Computers & Industrial Engineering 223 (2027) 112459

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Enhancing usability and attention alignment in AR-based construction assistance through formative multimodal evaluation

Xiang Yuan [a], Qipei Mei [b], Xinming Li [a,*]

[a] Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada
[b] Department of Civil & Environmental Engineering, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada

ARTICLE INFO

Keywords:
Augmented reality
Task assistance
Feedback-driven redesign
Multimodal evaluation
Eye tracking
Visual attention

ABSTRACT

Augmented reality (AR) is increasingly used to support construction and industrial procedural work through field-integrated guidance. However, technical feasibility does not by itself ensure usable and attention-compatible interaction in AR systems. Existing evaluations commonly operate at the outcome level, but the connection between user feedback, interface redesign, and multimodal validation remains insufficiently traceable. This study developed and evaluated a feedback-driven redesign workflow for a construction scaffolding AR task assistance system. System A was tested through task execution and semi-structured interviews, and the resulting feedback was coded, prioritized, and translated into interface modifications in System B. A two-cohort counterbalanced within-subject evaluation compared the two versions using behavioral, subjective, and AOI-based eye-tracking evidence. Results showed that System B improved task performance, usability, and visual attention organization. The convergence of these evidence sources within the counterbalanced design supports the interpretation that the observed changes reflect genuine shifts in attention-interface interaction. The study contributes a traceable formative multimodal workflow for construction AR task assistance development, applicable to other industrial task-assistance contexts involving procedural guidance, hands-busy interaction, and attention-intensive work.

1. Introduction

As the construction industry increasingly embraces human-centered automation, augmented reality (AR) has emerged as a promising tool for integrating digital information with real-world construction activities (Kolaei et al., 2022). By overlaying task-relevant information within users' field of view, AR enables in-situ access during task execution, with potential benefits for operational accuracy and safety (Tan et al., 2022). Head-mounted display (HMD) based AR further supports hands-free and context-aware guidance in dynamic construction environments (Hajirasouli et al., 2022; Maio et al., 2024). These capabilities position AR as a relevant form of digital work support in construction and related industrial settings, where task information delivery directly affects efficiency, safety, and system acceptance (Eswaran & Raju Bahubalendruni, 2023).

Despite this potential, reliable usability remains difficult to achieve in complex construction environments (Davila Delgado et al., 2020). Many AR applications prioritize spatial tracking, registration, and rendering performance, while giving comparatively less consideration to how interface design aligns with users' cognitive and attentional patterns (Smith et al., 2022). As a result, systems that perform well under controlled conditions may experience reduced usability on construction sites characterized by visual clutter, task interruptions, and competing attentional demands (Stefanidi et al., 2022; Yue et al., 2024). Prior studies indicate that misalignment among information density, interface layout, and attention allocation can contribute to information overload, increased cognitive effort, and misinterpretation of visual cues, potentially compromising task accuracy and safety (Beh et al., 2022). Attentional organization therefore emerges as a critical mediator between AR interface design and task-level usability outcomes, a concern shared across construction and broader industrial human-system work settings.

This tension between system design and human cognitive processes has also contributed to divergent findings (Ogunseiju et al., 2022; Chu et al., 2020). Some studies report substantial improvements in efficiency and perceived workload, while others highlight limited benefits or user frustration when AR tools disrupt natural task flow (Egger & Masood, 2020). These mixed outcomes show that technical innovation alone is

* Corresponding author.
E-mail address: xinming.li@ualberta.ca (X. Li).

https://doi.org/10.1016/j.cie.2026.112459
Received 20 February 2026; Received in revised form 28 June 2026; Accepted 12 October 2026
Available online 13 October 2026
0360-8352/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

insufficient to ensure effective AR adoption. The degree to which AR interface design supports users' attentional organization during task execution plays a critical role in shaping usability outcomes and interaction quality across industrial application domains (Vanneste et al., 2020; Ntoa, 2025).

Most AR usability studies assess system effectiveness through post-task questionnaires, completion time, and error rates (Lima & Hwang, 2024). These outcome-oriented measures capture overall task results and user evaluations but offer limited evidence about where interaction breakdowns occur or how visual attention is distributed during execution (Qu et al., 2024). In response, recent studies have advocated for process-oriented and multimodal evaluation approaches capable of capturing users' behavioral and attentional dynamics in real time (Zhang et al., 2025). Eye-tracking and related sensing techniques provide objective access to visual attention patterns and have been increasingly adopted for interface analysis (Dillen et al., 2024). However, their applications in construction AR have largely been limited to post-hoc assessment, with limited integration into iterative redesign processes (Abbas et al., 2020).

Existing research therefore leaves a methodological gap in the improvement of attention-interface alignment. While user feedback, performance metrics, and attentional data are increasingly available, they are seldom integrated through a structured process that identifies attention-related usability problems, translates them into interface modifications, and evaluates the resulting changes. This gap also concerns, more broadly, how human feedback and process-level sensing data can be translated into actionable improvements in industrial human-system work support.

To address this gap, this study adopts a formative and user-centered design perspective to examine how feedback-driven iteration can enhance both usability and attention alignment in AR system design. Qualitative feedback is used as the primary driver of iterative redesign, while multimodal data are leveraged as diagnostic evidence to interpret changes in user interaction and attentional organization. The study combines a traceable test-feedback-redesign process with a two-cohort counterbalanced within-subject evaluation, in which the Original Development Cohort completed System A before System B and provided formative feedback for redesign, while the independently recruited Validation Cohort completed the systems in reverse order to reduce fixed-order bias. All twelve participants were included in a combined paired comparison. Accordingly, this study is guided by the following research questions:

(1) **RQ1:** How do users' attentional behaviors and interaction patterns manifest when performing construction assembly tasks using an AR-based task assistance system?
(2) **RQ2:** How can qualitative user feedback be systematically translated into iterative AR interface refinements targeting attention-related usability issues?
(3) **RQ3:** How do multimodal evaluation results provide diagnostic evidence to interpret changes in usability and attentional organization following interface refinement?

The study makes three contributions. Theoretically, it advances a process-oriented understanding of AR usability by treating attention-interface alignment as a dynamic and diagnosable aspect of interaction. Methodologically, it presents a traceable test-feedback-redesign pathway that integrates qualitative feedback, interface refinement, and multimodal comparative evaluation. Practically, it demonstrates this framework in construction task assistance and offers transferable evaluation logic for industrial workflows involving procedural guidance and worker attention, including manufacturing assembly, aviation maintenance, and infrastructure inspection.

2. Literature review

2.1. AR task assistance systems in construction industry

AR has been applied across the construction lifecycle to support design coordination, on-site assembly, inspection, and safety management (Amin et al., 2023; Fazel & Adel, 2024). Related manufacturing, aviation, and healthcare studies also demonstrate AR's technical maturity and provide transferable insights for construction-oriented adaptation (Lu et al., 2020; Burian et al., 2023). In industrial assembly and maintenance contexts, AR task assistance has been widely used to deliver procedural guidance, reduce search effort, and support hands-free work execution (Yang et al., 2020; Yuan et al., 2024).

Construction research has expanded these capabilities across task-level application scenarios and system-level integration with digital infrastructures. Application-level studies have explored layout inspection, component installation, and safety management (Guray & Kismet, 2022; Einizinab et al., 2023), reporting improvements in operational outcomes such as enhanced defect identification (Chi et al., 2022). At the system level, AR integration with BIM and cloud platforms has supported maintenance tracking and spatially registered visualization (Khan et al., 2024), while multimodal systems have begun incorporating voice interaction and semantic image recognition for context-aware guidance (Chen, Hou, et al., 2024). Together, these studies establish the feasibility of AR task assistance, while also showing that operational value depends on how information presentation and interaction mechanisms fit the worker's task flow. Feasibility-focused reports, however, often describe system functions more fully than the human evidence used to shape them.

From a worker-oriented perspective, AR systems can be broadly categorized into real-time task assistance systems and skills training systems (Morales Méndez & del Cerro Velázquez, 2024). Task assistance systems deliver step-by-step virtual instructions at the point of work, often through HMDs to enable hands-free interaction in dynamic and tool-intensive environments (Moencks et al., 2022; Alkan & Basaga, 2023). They are commonly assessed using short-term indicators such as task completion time, error rates, and usability metrics (Daling & Schlittmeier, 2022). Skills training systems focus on knowledge acquisition and safety awareness through immersive simulations (Gong et al., 2024). They emphasize long-term outcomes including knowledge retention, safety incident reduction, and learning curves (Olbina & Glick, 2023; Wu et al., 2020). Given the present study's focus on real-time task assistance, this review concentrates on evaluation approaches that support iterative system refinement rather than long-term skill validation.

A central limitation of existing task-assistance research concerns the treatment of this interface-attention relationship. Many systems are developed from technical requirements and predefined task sequences and are subsequently examined through a single summative evaluation (Barbieri et al., 2024; Lee et al., 2020). User-centered approaches have also used interviews, task observations, and prototype testing to improve industrial AR systems (Ariansyah et al., 2022). These studies confirm that formative user involvement is already present in the field. However, they often provide limited detail on which interaction problems concerned attention allocation, how these problems were prioritized, and how the resulting interface modifications addressed them (Syed et al., 2023). Consequently, existing studies provide evidence of technical feasibility and user-centered refinement, while the process through which attention-related usability problems are diagnosed, modified, and subsequently evaluated remains insufficiently documented (Gualtieri et al., 2024).

2.2. Feedback-driven iteration and multimodal evaluation for AR task assistance

Feedback-driven iteration is central to improving system design

Table 1
Comparison of Redesign and Multimodal Evaluation Approaches in Industrial and Construction AR Task Assistance.

Study	Application context and AR assistance	Feedback-driven redesign and iterative validation	Evaluation evidence	Key findings	Remaining limitation
(Ariansyah et al., 2022)	Gearbox maintenance using video/animation and gesture/voice AR guidance	No; post-test feedback was not implemented	Time; errors; NASA-TLX; usability; eye tracking; physiology	Animation improved performance; video increased attention switching	Multimodal but summative; no redesign cycle
(Barbieri et al., 2024)	Industrial assembly inspection using AR annotation and information sharing	Partial; user feedback informed modifications and final validation	SEQ; TAM; interviews	UCD produced a well-accepted inspection tool	No formal prioritization, version comparison or attention measures
(Chu et al., 2020)	Dougong assembly using AR part identification, guidance, and step verification	Partial; preliminary observations informed functions and final testing	Time; errors; NASA-TLX; interviews	AR reduced assembly errors but increased completion time	Redesign process and issue-change mapping insufficiently reported
(Li et al., 2023)	Parcel sorting using 2D or spatial 3D AR interfaces	No; post-test comments did not drive redesign	Eye tracking; NASA-TLX; SUS; thematic feedback	3D interfaces reduced cognitive load and improved usability	Parallel interface comparison; no iterative validation
(Maio et al., 2024)	Industrial order picking and production monitoring using spatial AR guidance	Partial; stakeholder feedback informed successive prototypes	Time; errors; ratings; observation; comments	AR supported spatial awareness; tablets enabled faster data interaction	Incomplete feedback-change traceability; no eye tracking
(Vanneste et al., 2020)	Industrial assembly and quality control using projected work instructions	No; observations were not followed by redesign	Time; errors; help seeking; stress; workload ratings	AR improved quality and reduced assistance needs	No formative diagnosis, iteration or attention measures
(Wu et al., 2024)	Construction operations using AR hazard warnings and situation-awareness support	No; interviews informed interpretation only	Eye tracking; avoidance performance; interviews	AR accelerated hazard awareness and reduced avoidance failures	No feedback-driven modification or version validation
(Drouot et al., 2022)	Industrial assembly using HMD-AR work instructions	No; predefined conditions only	Time; errors; eye tracking; NASA-TLX; dual task	AR increased time and workload in some conditions	No usability scale, formative feedback or redesign
(Chu & Liu, 2023)	Human-robot collaborative assembly using visual and vibrotactile AR cues	No; interviews did not produce a revised version	Performance; gaze; usability; safety; trust; interviews	Multimodal cues improved interaction awareness and perceived safety	No formative involvement or feedback-driven iteration

because usability problems often emerge from the interaction between interface design, task context, and user attention during operation (Fernandes et al., 2022). Formative user feedback and task observation provide practical evidence for identifying interface bottlenecks and informing redesign decisions, and this iterative refinement logic has been recognized as essential for achieving sustained usability in complex work environments (Bernhaupt et al., 2016).

Existing AR usability evaluations largely rely on standardized self-report instruments such as the System Usability Scale (SUS) and related questionnaire- or model-based measures, often supplemented by aggregate task performance metrics including completion time and error rate (Lee et al., 2020; Lewis, 2018). Although these measures provide efficient summaries of perceived usability and task outcomes, they offer limited insight into context-specific interaction breakdowns and moment-to-moment attentional demands during task operations (Abbas et al., 2020). Task performance metrics capture observable operational outcomes but provide only indirect evidence of underlying cognitive processes and interaction strategies, especially in AR contexts where perception–action coupling is critical (Li, Zhou, et al., 2023; Akin et al., 2021). Moreover, usability is often assessed at a single time point, constraining the ability to interpret how iterative design changes influence interaction quality and attentional organization across successive system versions (Li, Zhou, et al., 2023). As a result, the diagnostic value of existing evaluation approaches for iterative AR system optimization remains limited (Guo et al., 2022).

These limitations have motivated growing interest in multimodal evaluation approaches that integrate behavioral, physiological, and attentional data to support process-oriented interpretation of user interaction (Rosch & Vogel-Walcutt, 2013). Such indicators have been shown to be more sensitive to usability variations in cognitively demanding environments than aggregated outcome metrics alone (Lee et al., 2018). Among these, eye-tracking has emerged as a particularly promising method for capturing visual attention allocation during interactive tasks (Liu & Li, 2024; Li, Duffy, et al., 2023). Compared with other physiological sensing techniques, it is more resilient to motion artifacts and can be deployed with minimal intrusion, making it suitable for construction scenarios involving frequent movement and tool use (Yang & Wang, 2023; Skaramagkas et al., 2023). By capturing how users allocate visual resources across interface elements, gaze data offer granular objective evidence of attention-interface alignment (Novák et al., 2024; Shin et al., 2024). For example, Comu et al., 2021 used fixation duration and time to first fixation to inform content optimization in immersive training. Shi et al., 2020 reported that gaze entropy strongly predicted performance differences across 2D, 3D, and immersive environments. Despite these advances, eye tracking remains underutilized as a diagnostic tool for iterative AR system refinement in construction contexts (Jeelani et al., 2019).

A review of related AR studies, summarized in Table 1, reveals two methodological patterns. A first group employed user-centered involvement to inform AR interface development, demonstrating that formative participation can generate practical design changes. However, these studies provide limited traceability from attention-related usability problems to prioritized modifications and comparative validation. A second group combined behavioral, subjective, and eye-tracking evidence to examine AR interaction under predefined conditions, showing that interface and interaction design influence performance, gaze behavior, and perceived usability. Qualitative findings in these studies generally supported result interpretation without driving an implemented redesign. Across both patterns, the direct connection between systematic qualitative feedback, traceable interface redesign, and attention-aware comparative evaluation remains insufficiently documented within a single study.

Collectively, these indicators form the attentional diagnostic foundation of the present study's evaluation framework. The present study differs from prior AR evaluations by integrating qualitative feedback analysis, redesign prioritization, implemented interface changes, task performance, usability assessment, and AOI-based eye-tracking metrics within one traceable test-feedback-redesign process, providing a replicable pathway for evidence-based AR system development in human-system work settings.

3. Methodology

This research adopted a structured framework for iterative AR-based task assistance systems (AR-TAS) development, integrating formative usability diagnosis, feedback-driven redesign, counterbalanced

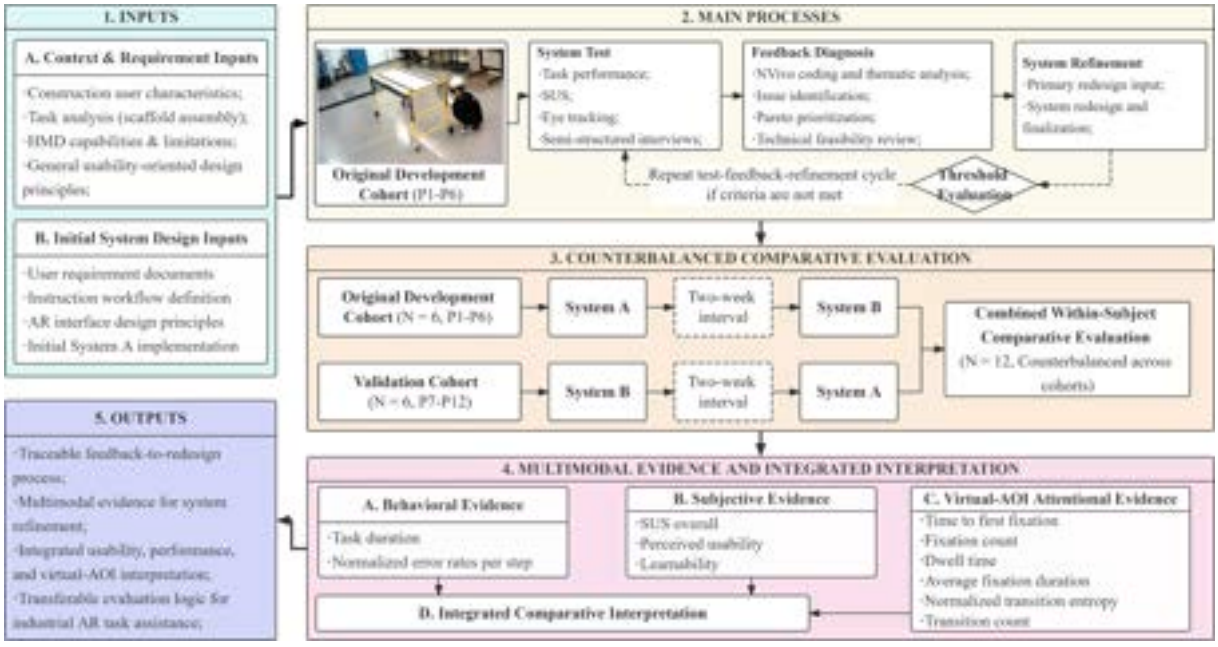

Fig. 1. Research Framework for Feedback-Driven Development and Multimodal Evaluation of AR-TAS.

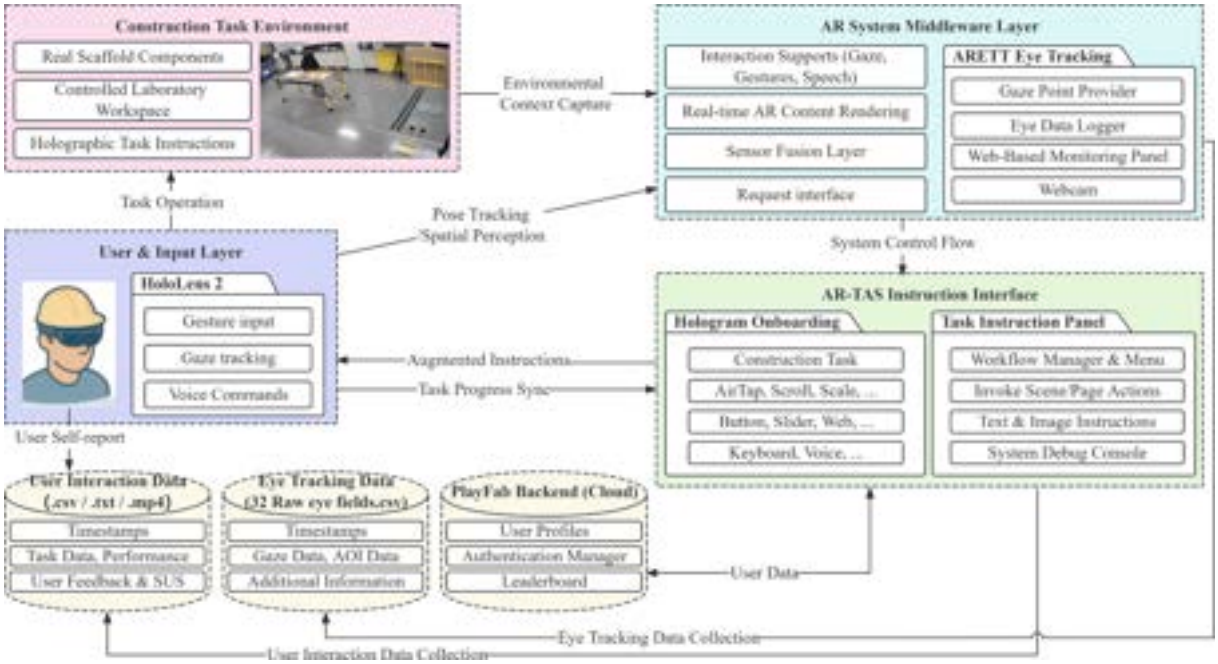

Fig. 2. System Architecture of Developed AR-TAS.

comparative evaluation, and multimodal evidence interpretation. As illustrated in Fig. 1, the framework comprises five components spanning system design inputs, iterative redesign, counterbalanced evaluation, multimodal interpretation, and replicable evaluation outputs. System A and System B operationalize a "test-feedback-redesign" loop that can support multiple iterative cycles. This study implemented one complete cycle, while the procedural structure allows further cycles if predefined criteria are not met. Qualitative feedback from semi-structured interviews constituted the primary redesign input, while behavioral, usability, and eye-tracking data provided diagnostic evidence for comparative evaluation.

To avoid open-ended refinement, a threshold-based iteration mechanism was applied. A SUS score exceeding 85, corresponding to the "best imaginable" usability category (Bangor et al., 2009), was adopted as the stopping criterion, as refinement beyond this level yields diminishing returns in perceived usability gains. Had this criterion not been met, an additional cycle would have been initiated. Task performance and AOI-based eye-tracking metrics provided complementary evidence for assessing whether usability gains were accompanied by changes in interaction efficiency and visual behavior (Chen, Zhao, et al., 2024).

3.1. AR task assistance system (AR-TAS) overview

The developed AR-TAS delivered real-time contextual task instructions to novice users performing construction assembly tasks through a HMD interface, supporting multimodal interaction through

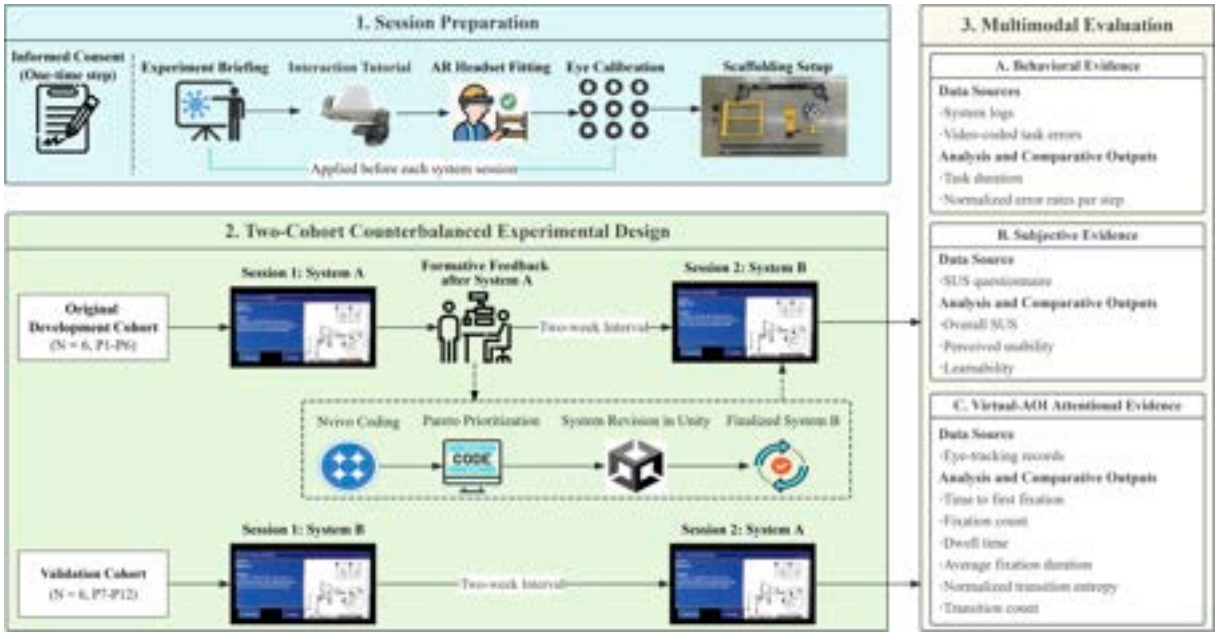

Fig. 3. Counterbalanced Experimental Procedure and Multimodal Comparative Evaluation.

gaze, gesture, and voice to facilitate hands-free task progression. As shown in Fig. 2, the system architecture consists of five functional layers: construction task environment, user and input layer, AR system middleware, AR-TAS instruction interface, and data collection and backend services. The middleware layer integrated the Augmented Reality Eye Tracking Toolkit (ARETT) for continuous gaze and AOI recording, using a collider-based dynamic AOI mechanism to maintain spatial accuracy despite user movement or object relocation (Kapp et al., 2021). The AR-TAS instruction interface included a hologram onboarding module for initial familiarization and a task panel presenting procedural steps, interactive elements, and system prompts. All data streams were logged through AR-TAS records, HoloLens sensors, and manual collection.

3.2. Experimental procedure

A scaffolding assembly task was selected as a representative construction activity requiring spatial reasoning, tool handling, and procedural compliance. Experiments were conducted at the University of Alberta within a 5×6 m workspace, with lighting and layout held constant across sessions. The experimental setup utilized a Metaltech multipurpose baker scaffold (working height: 1.83 m) with standard braces, platform, and locking mechanisms. The AR-TAS ran on a Trimble XR10 with HoloLens 2 (52° FOV), recording gaze behavior at approximately 30 Hz with an average spatial accuracy of 1.5° (Leharanger et al., 2023). Eye-tracking calibration was verified before each session, and data segments with tracking loss or motion artifacts were removed during preprocessing. Throughout each session, gaze, AOI-hit, and calibration status were monitored via the ARETT web interface, allowing any tracking or calibration loss to be corrected between assembly steps.

This study employed a two-cohort within-subject counterbalanced design. Seventeen participants with no or limited construction experience were recruited under approved ethics protocol (No. Pro00132621). Two participants withdrew voluntarily and three were excluded due to persistent eye-tracking calibration failure, resulting in a final analytical sample of $n = 12$ comprising eight males and four females, with a mean age of 28.6 years.

The first six participants formed the Original Development Cohort and completed System A followed by System B, with a two-week interval separating the two sessions. This sequence reflects the formative design logic, as System B was developed based on usability limitations identified in System A. The six subsequently recruited participants formed the Validation Cohort and completed the procedure in reverse order with the same two-week interval. This cohort had no prior exposure to System A and did not contribute to the redesign of System B, providing an independent reverse-order comparison to reduce fixed-order bias. No participants reported cybersickness or discomfort during the sessions.

The full experimental procedure is summarized in Fig. 3. Informed consent was obtained once before participation, and each session began with a briefing, interaction tutorial, headset fitting, and eye-tracking calibration before the scaffolding assembly task. Both systems covered the same seven physical assembly stages shown in Fig. 4. Task performance data, SUS responses, eye-tracking records, and video recordings were collected under both system conditions. Semi-structured interviews conducted after System A with the Original Development Cohort provided qualitative input for NVivo coding, Pareto prioritization, and System B refinement.

3.3. Usability issues prioritization

To identify and prioritize usability issues in System A, semi-structured interviews were combined with an evidence-based

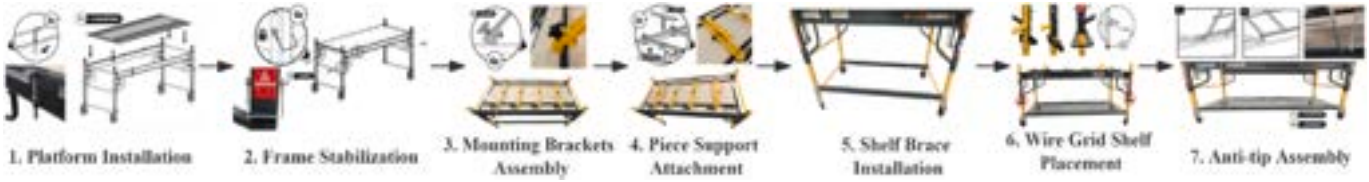

Fig. 4. Seven Major Stages of the Scaffolding Assembly Task.

Table 2
Semi-Structured Interview Agenda for Identifying Usability Issues.

Focus Area	Sample Questions and Probes
Overall Satisfaction	1. How satisfied are you with the system overall? Please rate on a scale of 1–10. 2. How do you think the system performs in terms of improving your work efficiency and performance?
Improvement Suggestions and Expectations	3. Which features do you think are missing or could be further enhanced? 4. Which steps do you think can be simplified or optimized? 5. Do you think there needs to be more guidance or prompts? Which specific guidance or prompts would be helpful? 6. Have you encountered any system crashes or errors that affected your operation? 7. Do you think the system needs to be more user-friendly? In which specific aspects?
Interview Conclusion	8. Do you have any other comments or suggestions that we haven't covered?

prioritization process. Post-task interviews were designed to capture both anticipated and emergent usability concerns through predefined guiding questions and open-ended probes, as shown in Table 2 (Kallio et al., 2016). Interview transcripts were anonymized and analyzed in NVivo 14 using an analytical coding approach informed by grounded-theory techniques. Open coding identified recurring usability-related concepts, which were refined through axial and selective coding into higher-level themes relevant to interface design and interaction flow. Two researchers coded independently and resolved discrepancies by discussion to enhance analytical consistency. The analysis targeted pragmatic sufficiency rather than theoretical saturation, and field notes from a pilot study were re-analyzed to verify the recurrence of key themes. Each theme was mapped to an interface design category, with the complete mapping documented in Table 5.

Feedback frequencies were normalized per participant and visualized using a heatmap to support redesign prioritization. A Pareto analysis was then applied to identify a subset of issues accounting for the majority of reported usability concerns. Given the small sample size, Pareto analysis served as a formative prioritization aid rather than a population-level test, with participant-level distributions inspected alongside the group mean to assess whether rankings reflected broadly shared concerns. Each candidate issue was further assessed for implementation feasibility on HoloLens 2, excluding recommendations constrained by hardware limitations or outside the defined system scope. Final redesign selection considered normalized feedback salience, cross-participant recurrence, and implementation feasibility, maintaining traceability between user feedback and implemented modifications.

3.4. Multimodal data collection and analysis

As illustrated in Fig. 5, the multimodal evaluation integrated three complementary evidence domains to examine system differences from outcome, perception, and interaction-process perspectives. Behavioral performance metrics quantified operational efficiency and reliability through task duration and normalized error rates per step. Subjective evidence captured participants' perceived usability and supported the identification of system-specific usability concerns. Eye-tracking evidence provided process-level insight into how users allocated and shifted visual attention across virtual instructional elements. Their

Table 3
Behavioral Error Categories Used in Task Performance Analysis.

Error Type	Interpretation
Search/Navigation Error	Inefficient or failed attempts to locate task-relevant components;
Assembly/Attachment Error	Incorrect operations during the physical joining of components;
Instruction Interpretation Error	Misinterpretation or cognitive misunderstanding of textual or visual guidance;
Safety-Critical Error	Actions violating safety instructions;
AR Interaction Error	User-system interaction breakdowns such as gesture misrecognition;

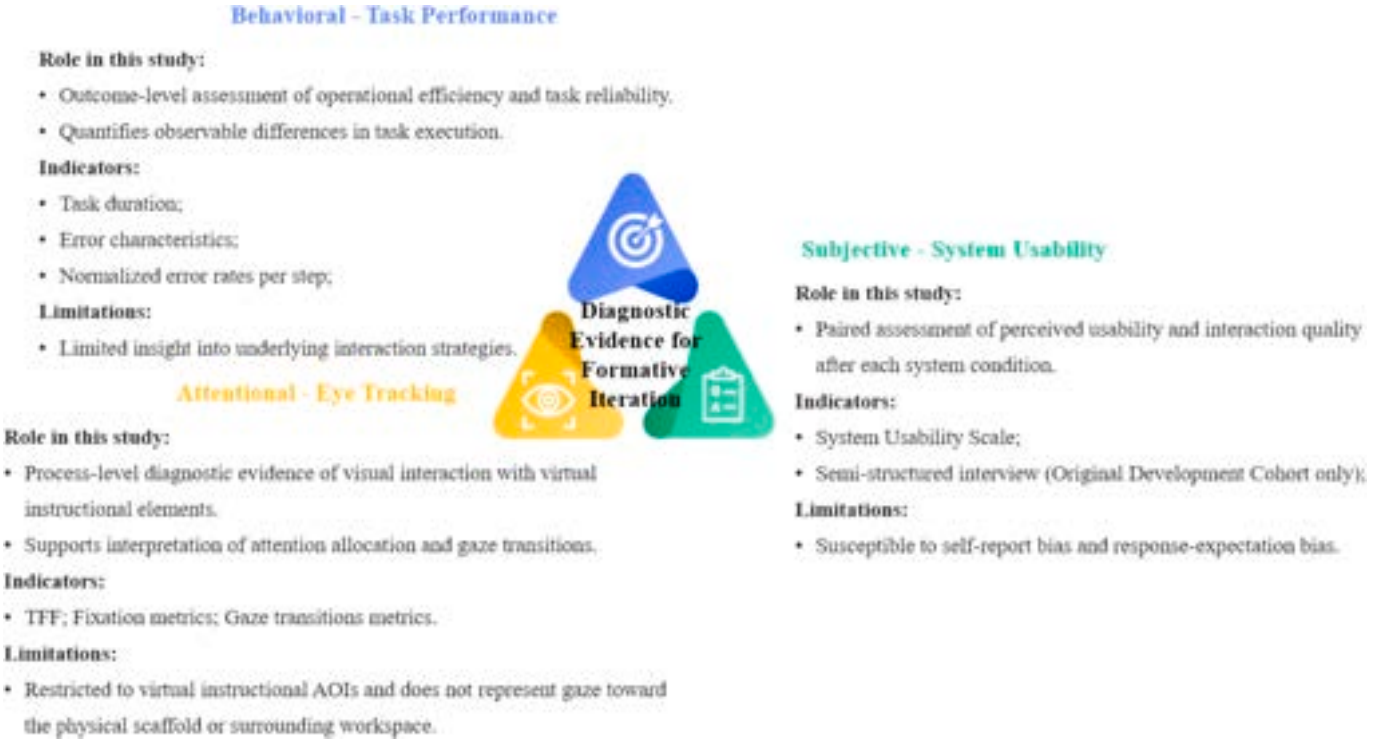

Fig. 5. Behavioral, Subjective, and Attentional Evidence Domains Used for Multimodal Evaluation.

Table 4
Core Eye-Tracking Fields Used in AOI-Based Visual Analysis.

Data category	Data fields	Description
Gaze validity and timing	eyeDataTimestamp, eyeDataRelativeTimestamp, isCalibrationValid, gazeHasValue	Timestamps and validity indicators used to order gaze samples, calculate durations, and retain calibrated gaze data.
Gaze direction and intersection	gazeOrigin_x/y/z, gazeDirection_x/y/z, gazePointHit, gazePoint_x/y/z	Records the three-dimensional gaze-ray origin, direction, and intersection position within the virtual scene.
Virtual AOI data	gazePointAOIHit, gazePointAOI_x/y/z, gazePointAOI_name	AOI intersection status, position, and semantic label used to associate gaze samples with virtual Text and Image AOIs. Valid samples without either AOI label were retained as aggregated outside-AOI gaze.

integration examined whether differences in task performance and perceived usability were accompanied by corresponding changes in visual behavior.

3.4.1. Operational efficiency metrics

Task performance was evaluated through task duration and error characteristics. Task duration was measured from task start to completion using timestamped system logs, supplemented by manual observation for error coding. To characterize errors, a five-category schema was inductively derived from video-coded task recordings, field observations, and user-reported root causes to support diagnostic interpretation of task breakdowns. As summarized in Table 3, each error instance was labeled by type, frequency, and duration, with coder consensus ensuring reliability. Interaction delays related to AR usability were retained, while hardware malfunctions were excluded to maintain internal validity. Because System B contained fewer scripted steps after workflow restructuring, error outcomes were analyzed using normalized errors per step. The normalized error rate was calculated as $Normalized\ Error\ Rate_c = \frac{Error\ Count_c}{Total\ Number\ of\ Steps}$.

Comparative analysis employed paired t-tests, with Wilcoxon signed-rank tests conducted as sensitivity analyses to confirm the robustness of results given the sample size. Task duration was treated as a single predefined primary outcome and reported with an unadjusted p value. Benjamini-Hochberg corrections were applied to normalized error rate comparisons across the five error categories as a single family of tests. Statistical results were interpreted together with effect sizes, 95% confidence intervals, and cross-modal convergence.

3.4.2. Perceived usability assessment

System usability was evaluated using the System Usability Scale (SUS), a standardized and widely validated instrument for interactive systems (Escalada-Hernandez et al., 2024). Participants completed the SUS immediately after each system session, yielding scores on a 0–100 scale computed using the standard procedure (Li, Zheng, et al., 2023). Following Lewis and Sauro's refinement, items 4 and 10 were grouped into the "learnability" subscale, while the remaining items contributed to the "perceived usability" dimension (Lewis & Sauro, 2009). SUS outcomes were compared using paired t-tests, with Wilcoxon signed-rank tests conducted as sensitivity analyses.

3.4.3. AOI-based visual analysis

AOIs were defined as semantically meaningful virtual regions, including Text AOIs (instructional texts and interface controls) and Image AOIs (graphical task illustrations). They were embedded into Unity 3D and dynamically tracked through ARETT collider-based detection. Gaze intersections were recorded using gazePointAOI output, linking each valid sample to the corresponding virtual interface element and task step. Before data collection, physical QR-code markers at known positions were used to verify spatial alignment between the AR scene and the recorded coordinate systems, confirming the spatial-reference accuracy of the gaze and AOI recordings. A custom Python-based preprocessing pipeline extended ARETT's default toolkit for segmented AOI-specific analysis. The pipeline incorporated logic from ARETT and Pupil Labs with parameters from related studies (Llanes-Jurado et al., 2020), and raw gaze samples were synchronized with task-step timestamps, AOI collider logs, and system interaction logs. The core data fields are summarized in Table 4.

Fixation detection followed a dispersion-based algorithm (I-DT), which classifies a fixation event when gaze samples remain within a spatial dispersion threshold over a minimum duration window. A dispersion threshold of 1.6° and a minimum duration of 250 ms were applied, consistent with parameters adopted in comparable AR eye-tracking studies using HoloLens-based platforms (Xu et al., 2026). Blink events were identified by requiring consecutive data points marked as invalid (gazeHasValue = False) within a single blink window, reducing false positives from intermittent signal dropout. Saccade events with velocities exceeding 150°/s were classified as noise and excluded from further analysis. Signal artifacts from tracking loss or head motion were removed, resulting in an average data loss of 2.4% with no systematic concentration in any task phase, indicating that missing data were unlikely to introduce directional bias.

To validate the reliability of the custom preprocessing pipeline, extracted features were benchmarked against results from a published HoloLens 2-ARETT study using the same dataset and identical model configurations (Bektaş et al., 2024). Classification accuracy remained comparable across both implementations, with differences of no more than 1.3 percentage points across all tested models, confirming the methodological consistency of the custom pipeline.

Six AOI-based features were extracted at the task-step level and categorized into static attention-allocation measures and dynamic transition measures. Valid gaze samples outside the virtual Text and Image AOIs were retained as an aggregated outside-AOI category and were not further analyzed as a distinct attentional domain. Features were selected based on their proven relevance in modeling task-oriented visual behaviors in immersive environments and interface evaluation (Yuan et al., 2024). Time to First Fixation (TFF) measures the latency from step onset to the first stable fixation on a target AOI (Yoo et al., 2023), computed as:

$$TFF_{AOI} = t_{first_fixation,target} - t_{step_start}$$

Shorter TFF indicates earlier orientation toward the target instructional element and reduced initial visual search (Fu et al., 2024; S. Wu et al., 2024). Fixation Count represents the total number of fixations directed at a given AOI within a task step:

$$Fixation\ Count = \sum_{i=1}^{n} \mathbb{I}(AOI_i = target)$$

Fewer fixations indicate reduced repeated visual search and more direct engagement with the target interface element (Upasani et al., 2024). Dwell Time captures the cumulative duration of all fixations on a target AOI (Noghabaei et al., 2021):

$$Dwell\ Time = \sum_{i=1}^{n} t_i \times \mathbb{I}(AOI_i = target)$$

Prolonged dwell time reflects greater attentional resource allocation to an interface element, which in AR instruction contexts may indicate effortful information processing (Comu et al., 2021). Average Fixation Duration is the mean duration per fixation on a target AOI:

System Improvement Suggestions	P1	P2	P3	P4	P5	P6	Avg.	R.
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	×

Fig. 6. Summary of Improvement Suggestions and Proportion in Original Development Cohort (%).

$$Average\ Duration = \frac{\sum_{i=1}^{n} t_i \times \mathbb{I}(AOI_i = target)}{Fixation\ Count}$$

Extended average fixation duration is commonly interpreted as greater information-processing demand, depending on the task and interface context (Jeelani et al., 2019; Devlin et al., 2022). Normalized Transition Entropy quantifies the distributional diversity of gaze shifts between AOIs, normalized to a 0–1 scale using the maximum possible entropy given the number of AOIs (Shiferaw et al., 2019):

$$H_{norm} = \frac{-\sum_{i,j} p_{i,j} \log_2 p_{i,j}}{\log_2 (N_{AOIs})}$$

Higher entropy indicates more evenly distributed transitions across AOIs, while lower entropy reflects more repetitive or concentrated switching patterns (Cui et al., 2024; Ma & Yao, 2025). Because entropy valence depends on task context, it was interpreted jointly with transition count and task performance outcomes and not as an isolated indicator. Transition Count represents the number of successive transitions

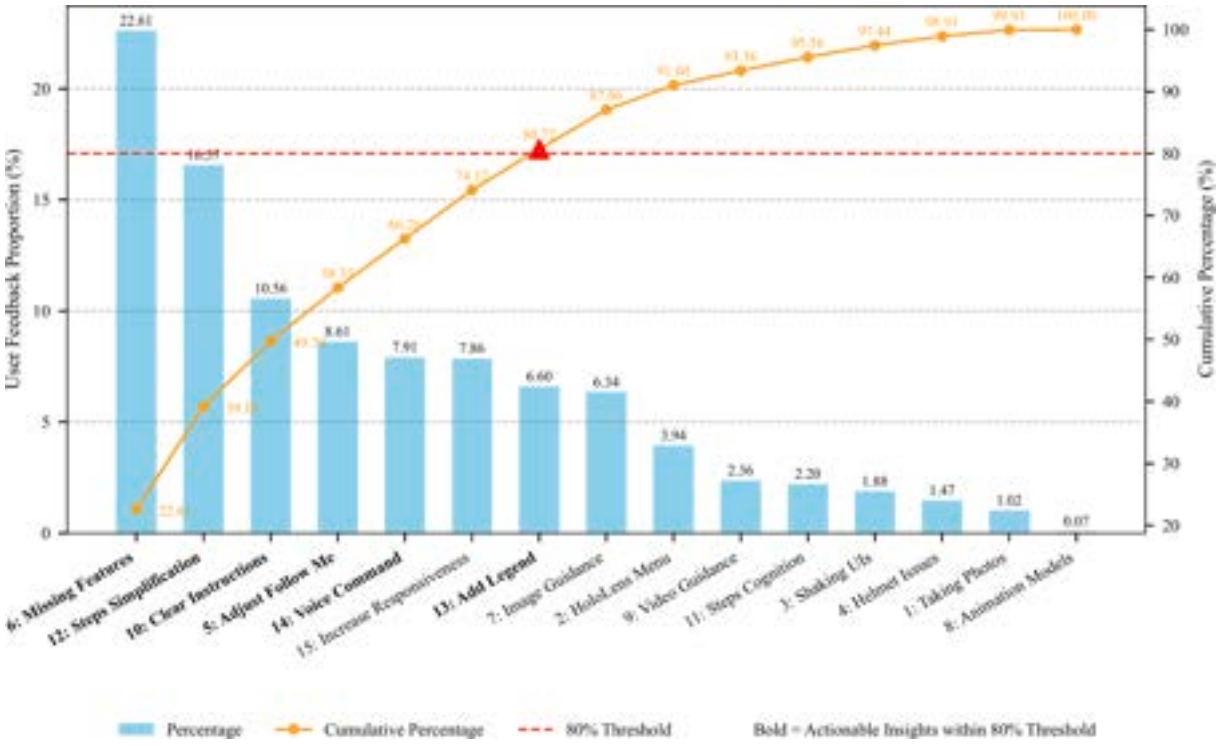

Fig. 7. Pareto Prioritization of Usability Issues Reported by the Original Development Cohort. ***Note:** Red dashed line indicates the 80% cumulative threshold. Bolded features indicate software- and content-level issues selected for redesign consideration within the current formative cohort. Item 15 was excluded due to hardware-level implementation constraints.

Table 5
Feedback-to-Redesign Mapping from System A Usability Issues to System B Modifications.

User-Identified Usability Issues in System A	Corresponding Modifications in System B
6. Missing Features – Searching for Objects (22.61%) Participants reported difficulty locating and confirming task-relevant components, leading to repeated visual searches. Full-scale object recognition was deferred as it required hardware-level sensing beyond the current scope (Yuan et al., 2023).	 Labeled visual overlays and categorized tooltips added to instruction panel to support component identification and reduce search ambiguity.
12. Steps Simplification (16.57%) Instructions were reported as overly detailed and repetitive, disrupting task flow.	The workflow script was restructured from 21 to 14 scripted steps by consolidating redundant operations. Step grouping and visual progress indicators were introduced.
10. Clear Instructions (10.56%) Participants reported difficulty interpreting ambiguous or inconsistent instructional content.	 Textual descriptions were revised for consistency and conciseness, and visual-textual alignment was improved to reduce interpretive effort.
5. Enhanced Features – Adjust Follow Me Function (8.61%) Participants reported that the Follow Me panel occasionally obstructed task-relevant views, reducing visibility during physically demanding assembly steps.	 Proximity thresholds and panel repositioning logic were recalibrated to keep instructions accessible while reducing visual obstruction.
14. Add Voice Command (7.91%) Participants requested more hands-free interaction options during tool-intensive operations, as manual navigation interrupted physical task flow.	 Basic voice commands were implemented via Unity and the native HoloLens speech API to support page navigation and selection.
13. Add a Legend (6.60%) Participants reported difficulty interpreting interface symbols and navigation cues without contextual reference.	 Instructional legends were introduced at task onset to explain key symbols and navigation commands.

in the retained AOI-entry sequence after consecutive identical raw AOI labels were collapsed (Frank et al., 2023):

$$Transition\ Count\ =\ \sum_{i=2}^{n}\mathbb{I}(AOI_i \neq AOI_{i-1})$$

Higher transition counts reflect greater attentional switching demands induced by interface structure, and have been associated with increased cognitive switching costs in immersive task environments (Peng et al., 2020). The combination of entropy and transition count captures both the diversity and frequency of gaze navigation across instructional elements.

A linear mixed-effects model (LMM) was employed using L-BFGS optimization algorithm. System version and task steps were designated as fixed effects, while participant identity was incorporated as a random effect to account for individual variability. LMM was selected for its capacity to handle nested hierarchical data structures and missing gaze observations (Li, Zheng, et al., 2023). Benjamini-Hochberg correction was applied across the ten AOI-based system-effect tests reported in Table 8, treating them as one family of parallel attentional outcomes.

4. Results

4.1. Usability diagnosis and optimized modifications

4.1.1. Identified usability problems

The analysis yielded six thematic categories and twenty-two sub-nodes, of which seven reflected general impressions and provided contextual insight but were not treated as direct redesign inputs. The remaining fifteen items represented actionable usability suggestions and are summarized in FIg. 6. Feedback frequencies were normalized per participant, with the group mean indicating overall prominence and the heatmap highlighting areas of concentrated concern. Three suggestions dominated participants' feedback: object search support (Item 6, 22.61%), step simplification (Item 12, 16.57%), and clarity of instructions (Item 10, 10.56%).

Notable inter-participant heterogeneity was also observed, reflecting differing interaction strategies and breakdown points despite a shared task structure. In particular, P2 allocated 40.53% of their feedback to object search difficulties (Item 6), which substantially elevated its group-level average of 22.61%. P4 similarly allocated 37.30% to Item 6. The prominence of this issue in the group ranking reflects strong and consistent concern among a subset of participants, and its selection for

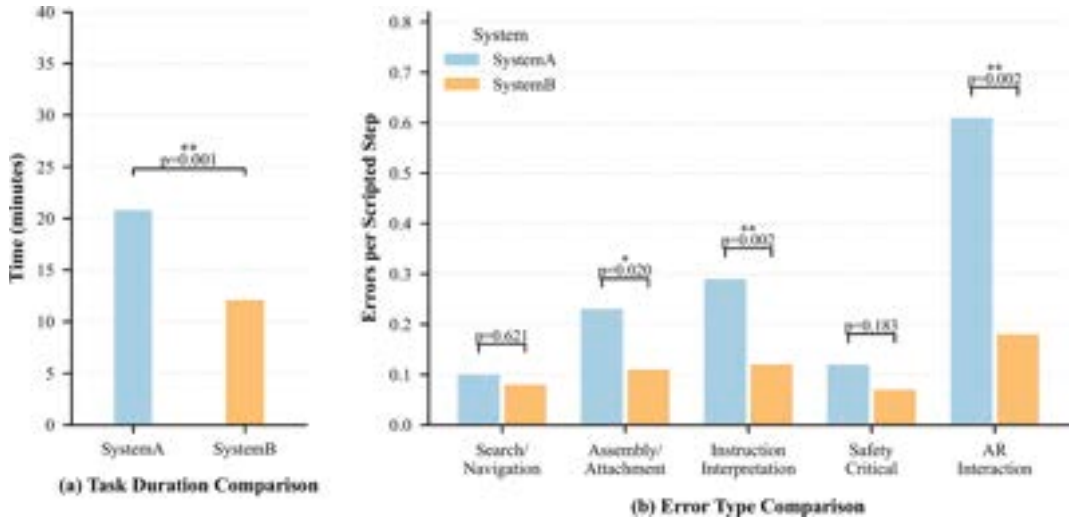

Fig. 8. Comparison of Task Duration and Normalized Error Rates Between System A and System B.

Table 6
Normalized Error Rates Per Step by Error Category.

Error category	System A Mean ± SD	System B Mean ± SD	t	BH-adjusted p	Cohen's dz	95% CI
Search/Navigation Errors	0.10 ± 0.12	0.08 ± 0.09	−0.51	0.621	−0.15	[−0.085, 0.053]
Assembly/Attachment Errors	0.23 ± 0.11	0.11 ± 0.10	−2.99	0.020*	−0.86	[−0.203, −0.031]
Instruction Interpretation Errors	0.29 ± 0.11	0.12 ± 0.11	−4.59	0.002**	−1.33	[−0.258, −0.091]
Safety Critical Errors	0.12 ± 0.06	0.07 ± 0.08	−1.56	0.183	−0.45	[−0.105, 0.018]
AR Interaction Errors	0.61 ± 0.32	0.18 ± 0.20	−4.73	0.002**	−1.37	[−0.628, −0.229]

Note: BH-adjusted p values were calculated across the five normalized error-rate categories. Statistical significance is denoted by ***p < 0.001, **p < 0.01, and *p < 0.05.

Table 7
Comparison of SUS Scores Between System A and System B.

Metrics	System A Mean ± SD	System B Mean ± SD	t	p	Cohen's dz	95% CI
SUS Overall Score	48.33 ± 18.38	86.25 ± 9.01	6.23	< 0.001***	1.80	[24.53, 51.31]
Usability Score	39.79 ± 15.72	70.21 ± 8.01	5.98	< 0.001***	1.73	[19.23, 41.60]
Learnability Score	8.54 ± 3.91	16.04 ± 2.49	4.78	< 0.001***	1.38	[4.05, 10.95]

redesign was further supported by cross-participant recurrence across five of the six participants. P6 emphasized voice command support (Item 14). Such variation highlights the value of participant-level inspection in formative diagnosis. Each suggestion was reviewed for feasibility on HoloLens 2, with hardware-constrained items labeled non-revisable (R. = ×) and software-level items retained for iteration (R. = ✓).

Based on normalized feedback proportions, a Pareto analysis was conducted to identify a minimal subset of issues accounting for the majority of reported usability concerns (Fig. 7). Suggestions were ordered by average salience, and cumulative contributions were calculated to form the Pareto curve. Given the small sample size, these rankings served as a practical prioritization aid and are not intended to represent a statistically stable distribution in a broader population. The cumulative proportion reached the 80% threshold at Item 13, with the top seven items jointly accounting for 80.7% of total feedback. These items served as candidate issues for redesign. Although Item 15 fell within the top 80% range, it was excluded due to hardware-level implementation constraints.

4.1.2. Design modifications based on user feedback

Table 5 documents the mapping from prioritized usability issues to implemented modifications in System B, providing a traceable account of how qualitative findings were translated into design actions. Six issues that were frequently reported and technically revisable were addressed.

4.2. Evaluation of optimization effects

4.2.1. Task performance comparison

Fig. 8 summarizes task duration and normalized error rates per step for System A and System B. As shown in Fig. 8 (a), participants completed the scaffolding assembly task in a shorter time when using System B (12.10 ± 6.17 min) compared to System A (20.82 ± 9.82 min), $t = -4.37$, $p = 0.001$. This reduction suggests more efficient task progression with the revised system. To further contextualize performance differences, task errors were categorized into five types. Because System B contained fewer steps than System A, 14 compared with 21, error counts were normalized by scripted step count to account for differences in workflow length. Normalized error rates are illustrated in Fig. 8 (b).

As summarized in Table 6, System B showed lower normalized error rates in all five categories. After Benjamini-Hochberg correction, instruction interpretation errors were significantly reduced ($t = -4.59$, $p = 0.002$), as were AR interaction errors ($t = -4.73$, $p = 0.002$). Assembly/attachment errors also showed a significant reduction ($t = -2.99$, $p = 0.020$). Reductions in search/navigation and safety–critical errors were observed in the same direction but did not reach

Table 8
Linear Mixed-Effects Model Results for AOI-Based Eye-Tracking Metrics.

Features	AOI	β B-A	SE	BH-adjusted p	95% CI
TFF (s)	Text	−1.325	0.723	0.074	[−2.741, 0.091]
	Image	−3.957	1.069	< 0.001***	[−6.052, −1.863]
Fixation Count	Text	−3.267	1.520	0.039*	[−6.247, −0.288]
	Image	−4.767	1.394	< 0.001***	[−7.499, −2.035]
Dwell Time (s)	Text	−6.192	1.199	< 0.001***	[−8.541, −3.843]
	Image	−5.999	1.209	< 0.001***	[−8.368, −3.630]
Average Duration (s)	Text	−0.494	0.075	< 0.001***	[−0.640, −0.348]
	Image	0.020	0.049	0.686	[−0.076, 0.115]
Normalized Transition Entropy	–	0.156	0.023	< 0.001***	[0.110, 0.202]
Transition Count	–	−24.199	4.893	< 0.001***	[−33.789, −14.610]

Note: BH-adjusted p values were calculated across the ten reported AOI-based LMM comparisons. Statistical significance is denoted by ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$.

statistical significance after correction. These findings indicate that the core performance improvements in System B were robust to the difference in step count between the two systems.

4.2.2. System usability enhancement

Table 7 summarizes the comparative results for overall SUS scores and subcomponents. System B achieved a substantially higher overall SUS score than System A, increasing from 48.33 ± 18.38 to 86.25 ± 9.01 ($t = 6.23$, $p < 0.001$). SUS comparisons were treated as independent primary outcomes and were not included in the Benjamini-Hochberg corrections. Further analysis of SUS subcomponents indicated that the perceived usability subscale increased from 39.79 ± 15.72 to 70.21 ± 8.01 ($t = 5.98$, $p < 0.001$), and the learnability subscale increased from 8.54 ± 3.91 to 16.04 ± 2.49 ($t = 4.78$, $p < 0.001$). Both subscales showed statistically significant improvements, indicating that participants perceived System B as more usable and easier to learn.

4.2.3. Visual attention restructuring

A step-level analysis of gaze behavior was conducted using six AOI-based eye-tracking features. Fig. 9 shows the dynamic changes of these metrics across all mapped steps. System A and System B exhibit distinct visual attention patterns, particularly during cognitively demanding operations including Steps 2–2, 3–2, and 6–2. In System A, these steps showed obvious peaks in TFF, fixation count, and dwell time, indicating extended visual search and prolonged engagement with instructional elements. System B demonstrated comparatively stable patterns across the same steps, suggesting more consistent attention distribution across task phases.

Table 8 summarizes the LMM results for AOI-based metrics. Model coefficients represent System B relative to System A, with negative values indicating lower values in System B.

Relative to System A, System B showed a non-significant decreasing trend in TFF for the Text AOI with $p = 0.074$, and a significant reduction in TFF for the Image AOI ($\beta = -3.957$, $p < 0.001$), indicating faster initial orientation toward image-based instructional information. Fixation count was significantly lower for both AOI types (Text: $\beta = -3.267$, $p = 0.039$; Image: $\beta = -4.767$, $p < 0.001$), as was dwell time (Text: $\beta = -6.192$, $p < 0.001$; Image: $\beta = -5.999$, $p < 0.001$), indicating less cumulative visual effort on instructional elements. Average fixation duration decreased significantly for the Text AOI ($\beta = -0.494$ s, $p < 0.001$), while no significant difference was observed for the Image AOI ($\beta = 0.020$ s, $p = 0.686$), suggesting that redesign effects were more pronounced for textual than for image-based instructional processing.

Dynamic transition metrics further highlighted differences in visual transition behavior. Normalized transition entropy was higher for System B ($\beta = 0.156$, $p < 0.001$), while transition count decreased ($\beta = -24.199$, $p < 0.001$). Because entropy valence depends on task context, this result was interpreted jointly with transition count and task performance. The co-occurrence of higher entropy with fewer transitions suggests a less repetitive and more evenly distributed gaze transition pattern, consistent with the reduced error rates and shorter task duration observed above. The AOI findings provide process-level diagnostic evidence of attention-interface alignment, supported by their convergence with behavioral and subjective results.

5. Discussion

5.1. Implications for task efficiency and usability

Following system iteration, the overall SUS score increased from 48.33 ± 18.38 in System A to 86.25 ± 9.01 in System B, indicating that System B was perceived as substantially more usable and consistent by the participating users. The reduction in score dispersion from 18.38 to 9.01 further suggests that the improved interface supported a more uniform interaction experience across participants, reflecting fewer individual difficulties related to navigation, interpretation, or interaction breakdowns.

The SUS subscale results further showed significant improvements in both perceived usability and learnability. Perceived usability increased from 39.79 ± 15.72 to 70.21 ± 8.01, while learnability increased from 8.54 ± 3.91 to 16.04 ± 2.49. These results suggest that the redesign improved interface clarity and interaction flow while supporting more accessible initial learning. Such an emphasis is consistent with usability perspectives in immersive systems that prioritize intuitive affordances and reduced interaction friction (Li, Zheng, et al., 2023).

Beyond subjective perceptions, iterative refinement was also associated with observable changes in task execution. Task duration decreased from 20.82 ± 9.82 min in System A to 12.10 ± 6.17 min in System B. Normalized error rates further indicated significant reductions in assembly/attachment errors ($t = -2.99$, $p = 0.020$), instruction interpretation errors ($t = -4.59$, $p = 0.002$), and AR interaction errors ($t = -4.73$, $p = 0.002$) in System B.

Eye-tracking results provide additional process-level insight. The reduction in average fixation duration on text AOIs ($\beta = -0.494$, $p < 0.001$) indicates that users extracted instructional information with less prolonged visual engagement. The convergence of subjective ratings, behavioral outcomes, and eye-tracking changes provides corroboration that the observed improvements reflect genuine shifts in attention-interface interaction, supporting the interpretation that system refinement enabled more stable and efficient task engagement.

The counterbalanced cohort structure helps address the possibility that observed gains were mainly driven by task familiarity, procedural memory, or reduced novelty effects. The Validation Cohort completed the procedure in reverse order and had no involvement in System A feedback or System B redesign. The convergence of results across cohorts reduces the likelihood that observed improvements were an artifact of a fixed A-to-B sequence, although residual familiarity effects cannot be fully excluded given that all participants completed the scaffolding task twice. Because system order was assigned by cohort, this cross-cohort comparison provides a descriptive robustness check and does not constitute a formal test of a cohort-by-system interaction.

5.2. Implications for visual attention allocation

5.2.1. Effects on static AOI engagement

Compared to System A, users interacting with System B showed a

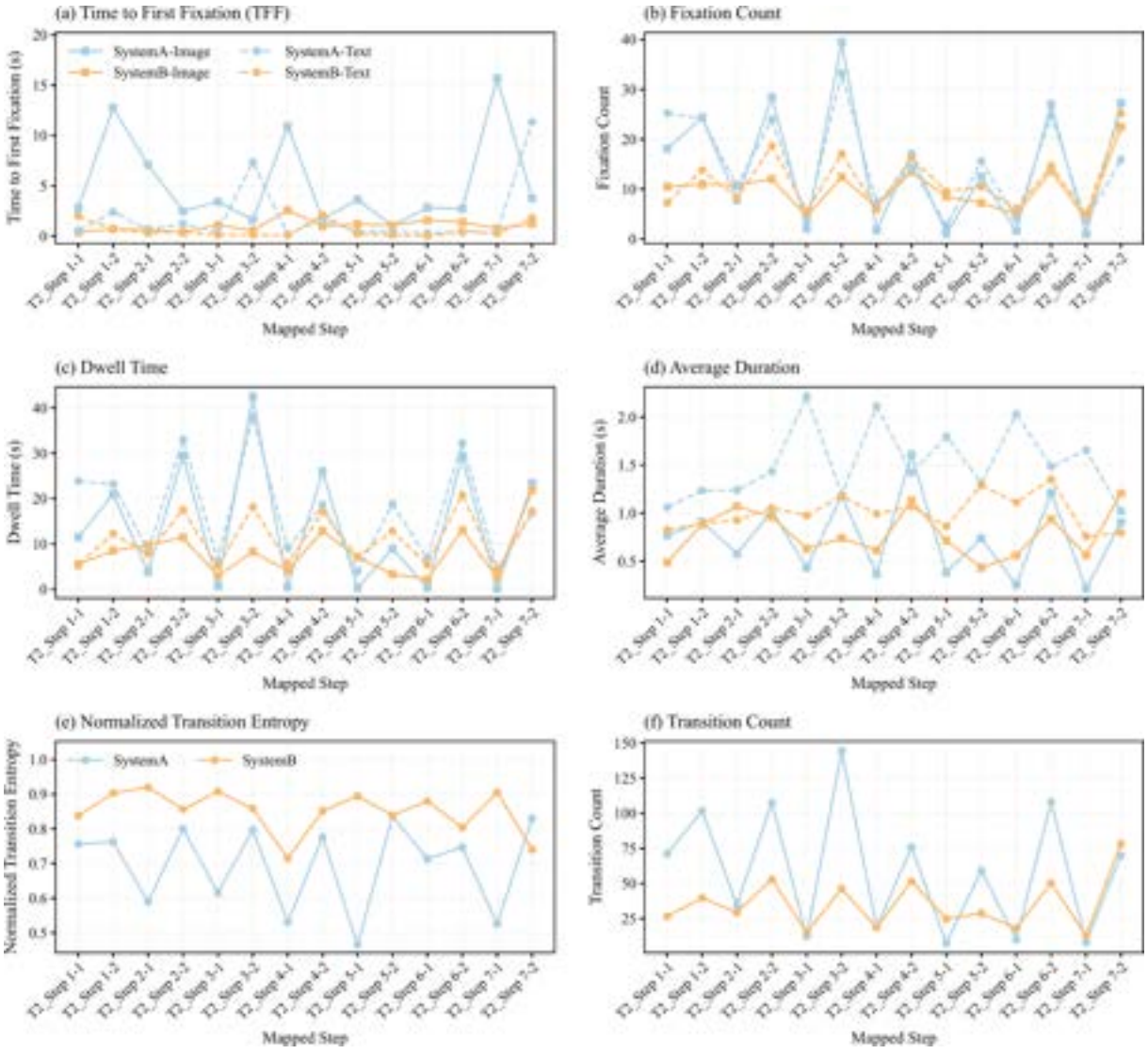

Fig. 9. Step-wise Comparison of Mean AOI-Based Eye-Tracking Metrics for System A and B.

significant reduction in TFF for image AOIs ($\beta = -3.957$, $p < 0.001$), while the reduction in TFF for text AOIs was a non-significant decreasing trend ($\beta = -1.325$, $p = 0.074$). Concurrent reductions in fixation counts and dwell time further indicate that users revisited interface elements less frequently and for shorter cumulative durations.

For textual AOIs in particular, System B presented substantial reductions in dwell time, fixation count, and average fixation duration. This pattern suggests that textual instructions could be interpreted with fewer re-fixations and less prolonged visual engagement. Such changes are consistent with the design changes introduced in System B, including semantic refinement and hierarchical restructuring of textual content, which aimed to reduce ambiguity and understanding effort during instruction interpretation. Prior eye-tracking research has similarly associated longer fixation durations with greater cognitive processing complexity and more difficult information extraction (Ozcelik et al., 2009).

In contrast, image AOIs showed a different attentional profile. Although System B users oriented toward images more quickly and fixated on them fewer times, the average fixation duration on image AOIs remained statistically unchanged ($\beta = 0.020$, $p = 0.686$). Interface refinements thus influenced the efficiency of locating and revisiting visual elements without affecting per-fixation processing duration for images. Qualitative feedback supports this interpretation, with reports such as "*I only read the text carefully when I encounter confusion; otherwise, I rely more on intuitive images.*" This division of roles highlights the function of images as high-density information carriers for spatial reasoning, with text serving a complementary explanatory role.

Across both AOI types, the reduction in fixation counts and dwell time suggests fewer redundant visual searches and more targeted information acquisition. System B showed shorter task duration and lower normalized error rates, indicating that reduced visual engagement was associated with more reliable task execution.

From a design perspective, these findings suggest that attention-aware refinement of virtual instructional content can support attentional economy in AR-assisted construction workflows, reducing unnecessary visual switching while preserving stable engagement with essential elements, which is particularly relevant where excessive scanning may undermine operational reliability.

5.2.2. *Effects on dynamic gaze transitions*

In addition to changes in static visual engagement, differences between system iterations were also reflected in dynamic gaze transition patterns. Under System B, normalized transition entropy was higher ($\beta = 0.156$, $p < 0.001$), while the total number of gaze transitions was lower ($\beta = -24.199$, $p < 0.001$), characterizing a shift in how users' gaze transitioned across virtual interface elements. Higher entropy with

fewer transitions indicates more evenly distributed and less repetitive gaze switching, an interpretation strengthened by convergence with shorter task duration and lower error rates.

Several interface refinements in System B offer reasonable explanations for these patterns. In System A, users were required to manually locate static AR panels, which may have constrained gaze to fixed visual anchors and limited the diversity of AOI transitions. System B's adaptive panel accompanied users' spatial workflows, potentially reducing repeated visual searches and physical repositioning. Similar principles have been reported in context-adaptive AR navigation systems, where selectively presented visual cues supported more autonomous interaction without imposing rigid viewing sequences (Seeliger et al., 2022).

The integration of voice command functionality introduced an alternative interaction channel that may have redistributed visual attention demands. Qualitative feedback supports this interpretation, with participants noting *"voice controls enabled more seamless, hands-free workflows and eliminated the need for repetitive visual confirmation"*. Eye-tracking metrics provide diagnostic evidence for how cross-modal interaction changes were reflected in gaze transition behavior, complementing the static AOI findings discussed above.

These findings suggest that interface refinement supported a more structured organization of gaze transitions, which is particularly relevant in construction AR contexts where excessive visual switching can interfere with task continuity and safety.

5.3. Limitations and future directions

This study has several limitations. First, the analytical sample of twelve novice users limits generalizability, and further validation in larger and more diverse populations remains necessary (Khan et al., 2024). Second, the within-subject design introduces residual risks of task familiarity, motor learning, and demand characteristics. The counterbalanced design and multimodal convergence partially mitigate these threats but do not eliminate them. Third, gaze outside the virtual Text and Image AOIs was retained as an aggregated category without differentiation into physical sub-regions, limiting the attention analysis to the virtual instructional interface. Fourth, the controlled laboratory setting limits transferability to operational construction sites, where conditions are more dynamic and subject to interruption.

These limitations point to three directions for future research. Field evaluation with larger and more diverse participant groups, including transitioning workers and experienced professionals, is a necessary next step. Differentiating outside-AOI gaze into physical sub-regions through markerless tracking or semantic scene understanding would provide a more complete account of attention allocation across virtual and physical task elements. Modeling the step-level association between gaze-transition measures and task errors would clarify whether the observed transition patterns are directly linked to task performance.

6. Conclusion

In this study, System A revealed several attention-related interaction problems during AR-assisted construction assembly. Users experienced delayed orientation toward relevant instructional elements, repeated visual checking, prolonged engagement with instructional content, frequent switching between information sources, and interruptions associated with interface operation. These patterns indicated that the original interface did not consistently support efficient coordination between digital guidance and physical task execution. Qualitative feedback was subsequently coded into usability themes, prioritized according to its frequency and recurrence across participants, and screened for technical feasibility. This process translated diagnosed problems into specific modifications to the information content, visual hierarchy, interface layout, task sequence, and interaction functions of System B. Following redesign, behavioral evidence indicated more efficient and reliable task execution, SUS results showed improved perceived usability and learnability, and AOI-based eye-tracking measures reflected more direct and efficient engagement with instructional elements. Considered within the counterbalanced design, the convergence of behavioral, subjective, and attentional evidence supports the interpretation that the observed changes reflect genuine shifts in attention-interface interaction, although residual learning and order effects cannot be completely excluded.

The central contribution of this study is a traceable formative multimodal workflow connecting qualitative diagnosis, systematic prioritization, feasibility screening, implemented redesign, and comparative evaluation. For construction AR-TAS development, this workflow provides a structured means of identifying attention-related usability problems and determining whether interface modifications are accompanied by improvements in user perception, operational performance, and visual interaction. The approach may also inform related industrial task-assistance contexts involving procedural guidance, hands-busy interaction, and attention-intensive work. The present findings demonstrate one complete redesign cycle under controlled laboratory conditions and do not establish universal effectiveness across tasks, users, or repeated iterations. Future research should evaluate the workflow in field environments with larger samples of construction and industrial professionals, longer periods of system use, and more diverse task conditions.

CRediT authorship contribution statement

Xiang Yuan: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qipei Mei:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Xinming Li:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by the Natural Sciences and Engineering Research Council of Canada through the Advance Alliance Grants [ALLRP 561120 - 20 and ALLRP 601811 - 24] with Alberta Innovates.

Data availability

Data will be made available on request.

References

Abbas, A., Seo, J., & Kim, M. (2020). Impact of mobile augmented reality system on cognitive behavior and performance during rebar inspection tasks. *Journal of Computing in Civil Engineering, 34*(6), Article 04020050. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000931

Akin, S., Ergun, O., Surer, E., & Gursel Dino, I. (2021). An immersive performative architectural design tool with daylighting simulations: A building information modeling-based approach. *Engineering, Construction and Architectural Management, 28*(4), 1319–1344. https://doi.org/10.1108/ECAM-07-2020-0562

Alkan, I. B., & Basaga, H. B. (2023). Augmented reality technologies in construction project assembly phases. *Automation in Construction, 156*. https://doi.org/10.1016/j.autcon.2023.105107

Amin, K., Mills, G., & Wilson, D. (2023). Key functions in BIM-based AR platforms. *Automation in Construction, 150*, Article 104816. https://doi.org/10.1016/j.autcon.2023.104816

Ariansyah, D., Erkoyuncu, J. A., Eimontaite, I., Johnson, T., Oostveen, A.-M., Fletcher, S., & Sharples, S. (2022). A head mounted augmented reality design practice for maintenance assembly: Toward meeting perceptual and cognitive needs of AR users. *Applied Ergonomics, 98*, Article 103597. https://doi.org/10.1016/j.apergo.2021.103597

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. https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf.

Barbieri, L., Marino, E., Bruno, F., Muzzupappa, M., & Colacino, B. (2024). User-centered design of an augmented reality inspection tool for industry 4.0 operators. *International Journal on Interactive Design and Manufacturing (IJIDeM)*.. https://doi.org/10.1007/s12008-024-01931-x

Beh, H. J., Rashidi, A., Talei, A., & Lee, Y. S. (2022). Developing engineering students' capabilities through game-based virtual reality technology for building utility inspection. *Engineering, Construction and Architectural Management, 29*(7), 2854–2877. https://doi.org/10.1108/ECAM-02-2021-0174.

Bektaş, K., Strecker, J., Mayer, S., & Garcia, K. (2024). Gaze-enabled activity recognition for augmented reality feedback. *Computers & Graphics, 119*, Article 103909. https://doi.org/10.1016/j.cag.2024.103909

Bernhaupt, R., Palanque, P., Manciet, F., & Martinie, C. (2016). User-test results injection into task-based design process for the assessment and improvement of both usability and user experience. In C. Bogdan, J. Gulliksen, S. Sauer, P. Forbrig, M. Winckler, C. Johnson, P. Palanque, R. Bernhaupt, & F. Kis (Eds.), *Human-Centered and Error-Resilient Systems Development* (pp. 56–72). Springer International Publishing. https://doi.org/10.1007/978-3-319-44902-9_5.

Burian, B. K., Ebnali, M., Robertson, J. M., Musson, D., Pozner, C. N., Doyle, T., Smink, D. S., Miccile, C., Paladugu, P., Atamna, B., Lipsitz, S., Yule, S., & Dias, R. D. (2023). Using extended reality (XR) for medical training and real-time clinical support during deep space missions. *Applied Ergonomics, 106*, Article 103902. https://doi.org/10.1016/j.apergo.2022.103902

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

Chen, L., Zhao, H., Shi, C., Wu, Y., Yu, X., Ren, W., Zhang, Z., & Shi, X. (2024). Enhancing multi-modal perception and interaction: An augmented reality visualization system for complex decision making. *Systems, 12*(1), Article 1. https://doi.org/10.3390/systems12010007

Chi, H.-L., Kim, M.-K., Liu, K.-Z., Thedja, J. P. P., Seo, J., & Lee, D.-E. (2022). Rebar inspection integrating augmented reality and laser scanning. *Automation in Construction, 136*, Article 104183. https://doi.org/10.1016/j.autcon.2022.104183

Chu, C.-H., Liao, C.-J., & Lin, S.-C. (2020). Comparing augmented reality-assisted assembly functions—A case study on dougong structure. *Applied Sciences, 10*(10), Article 10. https://doi.org/10.3390/app10103383

Chu, C.-H., & Liu, Y.-L. (2023). Augmented reality user interface design and experimental evaluation for human-robot collaborative assembly. *Journal of Manufacturing Systems, 68*, 313–324. https://doi.org/10.1016/j.jmsy.2023.04.007

Comu, S., Kazar, G., & Marwa, Z. (2021). Evaluating the attitudes of different trainee groups towards eye tracking enhanced safety training methods. *Advanced Engineering Informatics, 49*, Article 101353. https://doi.org/10.1016/j.aei.2021.101353

Cui, Z., Sato, T., Jackson, A., Jayarathna, S., Itoh, M., & Yamani, Y. (2024). Gaze transition entropy as a measure of attention allocation in a dynamic workspace involving automation. *Scientific Reports, 14*(1), 23405. https://doi.org/10.1038/s41598-024-74244-4

Daling, L. M., & Schlittmeier, S. J. (2022). Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: A scoping review. *Human Factors, 00187208221105135*. https://doi.org/10.1177/00187208221105135

Davila Delgado, J. M., Oyedele, L., Beach, T., & Demian, P. (2020). Augmented and virtual reality in construction: Drivers and limitations for industry adoption. *Journal of Construction Engineering and Management, 146*(7), Article 04020079. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001844

Devlin, S. P., Brown, N. L., Drollinger, S., Sibley, C., Alami, J., & Riggs, S. L. (2022). Scan-based eye tracking measures are predictive of workload transition performance. *Applied Ergonomics, 105*, Article 103829. https://doi.org/10.1016/j.apergo.2022.103829

Dillen, A., Omidi, M., Ghaffari, F., Romain, O., Vanderborght, B., Roelands, B., Nowé, A., & De Pauw, K. (2024). User evaluation of a shared robot control system combining BCI and eye tracking in a portable augmented reality user interface. *Sensors, 24*(16), Article 16. https://doi.org/10.3390/s24165253

Drouot, M., Le Bigot, N., Bricard, E., de Bougrenet, J.-L., & Nourrit, V. (2022). Augmented reality on industrial assembly line: Impact on effectiveness and mental workload. *Applied Ergonomics, 103*, Article 103793. https://doi.org/10.1016/j.apergo.2022.103793

Egger, J., & Masood, T. (2020). Augmented reality in support of intelligent manufacturing – A systematic literature review. *Computers & Industrial Engineering, 140*, Article 106195. https://doi.org/10.1016/j.cie.2019.106195

Einizinab, S., Khoshelham, K., Winter, S., Christopher, P., Fang, Y., Windholz, E., Radanovic, M., & Hu, S. (2023). Enabling technologies for remote and virtual inspection of building work. *Automation in Construction, 156*, Article 105096. https://doi.org/10.1016/j.autcon.2023.105096

Escalada-Hernandez, P., Soto-Ruiz, N., Ballesteros-Egüés, T., Larrayoz-Jiménez, A., & Martín-Rodríguez, L. S. (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., & Raju Bahubalendruni, M. V. A. (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

Fazel, A., & Adel, A. (2024). Enhancing construction accuracy, productivity, and safety with augmented reality for timber fastening. *Automation in Construction, 166*, Article 105596. https://doi.org/10.1016/j.autcon.2024.105596

Fernandes, Í., Rocha, S., Portela, C., Braz Junior, G., Almeida, J., Silva, A., Viana, D., Rabelo, J., Paiva, A., & Rivero, L. (2022). Defining an A/B testing process for usability and user experience evaluation through the analysis of the results of a literature review. In M. Kurosu, S. Yamamoto, H. Mori, M. M. Soares, E. Rosenzweig, A. Marcus, P.-L.-P. Rau, D. Harris, & W.-.-C. Li (Eds.), *HCI International 2022—Late Breaking Papers. Design, User Experience and Interaction* (pp. 204–213). Springer International Publishing. https://doi.org/10.1007/978-3-031-17615-9_14.

Frank, W., Mühl, K., Rosner, A., & Baumann, M. (2023). Advancing knowledge on situation comprehension in dynamic traffic situations by studying eye movements to empty spatial locations. *Human Factors, 65*(8), 1674–1688. https://doi.org/10.1177/00187208211063693

Fu, H., Xia, Z., Tan, Y., & Guo, X. (2024). Influence of cues on the safety hazard recognition of construction workers during safety training: evidence from an eye-tracking experiment. *Journal of Civil Engineering Education, 150*(1), Article 04023009. https://doi.org/10.1061/JCEECD.EIENG-1882

Gong, P., Lu, Y., Lovreglio, R., Lv, X., & Chi, Z. (2024). Applications and effectiveness of augmented reality in safety training: A systematic literature review and meta-analysis. *Safety Science, 178*, Article 106624. https://doi.org/10.1016/j.ssci.2024.106624

Gualtieri, L., Öhler, M., Revolti, A., & Dallasega, P. (2024). A visual management and augmented-reality-based training module for the enhancement of short and long-term procedural knowledge retention in complex machinery setup. *Computers & Industrial Engineering, 196*, Article 110478. https://doi.org/10.1016/j.cie.2024.110478

Guo, F., Chen, J., Li, M., Lyu, W., & Zhang, J. (2022). Effects of visual complexity on user search behavior and satisfaction: An eye-tracking study of mobile news apps. *Universal Access in the Information Society, 21*(4), 795–808. https://doi.org/10.1007/s10209-021-00815-1

Guray, T. S., & Kismet, B. (2022). VR and AR in construction management research: Bibliometric and descriptive analyses. *Smart and Sustainable Built Environment, 12*(3), 635–659. https://doi.org/10.1108/SASBE-01-2022-0015.

Hajirasouli, A., Banihashemi, S., Drogemuller, R., Fazeli, A., & Mohandes, S. R. (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.

Jeelani, I., Albert, A., Han, K., & Azevedo, R. (2019). Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. *Journal of Construction Engineering and Management, 145*(1), Article 04018115. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001589

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kapp, S., Barz, M., Mukhametov, S., Sonntag, D., & Kuhn, J. (2021). ARETT: Augmented reality eye tracking toolkit for head mounted displays. *Sensors, 21*(6), Article 6. https://doi.org/10.3390/s21062234

Khan, S., Panuwatwanich, K., & Usanavasin, S. (2024). Integrating building information modeling with augmented reality: Application and empirical assessment in building facility management. *Engineering, Construction and Architectural Management, 31*(7), 2809–2828. https://doi.org/10.1108/ECAM-12-2021-1146

Kolaei, A. Z., Hedayati, E., Khanzadi, M., & Amiri, G. G. (2022). Challenges and opportunities of augmented reality during the construction phase. *Automation in Construction, 143*, Article 104586. https://doi.org/10.1016/j.autcon.2022.104586

Lee, B. C., Chung, K., & Kim, S.-H. (2018). Interruption cost evaluation by cognitive workload and task performance in interruption coordination modes for human–computer interaction tasks. *Applied Sciences, 8*(10), Article 10. https://doi.org/10.3390/app8101780

Lee, J. G., Seo, J., Abbas, A., & Choi, M. (2020). End-users' augmented reality utilization for architectural design review. *Applied Sciences, 10*(15), Article 15. https://doi.org/10.3390/app10155363

Leharanger, M., Rodriguez Martinez, E. A., Balédent, O., & Vandromme, L. (2023). Familiarization with mixed reality for individuals with autism spectrum disorder: An eye tracking study. *Sensors, 23*(14), Article 14. https://doi.org/10.3390/s23146304

Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human-Computer Interaction, 34*(12), 1148–1156. https://doi.org/10.1080/10447318.2017.1418805

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In M. Kurosu (Ed.), *Human Centered Design* (Vol. 5619, pp. 94–103). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-02806-9_12.

Li, D., Zhou, H., Zhou, S., Huang, G., Ma, X., Zhao, Y., Wang, W., & Ng, S. T. (2023). An eye-tracking-based approach to evaluate the usability of government portal websites in pilot smart cities. *Engineering, Construction and Architectural Management*. https://doi.org/10.1108/ECAM-09-2023-0932. ahead-of-print(ahead-of-print).

Li, S., Duffy, M. C., Lajoie, S. P., Zheng, J., & Lachapelle, K. (2023). Using eye tracking to examine expert-novice differences during simulated surgical training: A case study. *Computers in Human Behavior, 144*, Article 107720. https://doi.org/10.1016/j.chb.2023.107720

Li, X., Zheng, C., Pan, Z., Huang, Z., Niu, Y., Wang, P., & Geng, W. (2023). Comparative study on 2D and 3D user interface for eliminating cognitive loads in augmented reality repetitive tasks. *International Journal of Human-Computer Interaction*, 1–17. https://doi.org/10.1080/10447318.2023.2276526

Lima, I. B., & Hwang, W. (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

Liu, D., & Li, C. (2024). Construction hazard-recognition training for generation Z students: A comparison study based on eye-movement data. *Engineering, Construction and Architectural Management.*. https://doi.org/10.1108/ECAM-09-2023-0882

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality. *Sensors, 20*(17), Article 17. https://doi.org/10.3390/s20174956

Lu, S., Sanchez Perdomo, Y. P., Jiang, X., & Zheng, B. (2020). Integrating eye-tracking to augmented reality system for surgical training. *Journal of Medical Systems, 44*(11), 192. https://doi.org/10.1007/s10916-020-01656-w

Ma, Y., & Yao, H. (2025). Research on understandability & cognitive load equilibrium of human-machine interface under time pressure. *Displays, 89*, Article 102891. https://doi.org/10.1016/j.displa.2024.102891

Maio, R., Araújo, T., Marques, B., Ramalho, P., Santos, A., Almeida, D., Dias, P., & Santos, B. S. (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*. https://doi.org/10.1145/3639701.3656316

Moencks, M., Roth, E., Bohné, T., & Kristensson, P. O. (2022). Augmented workforce: Contextual, cross-hierarchical enquiries on human-technology integration in industry. *Computers & Industrial Engineering, 165*, Article 107822. https://doi.org/10.1016/j.cie.2021.107822

Morales Méndez, G., & del Cerro Velázquez, F. (2024). Augmented reality in industry 4.0 assistance and training areas: A systematic literature review and bibliometric analysis. *Electronics, 13*(6), Article 6. https://doi.org/10.3390/electronics13061147

Noghabaei, M., Han, K., & Albert, A. (2021). Feasibility study to identify brain activity and eye-tracking features for assessing hazard recognition using consumer-grade wearables in an immersive virtual environment. *Journal of Construction Engineering and Management, 147*(9), Article 04021104. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002130

Novák, J.Š., Masner, J., Benda, P., Šimek, P., & Merunka, V. (2024). Eye tracking, usability, and user experience: A systematic review. *International Journal of Human-Computer Interaction, 40*(17), 4484–4500. https://doi.org/10.1080/10447318.2023.2221600

Ntoa, S. (2025). Usability and user experience evaluation in intelligent environments: A review and reappraisal. *International Journal of Human-Computer Interaction, 41*(5), 2829–2858. https://doi.org/10.1080/10447318.2024.2394724

Ogunseiju, O. R., Gonsalves, N., Akanmu, A. A., Bairaktarova, D., Bowman, D. A., & Jazizadeh, F. (2022). Mixed reality environment for learning sensing technology applications in Construction: A usability study. *Advanced Engineering Informatics, 53*, Article 101637. https://doi.org/10.1016/j.aei.2022.101637

Olbina, S., & Glick, S. (2023). Using integrated hands-on and virtual reality (VR) or augmented reality (AR) approaches in construction management education. *International Journal of Construction Education and Research, 19*(3), 341–360. https://doi.org/10.1080/15578771.2022.2115173

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*(2), 445–453. https://doi.org/10.1016/j.compedu.2009.03.002

Peng, J. S., Wang, C. W., Fu, R., & Yuan, W. (2020). Extraction of parameters for lane change intention based on driver's gaze transfer characteristics. *Safety Science, 126*, Article 104647. https://doi.org/10.1016/j.ssci.2020.104647

Qu, J., Bu, L., Zhao, L., & Wang, Y. (2024). A training and assessment system for human-computer interaction combining fNIRS and eye-tracking data. *Advanced Engineering Informatics, 62*, Article 102765. https://doi.org/10.1016/j.aei.2024.102765

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work, 15*(3), 313–327. https://doi.org/10.1007/s10111-012-0234-7

Seeliger, A., Weibel, R. P., & Feuerriegel, S. (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

Shi, Y., Du, J., & Zhu, Q. (2020). The impact of engineering information format on task performance: Gaze scanning pattern analysis. *Advanced Engineering Informatics, 46*, Article 101167. https://doi.org/10.1016/j.aei.2020.101167

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews, 96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

Shin, H., Oh, C., Kim, N. Y., Choi, H., Kim, B., & Ji, Y. G. (2024). Evaluating and eliciting design requirements for an improved user experience in live-streaming commerce interfaces. *Computers in Human Behavior, 150*, Article 107990. https://doi.org/10.1016/j.chb.2023.107990

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering, 16*, 260–277. https://doi.org/10.1109/RBME.2021.3066072.

Smith, A., Duff, C., Sarlo, R., & Gabbard, J. L. (2022). Wearable augmented reality interface design for bridge inspection. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 497–501). https://doi.org/10.1109/VRW55335.2022.00111

Stefanidi, Z., Margetis, G., Ntoa, S., & Papagiannakis, G. (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.

Syed, T. A., Siddiqui, M. S., Abdullah, H. B., Jan, S., Namoun, A., Alzahrani, A., Nadeem, A., & Alkhodre, A. B. (2023). In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns. *Sensors, 23*(1), Article 1. https://doi.org/10.3390/s23010146

Tan, Y., Xu, W., Li, S., & Chen, K. (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), Article 10. https://doi.org/10.3390/buildings12101529

Upasani, S., Srinivasan, D., Zhu, Q., Du, J., & Leonessa, A. (2024). Eye-tracking in physical human–robot interaction: Mental workload and performance prediction. *Human Factors, 66*(8), 2104–2119. https://doi.org/10.1177/00187208231204704

Vanneste, P., Huang, Y., Park, J. Y., Cornillie, F., Decloedt, B., & Van den Noortgate, W. (2020). Cognitive support for assembly operations by means of augmented reality: An exploratory study. *International Journal of Human-Computer Studies, 143*, Article 102480. https://doi.org/10.1016/j.ijhcs.2020.102480

Wu, W., Sandoval, A., Gunji, V., Ayer, S. K., London, J., Perry, L., Patil, K., & Smith, K. (2020). Comparing traditional and mixed reality-facilitated apprenticeship learning in a wood-frame construction lab. *Journal of Construction Engineering and Management, 146*(12), Article 04020139. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001945

Wu, S., Chen, H., Hou, L., Zhang, G. (Kevin), & Li, C.-Q. (2024). Using eye-tracking to measure worker situation awareness in augmented reality. *Automation in Construction, 165*, 105582. https://doi.org/10.1016/j.autcon.2024.105582.

Xu, W., Chen, K., & Tan, Y. (2026). Reducing visual fatigue by optimizing HCI interface and work–rest scheduling using eye tracking in AR-based building defect inspection. *Advanced Engineering Informatics, 72*, Article 104421. https://doi.org/10.1016/j.aei.2026.104421

Yang, C.-K., Chen, Y.-H., Chuang, T.-J., Shankhwar, K., & Smith, S. (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

Yang, F.-Y., & Wang, H.-Y. (2023). Tracking visual attention during learning of complex science concepts with augmented 3D visualizations. *Computers & Education, 193*, Article 104659. https://doi.org/10.1016/j.compedu.2022.104659

Yoo, D. W., Tarashiyoun, H., & Moghaddam, M. (2023). Modeling gaze behavior for real-time estimation of visual attention and expertise level in augmented reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct* (pp. 487–492). https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00106 (ISMAR-Adjunct).

Yuan, X., Mei, Q., & Li, X. (2023). *Integrating real-time object detection into an AR-driven task assistance prototype: An approach towards reducing specific motions in therbligs theory. In CONVR 2023: Proceedings of the 23rd International Conference on Construction Applications of Virtual Reality* (pp. 121–132). Firenze University Press. https://doi.org/10.36253/979-12-215-0289-3.12.

Yuan, X., Ng, K. K. H., Li, Q., Yiu, C. Y., Lau, C. K., Fung, K. H., & Ng, L. H. (2024). Exploring the human-centric interaction paradigm: Augmented reality-assisted head-up display design for collaborative human-machine interface in cockpit. *Advanced Engineering Informatics, 62*, Article 102656. https://doi.org/10.1016/j.aei.2024.102656

Yue, H., Ye, G., Liu, Q., Yang, X., Xiang, Q., & Luo, Y. (2024). Impact of cognitive fatigue on attention and the implications for construction safety: A neuroscientific perspective. *Journal of Construction Engineering and Management, 150*(8), Article 04024102. https://doi.org/10.1061/JCEMD4.COENG-14711

Zhang, Z., Guo, B. H. W., Feng, Z., & Goh, Y. M. (2025). Investigating situation awareness transition in construction hazard recognition: A multimodal study of cognitive and neural mechanisms. *Advanced Engineering Informatics, 68*, Article 103712. https://doi.org/10.1016/j.aei.2025.103712