR technology across diverse industrial settings.

REFERENCES

Bangor, A., P. Kortum, and J. Miller. 2009. "Determining what individual SUS scores mean: Adding an adjective rating scale." *Journal of usability studies*, 4 (3): 114–123. Usability Professionals' Association Bloomingdale, IL.

Bangor, A., P. T. Kortum, and J. T. Miller. 2008. "An Empirical Evaluation of the System Usability Scale." *International Journal of Human-Computer Interaction*, 24 (6): 574–594. https://doi.org/10.1080/10447310802205776.

Byrne, D. 2022. "A worked example of Braun and Clarke's approach to reflexive thematic analysis." *Qual Quant*, 56 (3): 1391–1412. https://doi.org/10.1007/s11135-021-01182-y.

Castelo, S., J. Rulff, E. McGowan, B. Steers, G. Wu, S. Chen, I. Roman, R. Lopez, E. Brewer, C. Zhao, J. Qian, K. Cho, H. He, Q. Sun, H. Vo, J. Bello, M. Krone, and C. Silva. 2024. "ARGUS: Visualization of AI-Assisted Task Guidance in AR." *IEEE Transactions on Visualization and Computer Graphics*, 30 (1): 1313–1323. https://doi.org/10.1109/TVCG.2023.3327396.

Chen, H., L. Hou, S. Wu, G. Zhang, Y. Zou, S. Moon, and M. Bhuiyan. 2024. "Augmented reality, deep learning and vision-language query system for construction worker safety." *Automation in Construction*, 157: 105158. https://doi.org/10.1016/j.autcon.2023.105158.

Di Pasquale, V., V. De Simone, S. Miranda, and S. Riemma. 2022. "Smart Operators: How Augmented and Virtual Technologies Are Affecting the Worker's Performance in Manufacturing Contexts." *Journal of Industrial Engineering and Management*, 15 (2): 233–235. https://doi.org/10.3926/jiem.3607.

Escalada-Hernandez, P., N. Soto-Ruiz, T. Ballesteros-Egüés, A. Larrayoz-Jiménez, and L. S. Martín-Rodríguez. 2024. “Usability and user expectations of a HoloLens-based augmented reality application for learning clinical technical skills.” *Virtual Reality*, 28 (2): 102. https://doi.org/10.1007/s10055-024-00984-3.

Eswaran, M., and M. V. A. Raju Bahubalendruni. 2023. “Augmented reality aided object mapping for worker assistance/training in an industrial assembly context: Exploration of affordance with existing guidance techniques.” *Computers & Industrial Engineering*, 109663. https://doi.org/10.1016/j.cie.2023.109663.

Hajirasouli, A., S. Banihashemi, R. Drogemuller, A. Fazeli, and S. R. Mohandes. 2022. “Augmented reality in design and construction: thematic analysis and conceptual frameworks.” *Construction Innovation*, 22 (3): 412–443. https://doi.org/10.1108/CI-01-2022-0007.

Lewis, J. R., and J. Sauro. 2018. “Item Benchmarks for the System.” 13 (3).

Lima, I. B., and W. Hwang. 2024. “Effects of Heuristic Type, User Interaction Level, and Evaluator’s Characteristics on Usability Metrics of Augmented Reality (AR) User Interfaces.” *International Journal of Human–Computer Interaction*, 40 (10): 2604–2621. https://doi.org/10.1080/10447318.2022.2163769.

lolambean. 2022. “MRTK2-Unity Developer Documentation - MRTK 2.” Accessed July 31, 2024. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05.

Maio, R., T. Araújo, B. Marques, P. Ramalho, A. Santos, D. Almeida, P. Dias, and B. S. Santos. 2024. “Pervasive Augmented Reality for Industry Scenarios: Lessons Learned and Insights Gained from a Comparative User Study.” *ACM International Conference on Interactive Media Experiences*, 246–255. Stockholm Sweden: ACM.

Moencks, M., E. Roth, T. Bohné, and P. O. Kristensson. 2022. “Augmented Workforce: contextual, cross-hierarchical enquiries on human-technology integration in industry.” *Computers & Industrial Engineering*, 165: 107822. https://doi.org/10.1016/j.cie.2021.107822.

Quandt, M., and M. Freitag. 2021. “A Systematic Review of User Acceptance in Industrial Augmented Reality.” *Front. Educ.*, 6. Frontiers. https://doi.org/10.3389/feduc.2021.700760.

Sauro, J., and J. R. Lewis. 2016. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Seeliger, A., R. P. Weibel, and S. Feuerriegel. 2022. “Context-Adaptive Visual Cues for Safe Navigation in Augmented Reality Using Machine Learning.” *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2022.2122114.

Stefanidi, Z., G. Margetis, S. Ntoa, and G. Papagiannakis. 2022. “Real-Time Adaptation of Context-Aware Intelligent User Interfaces, for Enhanced Situational Awareness.” *IEEE Access*, 10: 23367–23393. https://doi.org/10.1109/ACCESS.2022.3152743.

Tan, Y., W. Xu, S. Li, and K. Chen. 2022. “Augmented and Virtual Reality (AR/VR) for Education and Training in the AEC Industry: A Systematic Review of Research and Applications.” *Buildings*, 12 (10): 1529. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/buildings12101529.

Wolf, J. 2022. “Towards Advanced User Guidance and Context Awareness in Augmented Reality-guided Procedures.” Doctoral Thesis. ETH Zurich.

Yang, C.-K., Y.-H. Chen, T.-J. Chuang, K. Shankhwar, and S. Smith. 2020. “An augmented reality-based training system with a natural user interface for manual milling operations.” *Virtual Reality*, 24 (3): 527–539. https://doi.org/10.1007/s10055-019-00415-8.

Yuan, X., Q. Mei, and X. Li. 2023. “Integrating Real-Time Object Detection into an AR-Driven Task Assistance Prototype: An Approach Towards Reducing Specific Motions in Therbligs Theory.” *Proceedings e report*, P. Capone, V. Getuli, F. Pour Rahimian, N. Dawood, A. Bruttini, and T. Sorbi, eds. Firenze University Press.