

# Evaluating Multilingual User Interfaces for AR+BIM Construction Training

Joseph T Kider Jr.  
School of Computing  
Clemson University  
Charleston, South Carolina, USA  
jkider@clemson.edu

Maria Mollica Redmon  
Modern Languages and Literatures  
University of Central Florida  
Orlando, Florida, USA  
maria.redmon@ucf.edu

Joseph LaViola  
University of Central Florida  
Orlando, Florida, USA  
jjl@cs.ucf.edu

Amirpouya Ghasemaghaei  
ISUE Lab  
University of Central Florida  
Orlando, Florida, USA  
aghaei.ap@ucf.edu

Lori C Walters  
Institute for Simulation & Training  
University of Central Florida  
Orlando, Florida, USA  
Lori.walters@ucf.edu

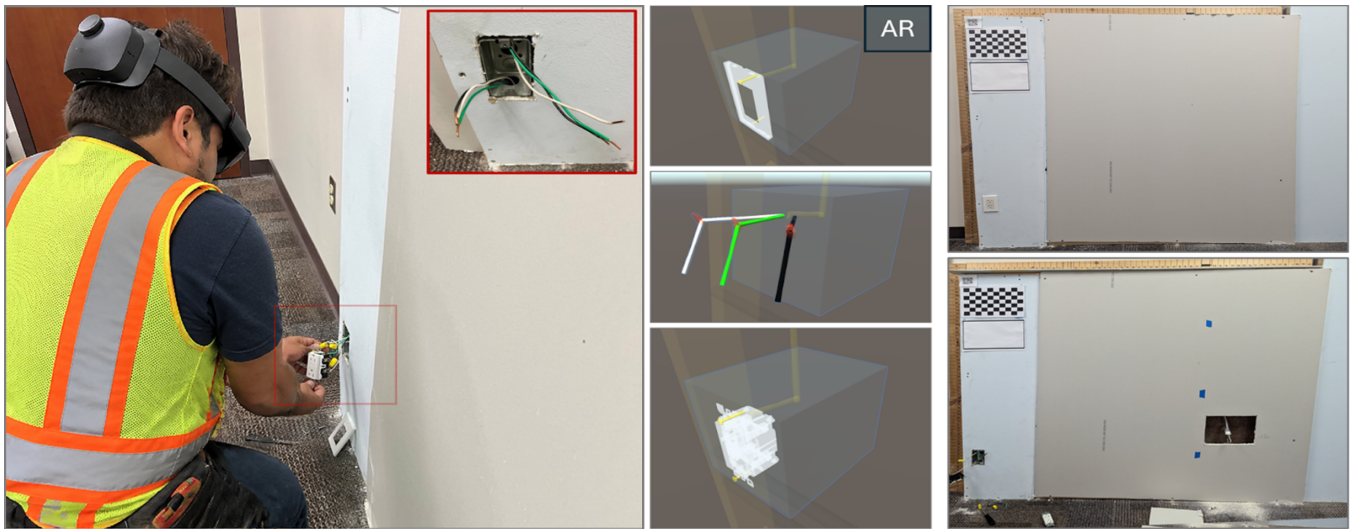


**Figure 1: Augmented reality + BIM-assisted electrical outlet installation workflow with native-language (L1) guidance. Left: AR-assisted outlet installation with workspace detection and translated instructions. Middle: AR visualizations of outlet placement, wire pose estimation, and device alignment, with L1 instructions to reduce bilingual switching. Right: Experimental setup (top) and installation panel (bottom).**

## Abstract

Augmented Reality (AR) interfaces integrated with Building Information Modeling (BIM) are increasingly used to support construction training and on-site tasks. Existing AR+BIM research has largely emphasized visual augmentation and spatial guidance, with limited attention to linguistic accessibility as a user interface design factor. This gap is significant in the U.S. construction industry, where many workers are native Spanish speakers who routinely interact with technical systems in a second language. This paper investigates how user interface language influences performance

and cognitive workload in AR+BIM-assisted construction training. We conducted a controlled study in which Spanish-speaking participants completed a multi-step electrical installation and modification task using identical AR+BIM interfaces presented in either Spanish (L1) or English (L2). In addition to completion time and procedural accuracy, we evaluated subjective workload, usability, and performance variability while accounting for individual bilingual language differences. By framing language as an interaction variable in XR interface design, we introduce a linguistically grounded perspective for evaluating AR instructional systems. We show that first- and second-language interfaces influence performance consistency and cognitive strategy use during AR-assisted construction tasks. This highlights the importance of considering language when designing and evaluating multilingual AR+BIM interfaces.



This work is licensed under a Creative Commons Attribution 4.0 International License.  
SUI '26, Bari, Italy

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2812-9/26/10  
<https://doi.org/10.1145/3822518.3830067>

## CCS Concepts

• **Human-centered computing** → **User studies; Mixed / augmented reality**; *Accessibility systems and tools*.

## Keywords

Bilingual Language Switching, Augmented Reality (AR), Building Information Modeling (BIM), Construction Training, Retrofitting.

### ACM Reference Format:

Joseph T Kider Jr., Maria Mollica Redmon, Joseph LaViola, Amirpouya Ghasemaghaei, and Lori C Walters. 2026. Evaluating Multilingual User Interfaces for AR+BIM Construction Training. In *ACM Symposium on Spatial User Interaction (SUI '26)*, October 10–11, 2026, Bari, Italy. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3822518.3830067>

## 1 Introduction

Augmented Reality (AR) systems integrated with Building Information Modeling (BIM) have emerged as effective interfaces for construction training, assembly guidance, and on-site retrofitting workflows [6, 51]. Research in augmented and extended reality has primarily focused on spatial registration accuracy, visualization fidelity, interaction techniques [32], safety, and task performance optimization [19]. These advances have demonstrated measurable gains in efficiency [25], reduced errors [61], and improved training transfer across a variety of construction scenarios [2, 69]. Most of this work, however, has examined spatial presentation and interaction techniques. Despite the inherently multimodal nature of AR [44], relatively little research has experimentally examined how interface language—whether presented in an AR user’s first (L1) or second language (L2) affects task performance and cognitive demand [41, 49]. Language is a key factor in shaping how users understand tasks, influencing performance, cognitive workload, and other operational factors, particularly due to bilingual switching costs arising from users’ alternation between their first and second languages during task execution [10].

In the United States, Hispanic workers comprise a substantial portion of the construction workforce, and many are native Spanish speakers with varying degrees of English proficiency. Despite this, safety documentation, technical specifications, and digital training systems are often provided primarily in English. When AR systems deliver procedural guidance in L2, workers must simultaneously manage bilingual switching [30], interpret domain-specific terminology, and perform physically grounded tasks in real time. This simultaneous demand increases the risk of errors, compromises safety, and slows task completion as workers divide their attention between understanding instructions and executing actions [21]. AR systems and workflows are often evaluated using frameworks grounded in human–computer interaction and cognitive load theory, showing that poorly designed overlays, cluttered interfaces, or excessive information density can increase extraneous cognitive load and subsequently degrade task performance [7, 16]. Language itself utilizes a separate processing channel that also consumes working memory. Comprehension in an L2 typically involves slower lexical retrieval, greater attentional demands, and, sometimes, implicit translation before actions can be executed [29]. In time-sensitive or safety-critical workflows, these additional cognitive demands can reduce efficiency and increase performance

variability, even if AR visual augmentation and interaction design are optimized. We aim to answer the following research questions:

**RQ1:** Does interface language (L1 Spanish v. L2 English) influence task performance in an AR-assisted construction task?

**RQ2:** Do individual language background factors moderate performance and workload outcomes within the L2 group?

**RQ3:** Does interface language affect perceived cognitive workload during AR-assisted construction tasks?

**RQ4:** How do outcomes from the AR L1/L2 group compare to those from prior AR studies in which instruction language was not experimentally manipulated?

This work makes four primary contributions. First, we identify interface language as a measurable factor in AR systems by examining how first- and second-language instruction influences task performance, consistency, and perceived demand beyond conventional AR usability metrics. Second, we present a controlled experiment on L1 and L2 AR interfaces, measuring their effects on task outcomes and user experience. Third, by integrating standardized bilingual measures (BSWQ [53], SILL [50], BLP [5], LEAP-Q [40]) into AR evaluation, we examine how individual language background and strategy use contribute to differences in AR task performance (Figure 1). Finally, we show that language-related factors contribute to differences in AR interaction outcomes beyond interface design and task characteristics. These findings highlight the importance of bilingual user characteristics in multilingual AR+BIM systems.

## 2 Related Work

Prior research has examined AR+BIM systems for supporting procedural guidance, training, and task execution in construction workflows. Studies in cognitive science and human performance show that language processing and working memory strongly influence how individuals understand and carry out multi-step instructions. Linguistic UI accessibility has received limited attention as a human factor in AR instructional systems.

### 2.1 AR and BIM for Construction

Emerging technologies such as Building Information Modeling and Augmented Reality are increasingly being adopted in the architecture, engineering, and construction industry to support visualization, coordination, and on-site decision-making [47, 51]. BIM provides a structured digital representation of building components and project information, enabling stakeholders to access spatial, geometric, and semantic data throughout the project lifecycle [3, 42, 70]. As BIM adoption has grown, research has examined additional ways to make this information available to on-site workers [6, 23, 58, 59]. This is important during construction, inspection, and retrofitting, where understanding the locations of hidden or planned building components is essential [15, 25, 57, 69]. AR systems can effectively display BIM information on construction sites and in buildings [25]. Integrating AR+BIM into training and on-site tasks enables embodied, hands-free interaction with contextually embedded data. This enhances the environment’s affordances [17, 48], expands perceptual capabilities [18], sense-making [65], and supports leveraging principles of 4E Cognition [9, 46] to reduce cognitive load and strengthen worker engagement with the work site [67].



**Figure 2: Using an AR+BIM L1 headset to guide an electrical outlet wiring and installation task. (1) The testing apparatus for the experiment overlaid with the augmented view behind the wall, (2) the participant aligns the AR-guided placement on the wall; (3) AR guidance supports cutting the drywall opening for the outlet box; and (4) the participant installs the electrical box following the AR instructions. The AR system provides both spatial guidance and L1 instructions to assist with accurate placement and step-by-step completion of the installation procedure.**

Unlike tablet-based or screen-mediated instruction, AR delivers procedural guidance spatially registered to physical components in the environment. This means a worker receiving step-by-step wiring instructions sees those instructions anchored to the actual outlet box, stud location, or wire run in their field of view, eliminating the need to shift attention between a handheld screen and the task surface. Prior work in AR+BIM has demonstrated the potential of these systems to support tasks such as layout verification [32], assembly guidance [38], inspection [71], progress monitoring [19, 56], and construction training. For example, AR interfaces [7] have been shown to improve task efficiency and reduce errors [2, 61] and reduce cognitive and physical demands during retrofitting training procedures when compared with traditional blueprint-based documentation. In many AR-supported workflows, spatial guidance is augmented with textual instructions, labels, or verbal cues [39, 55]. For example, AR can project step-by-step assembly instructions onto components, overlay wiring diagrams on electrical panels, display part numbers directly on structural elements, highlight connection points through sensor-based alerts, and provide voice prompts to guide maintenance and safety procedures [6]. Effective language use is crucial on these platforms and often overlooked in AR+BIM systems, as clear textual or verbal instructions directly affect how workers interpret, execute, and coordinate complex tasks.

## 2.2 Language Processing and Working Memory

Research in bilingualism and cognitive science has examined how individuals process and manage multiple languages during complex tasks. Bilingual language processing involves simultaneously activating both language systems, requiring cognitive control to use the target language and suppress processing and interference from the non-target language(s). A key phenomenon is the Bilingual Switch Cost [30], the cognitive cost of switching between languages or tasks, often reflected in slower responses or reduced accuracy [29]. This effect shows the connection between language control and executive functions [21]. Computational theories consider factors such as task type (linguistic vs. non-linguistic) and switching demands (difficulty or frequency) to explain how bilingualism affects cognitive processing [34]. Several theories and models have been proposed to explain bilingual language processing, including the

Visual World Paradigm (VWP) [33], the Inhibitory Control Model (ICM) [21], and Bilingual Interactive Activation (BIA and BIA+) [11, 12]. Additional frameworks relevant to bilingual task performance include code-switching theories, additive vs. subtractive bilingualism [28], and the Developmental/Modified Interdependence Hypothesis (Interdependence Theory) [26].

These models propose that bilingual switching costs arise from the need to regulate competing lexical systems [29] and manage language selection during tasks involving multiple languages [21]. Working memory is crucial in this process since understanding language and completing tasks rely on the same mental resources. Such memory models of language comprehension propose that language information must be maintained while it is integrated with task goals and environmental cues [22]. L2 processing may require more attention and working memory than L1 processing [45]. This is especially valid when users must understand instructions while performing spatial or procedural tasks.

## 2.3 Language Accessibility in XR Interfaces

Language barriers often create challenges in the workforce. Goodrum et al. [20] found that these barriers contribute to higher rates of injury and death among Hispanic workers [54]. Dong and Platner [13] further showed that Spanish-speaking workers are more likely to perform hazardous tasks [14] while often struggling to understand instructions, hazards, and safety warnings because of limited English proficiency [43]. These barriers create a disconnect in the mental models required for effective communication between workers and employers [36], leading to fewer career opportunities, lower average earnings, limited training opportunities, and consequently fewer role models for Spanish-speaking workers in the built environment [64]. Recent advances in speech and text translation have substantially improved multilingual communication through automated speech recognition [4], optical character recognition [62], machine translation [60], and natural language processing [35]. Supported by large multilingual datasets [1, 72], these deep learning approaches now enable increasingly accurate real-time translation.

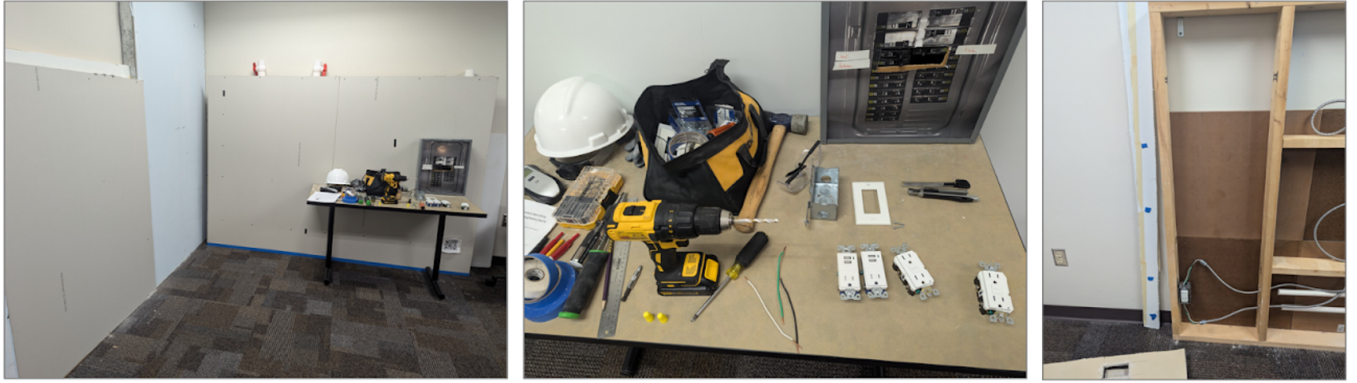


**Figure 3: Experimental setup and materials used in the study. (Left) Workspace for the construction task. (Center) Tools, hardware, and components provided. (Right) Experimental apparatus.**

Recent advances in large language models (LLMs) [63, 73] and neural speech translation have improved the accuracy and flexibility of real-time multilingual communication tools [52]. Platforms such as *Microsoft Azure AI Translator* (*Azure Cognitive Services*), OpenAI’s GPT-based language models, and *Google Cloud Translation* provide APIs for speech recognition, machine translation, and text generation across dozens of languages. These services integrate with XR frameworks, including the *HoloLens 2 Cognitive Services* API, enabling real-time transcription and translation of spoken or written instructions displayed as spatial overlays or audio feedback in augmented reality environments [31]. Using augmented reality [68], translated text and audio feedback can be presented to users, improving situational awareness and transparency [41]. Users must interpret spatial information, instructions, and interface elements while interacting with immersive environments [37]. Prior VR/AR research focused on second-language learning [10, 27] and has explored real-time speech translation, captioning, and multilingual interface design to support communication and comprehension in collaborative and training settings [49, 66]. Bilingual language processing creates cognitive demands that can affect task performance. However, language accessibility in AR+BIM interfaces remains underexplored, particularly for construction training and field tasks.

### 3 Methodology

This section describes the study design and presents our hypotheses. We outline the software systems and materials used in the study, which are available for review and replication (*Supplement*).

#### 3.1 Participants

This study recruited 24 Spanish-speaking non-expert participants (19 male, 5 female; age 18–45 years,  $M = 25.9$ ,  $SD = 10.9$ ). Participants had normal or corrected-to-normal vision, reported no sensory or motor impairments, and had no history of virtual reality-related sickness. They were recruited from a large southeastern U.S. university and represented diverse academic backgrounds. Eligibility required Spanish as the participant’s first language (L1), defined by (a) Spanish age of acquisition (AoA) before three years, (b) self-reported Spanish as the dominant language at home, and (c)

a LEAP-Q proficiency balance score indicating Spanish dominance or near-balance. All participants completed the study.

To address RQ4, the results are compared with a prior cohort study that used the same apparatus, installation task, NASA-TLX, and SUS instruments [57]. That study included 66 non-expert students (45 male, 20 female, 1 other gender; age 18–45 years,  $M = 19.62$ ,  $SD = 1.99$ ) who were not screened for first language and were assigned to AR or Paper instruction conditions without linguistic control. Two participants (one from each condition) were excluded because they did not complete the study. This cohort serves as the RQ4 baseline, enabling comparison of workload and usability outcomes between language-controlled and uncontrolled AR cohorts to evaluate whether interface language explains variation previously attributed to interface or task factors (Table 1).

**Table 1: Summary of experimental conditions.**

| Condition          | Language     | Medium               | Cohort  | Sample   |
|--------------------|--------------|----------------------|---------|----------|
| L1 Spanish AR      | Spanish (L1) | HoloLens 2 AR        | Current | $n = 12$ |
| L2 English AR      | English (L2) | HoloLens 2 AR        | Current | $n = 12$ |
| AR Uncontrolled    | Uncontrolled | HoloLens 2 AR        | Prior   | $n = 33$ |
| Paper Uncontrolled | Uncontrolled | Printed Instructions | Prior   | $n = 33$ |

#### 3.2 Study Design

In a controlled laboratory experiment, participants investigated the impact of bilingual switching in an AR+BIM augmentation for built-environment retrofitting tasks. Participants completed a structured electrical installation procedure using an AR guidance system that provided step-by-step instructions and visual augmentations. The primary independent variable was the user interface (UI) language, defined by the availability and use of L1 (the participant’s native language, Spanish) or L2 (English) guidance within the AR interface. Participants were randomly assigned to one of two experimental conditions: L1 ( $n = 12$ ) or L2-bilingual switching ( $n = 12$ ). The total sample comprised 24 participants. Given the sample size, the study should be considered exploratory and was primarily intended to estimate the magnitude and direction of potential effects rather than to provide high statistical power between conditions.

A between-subjects design was used to compare the effects of the two language conditions on task performance, instruction comprehension, and cognitive workload during the procedural training task while minimizing learning and carryover effects between conditions. A within-subjects design was not adopted because repeated exposure to the installation workflow and AR interface could improve performance through practice and familiarity, making it more difficult to attribute observed differences to the language condition alone. The experimental task consisted of a two-stage GFCI electrical outlet installation: (1) replacement and rewiring of an existing outlet and (2) installation of a new junction box and outlet. Throughout the task, the AR system provided spatially registered instructions, tool identification cues, and step-by-step procedural guidance aligned with the physical wall apparatus.

Before beginning the task, participants watched a two-minute video outlining the study's purpose and providing a general overview of retrofitting a structure, its benefits, and the task procedure. They also received a verbal explanation of the relevant building components, drywall, wall studs, PVC piping, wiring, and electrical outlets, and training on how to use their assigned AR system. The video and instructions did not reference the study groups or prior study. This training minimized potential differences in prior experience with AR or construction tasks.

Participants completed pre- and post-study questionnaires on a computer using *Qualtrics*. These questionnaires collected demographic information and assessed familiarity with AR and construction. They measured language expertise, skill, and fluency, as well as system usability and perceived task workload. The study was approved by the *University of Central Florida Institutional Review Board* (STUDY00006983). For safety, none of the wiring participants handled was connected to power, and all drywall cutting and drilling were performed by the study administrator or used markings provided by participants.

A reusable wall apparatus was constructed for this study (Figures 1 and 2). The structure simulated a realistic retrofit installation scenario while protecting participants and existing infrastructure, enabling repeated experiments. It contained pre-installed wiring and embedded pipes within the wall cavity that participants had to identify and avoid during installation planning (Figure 3). The experimental procedure consisted of two tasks. In the first, participants replaced an existing lower-left outlet with a GFCI outlet by removing the faceplate and outlet, identifying line, load, and ground wires, and preparing connections using pigtails and twist-on connectors. In the second, participants installed a new outlet on the right side of the wall by locating a vertical stud and selecting a mounting position approximately 12 inches above the drywall base. Using AR augmentation, participants identified and marked the stud location with painter's tape, marked the outlet position, and outlined the drywall section for removal to accommodate the junction box and drilling clearance. To reduce study duration and preserve apparatus reusability, final drilling and electrical wiring were not performed for all participants to minimize material use and cost. The tasks are summarized in Table 2.

**Table 2: Brief description of required tasks.**

| Task | Description                                                                                                                                                                                                                                     |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1–2  | Locate and identify all required tools and materials (electric drill, jab saw, Phillips screwdriver, painter's tape, safety gloves, stud finder, wires/connectors, GFCI outlet, junction box, mounting screws).                                 |
| 3    | Turn off the power at the breaker box by switching off the marked breaker.                                                                                                                                                                      |
| 4    | Twist preparation step: Using the AR instructions, identify the electrical wires and begin twisting the appropriate pairs of wires together with connectors to prepare the wiring configuration before proceeding with the outlet installation. |
| 5    | Locate the ideal vertical stud using AR augmentation.                                                                                                                                                                                           |
| 6    | Using AR guidance, identify the target installation location for the new outlet on the right side of the wall (approximately 12 inches above the base and 5–10 inches from the right edge).                                                     |
| 7    | Mark the center height of the junction box and the AR-indicated stud edge on the wall using a pencil.                                                                                                                                           |
| 8–9  | Mark the corners of the drywall section to be removed with pencil and tape, connecting them with straight lines to outline the cutout area.                                                                                                     |
| 10   | Verify there is sufficient clearance in the wall opening for the drill to reach the mounting points; if necessary, slightly enlarge the opening. Inform the study administrator once adequate space is confirmed.                               |

### 3.3 AR System

An AR+BIM software system was extended for this study to support language changes. Participants used a *Microsoft HoloLens 2* head-worn display (FoV: 52° diagonal, resolution: 1440 × 936 per eye, refresh rate: 60 Hz) with hand tracking. The AR application was developed in *Unity* (2020.3.34f1) using the *Microsoft Mixed Reality Toolkit*. BIM data were designed in *Revit* 2023 and imported into *Unity* as an FBX file to digitally replicate the experimental apparatus. The AR system provided interactive guidance during retrofit activities by displaying virtual elements, instructions, and task cues (Figures 4 and 5). Before each session, virtual models were aligned with the physical wall using a pre-positioned QR code. The code was scanned 100 times to estimate average position and rotation, followed by manual fine-tuning by the study administrator when needed. Low-opacity virtual models revealed hidden structural components while reducing visual clutter, providing hands-free access to building information. This supported installation planning, reduced measurement requirements, and highlighted critical details.

We designed the system to support bilingual language manipulation by presenting identical procedural content in either Spanish (L1) or English (L2), depending on the participant's assigned condition. All Spanish interface instructions were translated by a fluent Spanish-language expert and independently reviewed for accuracy, clarity, and consistency with the English version. Participants could interact with these instructions through a toggleable menu placed about 0.4 meters in front of their heads. The menu adjusted its position if it moved more than 60 degrees from the user's forward gaze. A 'Pin' button at the top-left of the menu lets participants lock it in

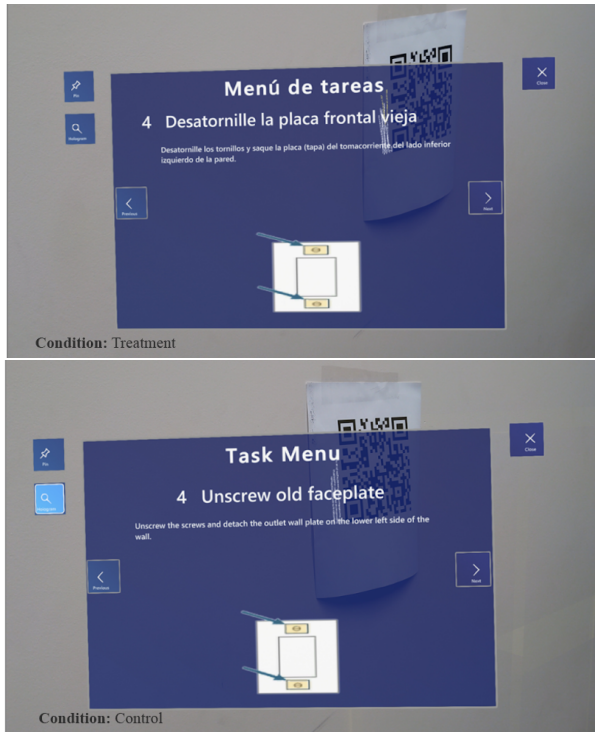


**Figure 4: Side-by-side AR UI comparison: Spanish (L1, top) and English (L2, bottom), both showing step-by-step instructions for removing an old faceplate during outlet installation.**

place. When pinned, they could move the menu's position, rotation, and size using one or both hands. Before starting the installation task, participants completed a training session lasting up to three minutes. During this time, they practiced pinning, adjusting, and unpinning the menu. Participants could finish this training early if they felt sufficiently confident.

### 3.4 Procedure

Participants first reviewed the consent form and provided informed consent. They were assigned a participant ID and study condition, then completed pre-study digital questionnaires. These questionnaires collected demographic information, screened for physical limitations, and assessed prior experience with AR, VR, BIM, and construction practices. Participants also completed two baseline language measures: the Language Experience and Proficiency Questionnaire (LEAP-Q) and the English-Spanish Bilingual Language Profile (BLP). After receiving task and safety instructions, participants completed AR calibration and began the experimental session, which lasted approximately 30–45 minutes.

The experiment included two sequential retrofit tasks using a test wall apparatus. Participants wore work gloves, and no tools or wiring were connected to live electricity. In Task 1, participants replaced a standard outlet with a GFCI outlet by removing the outlet and wall plate, identifying the wiring configuration, disconnecting existing wires, and installing the new outlet. In Task 2, participants planned installation of a second GFCI outlet by identifying the installation location, marking the cutout area, and verifying clearance

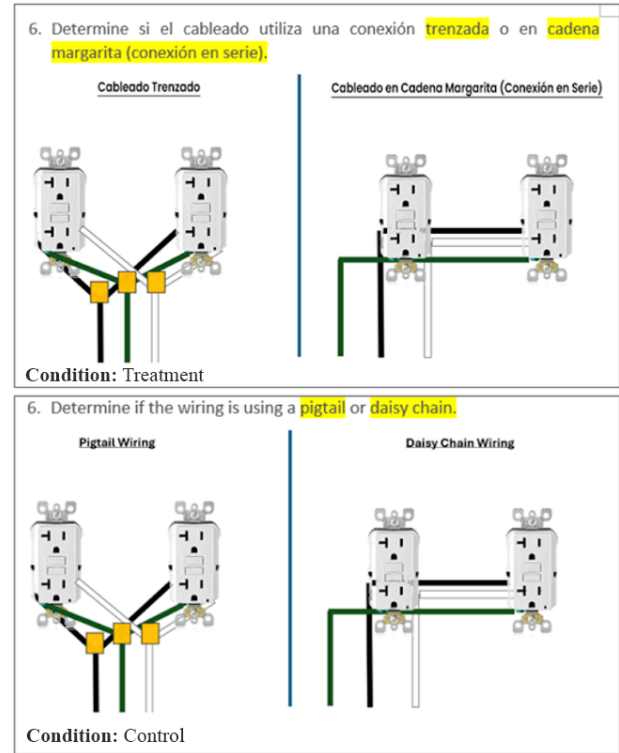


**Figure 5: Side-by-side example of bilingual wiring instructions in the AR system for a complex wire installation step.**

for future mounting and drilling. Following prior outlet installation studies, the procedure emphasized realistic task interactions while maintaining safety; participants did not perform higher-risk drilling or keyhole saw operations. After completing the tasks, participants completed four post-task questionnaires: the System Usability Scale (SUS) [8], the NASA Task Load Index (NASA-TLX) [24], and two language-related instruments (BSWQ and SILL).

### 3.5 Measures

Primary outcome measures focused on task effectiveness under bilingual AR instruction. Four bilingual background measures were included to contextualize the Bilingual Switching Cost. LEAP-Q characterized participants' language history and self-reported proficiency, while the English-Spanish Bilingual Language Profile assessed relative language dominance. The BSWQ measured individual differences in bilingual switching, and the SILL captured self-reported language-learning strategies. Perspective-corrected photos of the apparatus were collected after each experiment to qualitatively evaluate task performance, including stud locations (Step 5), outlet placement (Steps 6–8), and cutout size and errors (Steps 9–10). Task completion timestamps were manually recorded using a custom Python script to calculate durations, and critical errors preventing task progression were documented. The SUS and NASA-TLX assessed perceived usability and workload across five dimensions: mental, physical, and temporal demands, effort, and performance. Retaining comparable task performance measures,

SUS, and NASA-TLX metrics allows evaluation of whether differences across AR cohorts are attributable to interface language rather than AR UI characteristics alone.

### 3.6 Hypotheses

For this experiment, our hypotheses focus on how interface language influences training consistency, cognitive workload, and task performance during AR-assisted construction tasks. Specifically, we examine whether providing instructions in participants' first language (L1) reduces cognitive load and variability in task execution compared with second-language (L2) instruction, and whether this affects overall task completion time.

**(1) Performance: H1a:** Participants receiving instructions in L1 (Spanish) will demonstrate shorter task completion times than those receiving instructions in L2 (English) (RQ1). **H1b:** Participants receiving L1 instruction will commit fewer procedural errors than those receiving L2 instruction (RQ1). **H1c:** Performance variability will be lower in the L1 condition than in the L2 condition (RQ1).

**(2) Language Effects (Individual Differences): H2a:** Greater L2 dominance (BLP) will be associated with reduced task time and lower workload in the L2 condition (RQ2). **H2b:** Higher bilingual switching (BSWQ) will correlate with reduced performance penalties under L2 instruction (RQ2). **H2c:** Strategic language learning behaviors (SILL) will moderate L2 task efficiency (RQ2).

**(3) Cognitive Workload: H3:** Participants in the L2 condition will report higher NASA-TLX scores compared to participants in the L1 condition (RQ3). **(4) Cross-Study Comparison: H4:** Performance and workload outcomes in the current AR cohort will differ from those observed in prior AR studies, suggesting that interface language may account for variance in task efficiency, usability, and perceived workload that was previously attributed to general system usability or task difficulty (RQ4).

## 4 Results

### 4.1 Explanatory value of the language measures

To characterize participants' bilingual background and identify the most informative language-related predictors for subsequent analyses, we computed statistics, internal consistency estimates, and intercorrelations for the LEAP-Q, BLP, BSWQ, and SILL measures. The sample ( $n = 24$ ) reflected a largely balanced bilingual profile with high proficiency in both languages, but with substantial individual variability in dominance and language experience.

**Table 3: Summary of bilingual profile measures and reliability estimates. Full variable table in the Supplement.**

| Instrument | Main finding                                                                                                                              |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| LEAP-Q     | Participants showed high proficiency in both English ( $M = 9.56$ ) and Spanish ( $M = 9.40$ ), with greater current exposure to English. |
| BLP        | Overall language dominance was near balanced ( $M = 2.23$ ).                                                                              |
| BSWQ       | Moderate levels of language switching were reported ( $M = 2.81$ ).                                                                       |
| SILL       | Participants reported moderate-to-high strategy use overall ( $M = 3.24$ , $\alpha = .92$ ).                                              |

LEAP-Q responses (Table 3) indicated very early acquisition of Spanish ( $M = 0.92$ ,  $SD = 1.32$ ) and later acquisition of English ( $M = 6.54$ ,  $SD = 3.64$ ). Self-rated proficiency was high in both languages (Spanish:  $M = 9.40$ ,  $SD = 1.00$ ; English:  $M = 9.56$ ,  $SD = 0.77$ ), and proficiency balance was close to zero ( $M = 0.15$ ,  $SD =$

1.42), indicating broadly comparable proficiency across languages. Current exposure was somewhat higher for English ( $M = 7.48$ ,  $SD = 1.67$ ) than for Spanish ( $M = 5.37$ ,  $SD = 1.96$ ), resulting in a modest English-leaning exposure balance ( $M = 2.11$ ,  $SD = 2.01$ ).

BLP scores (Table 3) showed a similar pattern. Mean global English ( $M = 91.64$ ,  $SD = 24.14$ ) and Spanish ( $M = 89.41$ ,  $SD = 19.67$ ) scores were comparable, while the weighted dominance index was near zero ( $M = 2.23$ ,  $SD = 40.74$ ) with substantial dispersion, supporting its use as a continuous predictor. BSWQ scores indicated moderate code-switching ( $M = 2.81$ ,  $SD = 0.62$ ), and SILL scores showed moderate strategy use ( $M = 3.24$ ,  $SD = 0.56$ ), with cognitive strategies highest ( $M = 3.82$ ,  $SD = 0.58$ ). Reliability was strongest for composite measures: BLP proficiency scales showed good internal consistency (English:  $\alpha = .84$ , Spanish:  $\alpha = .83$ ), BSWQ reliability was modest ( $\alpha = .63$ ), and SILL reliability was excellent ( $\alpha = .92$ ). BSWQ subscales, particularly controlled switching ( $\alpha = .44$ ), were less reliable and interpreted cautiously. Correlation analysis showed strong association between BLP dominance and LEAP-Q proficiency balance ( $r = .75$ ), suggesting overlap in dominance measurement. LEAP-Q exposure balance was moderately related to BLP dominance ( $r = .44$ ) and proficiency balance ( $r = .65$ ). BSWQ and SILL scores showed weak-to-moderate relationships with dominance measures. These results (Table 3) support treating the language questionnaires as complementary rather than interchangeable. We use LEAP-Q primarily for participant characterization; BLP-weighted dominance as the principal dominance/balance predictor; BSWQ total as the primary switching measure; and SILL overall as the primary strategy measure in subsequent analyses as the primary language-related explanatory variables.

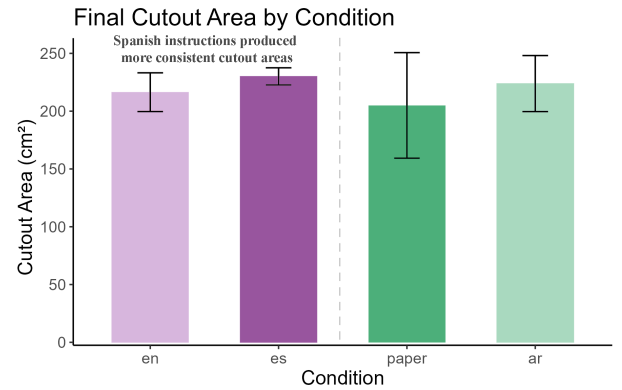


**Figure 6: Mean cutout area by condition with 95% confidence intervals. The ES (L1) condition showed lower variance.**

### 4.2 Effects of L1 versus L2 interface on performance and experience

Figure 6 presents the mean cutout area ( $\text{cm}^2$ ) by condition, with error bars representing 95% confidence intervals. The four conditions (EN, ES, Paper, and AR) are compared using the cutout areas shown in Figure 7. Cutout area did not differ significantly between the English condition ( $M = 210\text{cm}^2$ ,  $SD = 25\text{cm}^2$ ) and Spanish condition ( $M = 225\text{cm}^2$ ,  $SD = 25\text{cm}^2$ ), Welch's  $t(15.15) = -1.64$ ,  $p = 0.121$ .

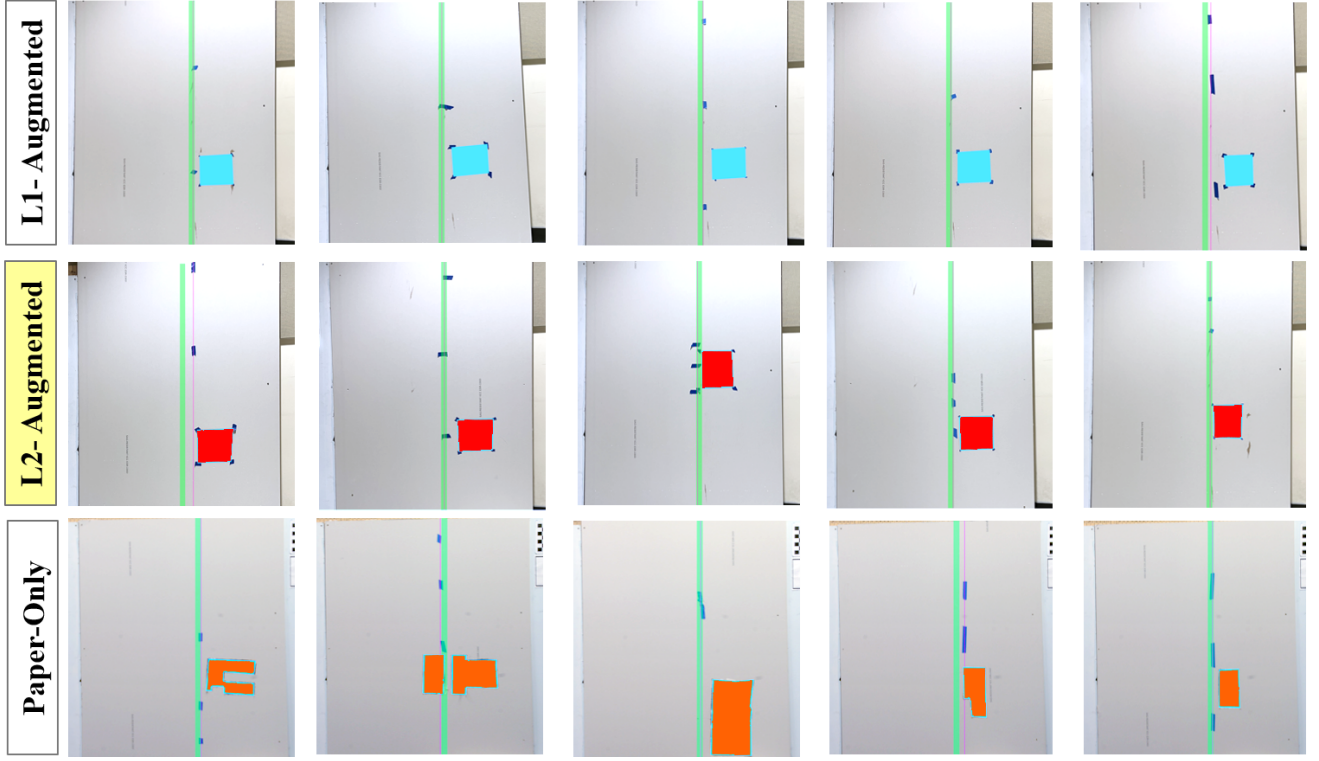


Figure 7: Examples of participant cutouts for the L1, L2 and Paper groups. Cutout areas are highlighted. The cutouts by the Paper participants were noticeably more irregular and somewhat varied for L2, which would be more difficult to patch.

The ES (L1) condition exhibited lower variability than the other conditions, suggesting more consistent cutout performance, whereas the Paper condition showed greater variability across participants.

To examine whether bilingual language experience predicts spatial selection behavior, we analyzed the relationship between cutout area and three language measures: BLP dominance, BSWQ switching frequency, and SILL strategy use. Pearson correlations and simple regressions, conducted separately for L1 Spanish and L2 English participants, showed that only SILL strategy use was significantly associated with cutout area for L2 participants ( $r = .62$ ,  $p = .031$ ), indicating that participants who reported greater use of language learning strategies selected larger visual regions to cutout. No significant relationships were observed for L1 participants or for BLP dominance and BSWQ switching measures.

To test whether these relationships differed across language groups, we fit a linear model including language condition and interactions with each language measure. The model was significant overall ( $F(7, 16) = 3.07$ ,  $p = .030$ ,  $R^2 = .57$ ). Strategy use (SILL overall) positively predicted cutout area ( $\beta = 27.69$ ,  $p = .024$ ), but this effect differed significantly between language groups ( $\beta = -40.82$ ,  $p = .014$ ), indicating that the relationship between strategy use and spatial selection was stronger for L2 participants. Switching frequency also showed a significant interaction with language group ( $\beta = 31.04$ ,  $p = .042$ ). Bilingual dominance (BLP) was not associated with cutout area ( $p = .64$ ). A direct comparison of mean cutout area between L1 and L2 groups revealed no significant difference ( $t(15.15) = -1.64$ ,  $p = .121$ ), suggesting that the observed effects

Table 4: Relationships between language measures and cutout area. Pearson correlations are shown separately for L1 and L2 participants.

| Predictor                                                              | L1 r               | L1 p            | L2 r        | L2 p        |
|------------------------------------------------------------------------|--------------------|-----------------|-------------|-------------|
| BLP Dominance                                                          | 0.07               | .827            | -0.34       | .274        |
| BSWQ Switching                                                         | 0.19               | .561            | -0.48       | .113        |
| SILL Strategy Use                                                      | -0.35              | .270            | <b>0.62</b> | <b>.031</b> |
| <i>Interaction Model (cutout area ~ language measures × condition)</i> |                    |                 |             |             |
| SILL Overall                                                           | $\beta = 27.69$    | $p = .024$      |             |             |
| SILL × Condition                                                       | $\beta = -40.82$   | $p = .014$      |             |             |
| BSWQ × Condition                                                       | $\beta = 31.04$    | $p = .042$      |             |             |
| BLP Dominance                                                          | $p = .637$ (ns)    |                 |             |             |
| L1 vs L2 Cutout Area                                                   | $t(15.15) = -1.64$ | $p = .121$ (ns) |             |             |

reflect individual differences in language experience rather than overall group differences (Table 4).

### 4.3 System Usability

We compared perceived usability and cognitive load between participants interacting with the system in English and Spanish. Results showed no significant differences between groups for either usability ( $t(22) = 1.28$ ,  $p = .21$ ) or overall cognitive load ( $t(22) = 0.99$ ,  $p = .34$ ). Mean SUS scores were slightly higher in the English condition ( $M = 3.58$ ,  $SD = 1.16$ ) than in the Spanish condition ( $M = 2.92$ ,  $SD = 1.38$ ). The TLX scores were somewhat higher for

**Table 5: Usability and cognitive load results showing comparisons between English and Spanish participants and regression models predicting usability and cognitive load (TLX).**

| Group comparison         |                   |                   |             |      |           |
|--------------------------|-------------------|-------------------|-------------|------|-----------|
| Measure                  | English Mean (SD) | Spanish Mean (SD) | t(22)       | p    | Cohen's d |
| SUS                      | 3.58 (1.16)       | 2.92 (1.38)       | 1.28        | .21  | 0.52      |
| TLX Total                | 3.71 (0.55)       | 3.32 (1.25)       | 0.99        | .34  | 0.40      |
| Regression predictors    |                   |                   |             |      |           |
| Predictor                | SUS $\beta$       | p                 | TLX $\beta$ | p    |           |
| Language group (Spanish) | -0.53             | .32               | -0.52       | .18  |           |
| BLP dominance            | 0.018             | .078              | 0.012       | .114 |           |
| BSWQ switching           | -0.49             | .29               | 0.16        | .64  |           |
| SILL strategies          | 0.59              | .28               | 0.42        | .29  |           |
| LEAP proficiency balance | -0.29             | .34               | -0.43       | .063 |           |

English ( $M = 3.71$ ,  $SD = 0.55$ ) than Spanish ( $M = 3.32$ ,  $SD = 1.25$ ), though this difference was not statistically significant. Effect sizes were small-to-moderate (Cohen's  $d = 0.52$  for SUS and  $d = 0.40$  for TLX), suggesting that language condition produced only modest differences in perceived usability and cognitive load (Table 5).

We also examined whether individual language characteristics predicted usability and cognitive load. Regression analyses showed a marginal positive relationship between bilingual language dominance and perceived usability ( $\beta = 0.018$ ,  $p = .078$ ), suggesting that stronger language dominance was associated with slightly higher usability ratings. Proficiency balance showed a marginal negative relationship with TLX ( $\beta = -0.43$ ,  $p = .063$ ), indicating lower cognitive load among participants with more balanced proficiency. Correlation analysis showed a trend toward higher cognitive load among participants reporting greater language-learning strategy use ( $r = .38$ ,  $p = .067$ ). Strong associations among TLX subscales, particularly effort and frustration ( $r = .82$ ,  $p < .001$ ), indicate that these components contributed substantially to reported workload (Table 5). SUS and TLX results are visualized in Figure 8 alongside prior language-uncontrolled groups [57].

Post-questionnaire analysis showed comparable usability and workload between the AR and paper interfaces. System Usability Scale scores did not differ significantly between L1 ( $M = 2.92$ ,  $SD = 1.38$ ) and Paper ( $M = 3.50$ ,  $SD \approx 1.2$ ,  $p = 0.539$ ), or between L2 ( $M = 3.58$ ,  $SD = 1.16$ ) and Paper ( $p = 0.996$ ). NASA-TLX scores showed no significant differences between L1 and Paper, with only the L2 versus Paper effort subscale reaching significance ( $p = 0.045$ ); other TLX measures were not significant. These findings suggest comparable usability and workload across interface modalities, although the higher L2 effort rating (L2 vs. Paper:  $M = 5.00$  vs.  $4.25$ ,  $p = 0.045$ ) may reflect additional cognitive demands associated with second-language processing.

## 5 Discussion

Our findings show that interface language is one factor influencing AR task performance, with effects that emerge primarily through performance consistency and user-specific language characteristics. The results provide partial support for our hypotheses.

- **H1a** (Task completion time; RQ1): *Not supported*. No statistically significant differences in task completion time were observed.
- **H1b** (Procedural errors; RQ1): *Not supported*. Error rates did not significantly differ between L1 and L2 interface conditions.

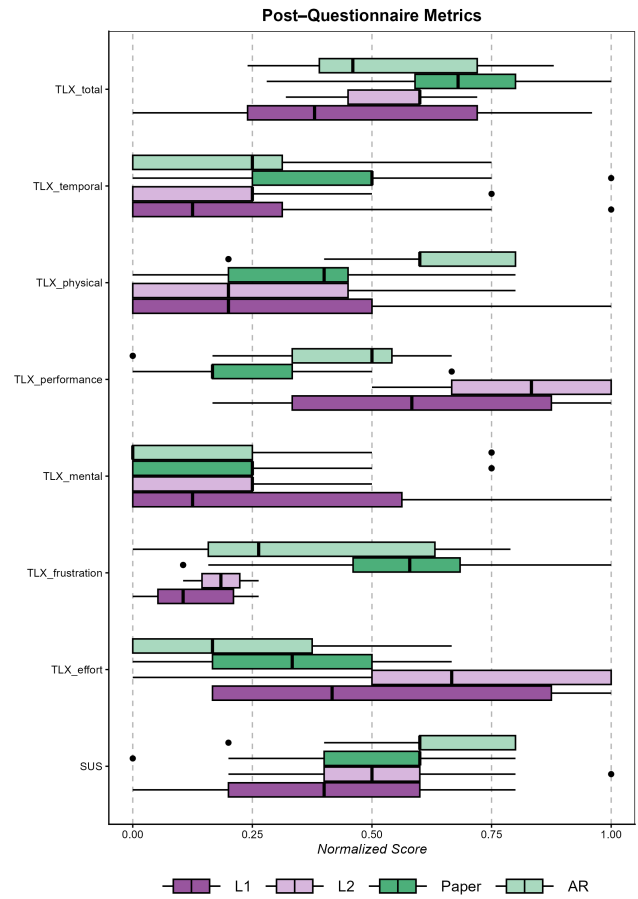


**Figure 8: SUS and NASA-TLX subscale scores, normalized to a 0–1 scale, for the current study's L1 and L2 conditions and the prior cohort's AR and Paper conditions. All raw scores were min-max normalized within their respective instrument's scale range (SUS: 0–100 → 0–1; TLX subscales: 0–100 → 0–1) to allow direct visual comparison across instruments and cohorts on a single axis.**

- **H1c** (Performance variability; RQ1): **Supported**. Participants in the L1 condition exhibited lower variability and more consistent spatial outcomes than those in L2.
- **H2a** (L2 dominance BLP; RQ2): **Partially supported**. Greater L2 dominance showed directional trends toward improved L2 performance and reduced workload, though effects were modest and not uniformly significant across measures.
- **H2b** (Switching frequency BSWQ; RQ2): **Partially supported**. Bilingual switching frequency showed limited correlation with performance and a minor moderating role in this task context.
- **H2c** (Language strategies SILL; RQ2): **Supported**. Strategic language behaviors significantly moderated L2 performance, with evidence of compensatory strategies influencing spatial decisions.
- **H3** (NASA-TLX differences; RQ3): *Not supported*. No significant differences in perceived workload were found between L1 and L2 conditions for these tasks.

- **H4** (Comparison to prior cohort; RQ4): **Supported.** The RQ4 comparison suggests that language background may contribute to variance often attributed to system usability, task difficulty, or workload in AR studies.

Mean performance did not differ significantly between L1 and L2 conditions (H1), but L1 instruction produced more consistent task performance by reducing variability (H1c). In construction-oriented AR+BIM applications, where reliable execution is critical, consistency may be as important as improvements in average performance. Results also show that language effects depend on individual bilingual strategies (H2c), even among highly proficient bilinguals. Because the study used a high-proficiency sample, observed effects may underestimate those in populations with greater L1/L2 proficiency asymmetry. These findings also inform AR evaluation. Language background is rarely treated as a controlled variable in AR studies, yet the RQ4 comparison (H4) demonstrates that it contributes to variability often attributed to interface design or task complexity. Future AR evaluations should consider brief language background measures, such as the LEAP-Q short form, to account for linguistic effects even when language is not the primary focus.

The TLX results show that language affected a specific component of workload rather than overall workload. Although total NASA-TLX scores did not differ between L1 and L2 conditions, the L2 group showed a small but significant increase in the Effort subscale relative to the Paper baseline ( $p = .045$ ). AR-related ergonomic benefits and language-related processing demands operate on partially independent dimensions. The strong correlation between effort and frustration ( $r = .82, p < .001$ ) further indicates that increased demand was concentrated in a specific workload component rather than distributed across the TLX subscales.

## 5.1 Implications for XR Design and Evaluation

Our findings suggest that native-language (L1) support should be considered a UI design requirement. XR systems should move beyond simple language toggles toward adaptive interfaces that incorporate individual bilingual profiles and context to tailor instructions and augmentations. As generative AI enables more widespread real-time translation, evaluating and integrating multilingual capabilities will be essential for developing dynamic XR interfaces that accommodate individual language preferences and dialects.

We focused specifically on text-based instructional content because textual and verbal instructions remain the primary mechanism by which most deployed AR+BIM systems convey procedural and safety information [6, 39]. Text is the modality in which bilingual switching costs are most directly documented and measured in the cognitive science literature [29]. Our results characterize the language-dependent component of AR+BIM instruction specifically, rather than AR interaction as a whole. In the future, extending AR 3D guidance would help isolate how much of the language effect we observed persists once richer non-linguistic supports are present.

Our contribution is not to show that L1 instruction universally outperforms L2 instruction, but to establish interface language as a measurable design variable in AR+BIM systems. Results show that language effects can emerge even among highly proficient bilinguals, with L1 instruction producing more consistent spatial outcomes (H1c) and individual language-management strategies

moderating L2 performance (H2c). AR interfaces should therefore consider both instruction language and user-specific bilingual characteristics. The LEAP-Q, BLP, BSWQ, and SILL provide a multidimensional view of bilingual experience: proficiency and dominance characterize language ability, switching reflects habitual language management, and strategy use captures how users regulate bilingual resources during tasks. These measures explain variability that may otherwise be attributed solely to AR usability or task difficulty.

The absence of significant differences in completion time (H1a), procedural errors (H1b), and overall NASA-TLX workload (H3) should be interpreted in the context of our balanced bilingual sample rather than as evidence that language has no effect. Prior bilingual-processing research suggests that language-related costs are often more apparent in variability and strategy-dependent measures than in mean performance, particularly among highly proficient bilinguals [22, 31, 48]. The results show that language effects in AR may be subtle and user-dependent, emerging through consistency and cognitive strategy rather than broad declines in task performance. Future multilingual AR evaluations should therefore incorporate measures of bilingual background and strategy use to better characterize how language interacts with interface design.

## 5.2 Limitations and Future Work

Several limitations should be considered. First, participants were highly proficient, largely balanced bilinguals (LEAP-Q self-rated proficiency: English  $M = 9.56, SD = 0.77$ ; Spanish  $M = 9.40, SD = 1.00$ ) recruited from a university setting rather than the construction workforce. This differs from many Spanish-speaking construction workers, who may have more asymmetric bilingual experiences, including lower English proficiency and less formal L2 acquisition. Because bilingual processing demands and language-switching costs vary with proficiency asymmetry [21, 29, 45], observed effects may underestimate differences in populations with lower L2 proficiency. Future work should validate these findings with construction workers across broader language backgrounds and task contexts. Second, this study examined text-based AR instruction only; future research should compare text, audio, and multimodal delivery within the same AR framework. The between-subjects design reduced language carryover effects but introduced variability from manual dexterity and prior experience, which future studies should control through measures of construction experience and fine-motor skills.

## 6 Conclusion

This study highlights an overlooked factor in AR+BIM research: interface language as an important design consideration in AR systems. Our results provide initial evidence that the language of instruction influences performance consistency and cognitive strategy use, highlighting language as an important factor in multilingual AR interaction. Future research should examine whether these effects extend to populations with lower L2 proficiency. We argue for L1-first AR design and for integrating language proficiency, bilingual switching costs, and cognitive load into evaluation. Without accounting for language, AR systems risk biased results and reduced accessibility for multilingual users.

## Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2326134.

## References

- [1] Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, 4218–4222. <https://aclanthology.org/2020.lrec-1.520/>
- [2] Sanyukta Arvikar, Pa Pa Win Aung, Gichun Cha, and Seunghye Park. 2025. Augmented Reality-based construction site management using optimized BIM and project schedule integration. *Automation in Construction* 175 (2025), 106204. doi:10.1016/j.autcon.2025.106204
- [3] Salman Azhar. 2011. Building Information Modeling (BIM): Trends, Benefits, Risks, and Challenges for the AEC Industry. *Leadership and Management in Engineering* 11, 3 (July 2011), 241–252. doi:10.1061/(ASCE)LM.1943-5630.0000127
- [4] Advait Bal, P Anjana, KT Gayathri Devi, Vishal S Upot, S Abhishek, and T Anjali. 2024. AR glasses for real-time translation and subtitling. In *International Conference on Information and Communication Technology for Intelligent Systems*. Springer, 279–288. doi:10.1007/978-981-97-6678-9\_25
- [5] David Birdsong, Laura M. Gertken, and Mark Amengual. 2012. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. <https://sites.la.utexas.edu/bilingual/>
- [6] Doug A Bowman, Joseph Gabbard, Daniel Auerbach, Nazila Roofigari-Esfahan, Kathryn Britt, Cory I Ilo, and Keerthana Adapa. 2022. BuildAR: a proof-of-concept prototype of intelligent augmented reality in construction. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 508–512. doi:10.1109/VRW55335.2022.00113
- [7] Doug A Bowman, Joseph L Gabbard, and Deborah Hix. 2002. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators & Virtual Environments* 11, 4 (2002), 404–424. doi:10.1162/105474602760204309
- [8] John Brooke. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7. Publisher: London, England.
- [9] Andy Clark and David J. Chalmers. 1998. The Extended Mind. *Analysis* 58, 1 (1998), 7–19. doi:10.1093/analys/58.1.7
- [10] Che Samihah Che Dalim, Mohd Shahrizal Sunar, Arindam Dey, and Mark Billinghurst. 2020. Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies* 134 (2020), 44–64. doi:10.1016/j.ijhcs.2019.10.002
- [11] Ton Dijkstra and Walter JB Van Heuven. 1998. The BIA model and bilingual word recognition. In *Localist Connectionist Approaches to Human Cognition*. Psychology Press, 189–225.
- [12] Ton Dijkstra and Walter JB Van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5, 3 (2002), 175–197. doi:10.1017/S1366728902003012
- [13] Xiuwen Dong and James W Platner. 2004. Occupational fatalities of Hispanic construction workers from 1992 to 2000. *American journal of industrial medicine* 45, 1 (2004), 45–54. doi:10.1002/ajim.10322
- [14] Xiuwen Sue Dong, Yurong Men, and Knut Ringen. 2010. Work-related injuries among Hispanic construction workers—Evidence from the Medical Expenditure Panel Survey. *American journal of industrial medicine* 53, 6 (2010), 561–569. doi:10.1002/ajim.20799
- [15] Steven Feiner, Blair MacIntyre, and Dorée Seligmann. 1993. Knowledge-based augmented reality. *Commun. ACM* 36, 7 (July 1993), 53–62. doi:10.1145/159544.159587
- [16] Joe L Gabbard and J Edward Swan II. 2008. Usability engineering for augmented reality: Employing user-based studies to inform design. *Transactions on Visualization and Computer Graphics* 14, 3 (2008), 513–525. doi:10.1109/TVCG.2008.24
- [17] James J. Gibson. 1977. The Theory of Affordances. In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, Robert E. Shaw and John Bransford (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 67–82.
- [18] James J. Gibson. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, USA.
- [19] Mani Golparvar-Fard, Femiokya Peña-Mora, and Silvio Savarese. 2009. D4AR—a 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction* 14, 13 (2009), 129–153. <https://www.itcon.org/2009/13>
- [20] Paul M Goodrum and Jiukun Dai. 2005. Differences in occupational injuries, illnesses, and fatalities among Hispanic and non-Hispanic construction workers. *Journal of Construction Engineering and Management* 131, 9 (2005), 1021–1028. doi:10.1061/(ASCE)0733-9364(2005)131:9(1021)
- [21] David W. Green. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1, 2 (1998), 67–81. doi:10.1017/S1366728998000133
- [22] John G Grundy and Kalinka Timmer. 2017. Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research* 33, 3 (2017), 325–340. doi:10.1177/0267658316678286
- [23] Ning Gu and Kerry London. 2010. Understanding and facilitating BIM adoption in the AEC industry. *Automation in Construction* 19, 8 (Dec. 2010), 988–999. doi:10.1016/j.autcon.2010.09.002
- [24] Sandra G Hart. 1986. NASA Task Load Index (TLX).
- [25] Steven Henderson and Steven Feiner. 2010. Exploring the benefits of augmented reality documentation for maintenance and repair. *Transactions on Visualization and Computer Graphics* 17, 10 (2010), 1355–1368. doi:10.1109/TVCG.2010.245
- [26] Philip Herdina and Ulrike Jessner. 2002. A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics. In *A Dynamic Model of Multilingualism*. Multilingual Matters. doi:10.21832/9781853595547
- [27] Xinyi Huang, Di Zou, Gary Cheng, and Haoran Xie. 2021. A systematic review of AR and VR enhanced language learning. *Sustainability* 13, 9 (2021), 4639. doi:10.3390/su13094639
- [28] Anita D. Ianco-Worrall. 1972. Bilingualism and cognitive development. *Child Development* (1972), 1390–1400. doi:10.2307/1127524
- [29] Nan Jiang. 2023. Language Switch and Control in Bilinguals. In *The Study of Bilingual Language Processing*, Nan Jiang (Ed.). Oxford University Press, Chapter 5, 144–187. doi:10.1093/oso/9780198852384.003.0005
- [30] Nan Jiang. 2023. *The study of bilingual language processing*. Oxford University Press.
- [31] Parul Khanna, Ravdeep Kour, and Ramin Karim. 2025. Human-centric Maintenance Process Through Integration of AI, Speech, and AR. doi:10.48550/arXiv.2511.13918
- [32] Julian Kreimeier, Hannah Schieber, Shiyu Li, Alejandro Martin-Gomez, and Daniel Roth. 2024. Visual Guidance for Assembly Processes. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 652–653. doi:10.1109/ISMAR-Adjunct64951.2024.00195
- [33] Helene Kreysa, Eva M. Nunnemann, and Pia Knoeferle. 2018. Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. *Acta Psychologica* 188 (2018), 220–229. doi:10.1016/j.actpsy.2018.05.001
- [34] Judith F Kroll and Annette MB De Groot. 2009. *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.
- [35] William D. Lewis. 2015. Skype Translator: Breaking Down Language and Hearing Barriers. In *Proceedings of Translating and the Computer* 37. AsLing, London, UK, 125–149. <https://aclanthology.org/2015.tc-1.9.pdf>
- [36] Mihai Lintean, Vasile Rus, and Roger Azevedo. 2012. Automatic detection of student mental models based on natural language student input during metacognitive skill training. *International Journal of Artificial Intelligence in Education* 21, 3 (2012), 169–190. doi:10.3233/JAI-2012-022
- [37] Chaoun Liu, Wenxuan Zhang, Yiran Zhao, Luu Anh Tuan, and Lidong Bing. 2025. Is translation all you need? a study on solving multilingual tasks with large language models. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 9594–9614. doi:10.18653/v1/2025.naacl-long.485
- [38] Tianyu Liu, Weiping He, Bokai Zheng, Jilong Bai, Xiaotian Zhang, Wenbo Pang, Yanli He, Mark Billinghurst, and Lili Wang. 2026. Exploring the core cognitive dynamics of MR-enabled in-situ assembly visualization. *Journal of Manufacturing Systems* 86 (2026), 67–81. doi:10.1016/j.jmsy.2026.02.018
- [39] Mark A. Livingston, Zhuming Ai, Kevin Karsch, and Gregory O. Gibson. 2011. User interface design for military AR applications. *Virtual Reality* 15, 2-3 (June 2011), 175–184. doi:10.1007/s10055-010-0179-1
- [40] Viorica Marian, Henrike K. Blumenfeld, and Marina Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50, 4 (2007), 940–967. doi:10.1044/1092-4388(2007)067
- [41] Florian Mathis, Yu Sun, and Adrian Preussner. 2024. Mrtranslate: Bridging language barriers in the physical world using a mixed reality point-and-translate system. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces (AVI '24)*. ACM, New York, NY, USA, Article 12, 9 pages. doi:10.1145/3656650.3656652
- [42] Petr Matějka and Aleš Tomek. 2017. Ontology of BIM in a Construction Project Life Cycle. *Procedia Engineering* 196 (2017), 1080–1087. doi:10.1016/j.proeng.2017.08.065
- [43] Nancy Nivison Menzel and Antonio P Gutierrez. 2010. Latino worker perceptions of construction risks. *American journal of industrial medicine* 53, 2 (2010), 179–187. doi:10.1002/ajim.20735
- [44] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.
- [45] Catherine Monnier, Julie Boiché, Pauline Armandon, Sophie Baudoin, and Stephanie Bellocchi. 2022. Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and*

- Bilingualism* 25, 6 (2022), 2229–2255. doi:10.1080/13670050.2021.1908220
- [46] Albert Newen, Leon De Bruin, and Shaun Gallagher. 2018. *The Oxford handbook of 4E cognition*. Oxford University Press. doi:10.1093/oxfordhb/9780198735410.001.0001
- [47] Mohammadali Noktehdan, Mehdi Shahbazpour, Mohammad Reza Zare, and Suzanne Wilkinson. 2019. Innovation Management and Construction Phases in Infrastructure Projects. *Journal of Construction Engineering and Management* 145, 2 (Feb. 2019), 04018135. doi:10.1061/(ASCE)CO.1943-7862.0001608
- [48] Donald A. Norman. 1988. *The Design of Everyday Things*. Basic Books, New York, NY, USA.
- [49] Jiyeon Oh and Jin-Woo Jeong. 2025. Exploring Interface Design of Translation System for Enhanced Immersion and Usability in Mixed Reality. In *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 719–720. doi:10.1109/ISMAR-Adjunct68609.2025.00156
- [50] Rebecca L. Oxford. 1990. *Language Learning Strategies: What Every Teacher Should Know*. Newbury House Publishers.
- [51] Riccardo Palmari, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torab-mostaedi. 2018. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing* 49 (Feb. 2018), 215–228. doi:10.1016/j.rcim.2017.06.002
- [52] Mahya Qorbani, Kamran Paynabar, and Mohsen Moghaddam. 2025. Teaching LLMs to See and Guide: Context-Aware Real-Time Assistance in Augmented Reality. doi:10.48550/arXiv.2511.00730
- [53] Antoni Rodríguez-Fornells, Ulrike M. Krämer, Urbano Lorenzo-Seva, Julia Festman, and Thomas F. Münte. 2012. Self-Assessment of individual differences in language switching. *Frontiers in Psychology* 2 (2012), 388. doi:10.3389/fpsyg.2011.00388
- [54] Cora Roelofs, Linda Sprague-Martinez, Maria Brunette, and Lenore Azaroff. 2011. A qualitative investigation of Hispanic construction worker perspectives on factors impacting worksite safety and risk. *Environmental health* 10, 1 (2011), 1–9. doi:10.1186/1476-069X-10-84
- [55] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented Reality: Principles and Practice*. Addison-Wesley Professional.
- [56] John Sermarini, Robert A Michlowitz, Lori C Walters, Roger Azevedo, et al. 2021. Embodied facade retrofit design with Augmented Reality in the built environment. In *Building Simulation Conference Proceedings*, Vol. 17.
- [57] John Sermarini, Robert A Michlowitz, Joseph J LaViola, Lori C Walters, Roger Azevedo, and Joseph T Kider. 2023. Investigating the impact of augmented reality and BIM on retrofitting training for non-experts. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4655–4665.
- [58] John Sermarini, Robert A Michlowitz, Joseph J LaViola, Lori C Walters, Roger Azevedo, and Joseph T Kider. 2024. Exploring augmented reality's role in enhancing spatial perception for building facade retrofit design for non-experts. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 233–243.
- [59] John Sermarini, Robert A Michlowitz, Joseph J LaViola, Lori C Walters, Roger Azevedo, and Joseph T Kider Jr. 2022. BIM driven retrofitting design evaluation of building facades. In *Proceedings of the 2022 ACM symposium on spatial user interaction*. 1–10.
- [60] Bokwang Shim, Suhyeon Seo, Sungil Oh, and Yangkyu Lim. 2025. Optimizing Real-Time Translation for AR Glasses: Enhancing Usability through Gesture and Content Algorithm. In *International Conference on Human-Computer Interaction*. Springer, 149–158. doi:10.1007/978-3-031-94150-4\_16
- [61] Ana Stanescu, Peter Mohr, Franz Thaler, Mateusz Kozinski, Lucchas Ribeiro Skreinig, Dieter Schmalstieg, and Denis Kalkofen. 2024. Error management for augmented reality assembly instructions. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 690–699. doi:10.1109/ISMAR62088.2024.00084
- [62] Jannis Strecker, Kimberly García, Kenan Bektaş, Simon Mayer, and Ganesh Ramanathan. 2023. SOCRAR: Semantic OCR through augmented reality. In *Proceedings of the 12th International Conference on the Internet of Things (IoT '22)*. ACM, New York, NY, USA, 25–32. doi:10.1145/3567445.3567453
- [63] Yiliu Tang, Jason Situ, Andrea Yaoyun Cui, Mengke Wu, and Yun Huang. 2025. LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, USA, Article 1054, 24 pages. doi:10.1145/3706598.3714224
- [64] Melissa K Thevenin and Jonathan W Elliott. 2015. Construction management students' mentors and role models: Developing a demographic profile. In *Proceedings of the 51st Annual International Conference of the Associated Schools of Construction*. <http://ascpro0.ascweb.org/archives/cd/2015/paper/CERT339002015.pdf>
- [65] Evan Thompson and Mog Stapleton. 2009. Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi* 28, 1 (2009), 23–30. doi:10.1007/s11245-008-9043-2
- [66] Sandhyarani Vadlamudi, KS Shashikala, MM Varun Chengappa, Mohammed Owais Arif, TY Bhargavi Devi, and Vidhya Siddhan. 2025. Augmented Reality Based Real Time Language Translation System. In *2025 2nd International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*. IEEE, 1–6. doi:10.1109/ICCAMS65118.2025.11234607
- [67] Francisco J. Varela, Eleanor Rosch, and Evan Thompson. 1992. *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, Cambridge, MA.
- [68] Lisa-Marie Vortmann, Pascal Weidenbach, and Felix Putze. 2022. AtAwAR translate: attention-aware language translation application in augmented reality for mobile phones. *Sensors* 22, 16 (2022), 6160. doi:doi.org/10.3390/s22166160
- [69] Anthony Webster, Steven Feiner, Blair MacIntyre, William Massie, and Theodore Krueger. 1996. Augmented reality in architectural construction, inspection and renovation. In *Proceedings of the 3rd Congress on Computing in Civil Engineering*, Vol. 1. New York, NY, American Society of Civil Engineers (ASCE), 913–919. <https://sites.cc.gatech.edu/fac/Blair.MacIntyre/papers/asce.pdf>
- [70] Xun Xu, Ling Ma, and Lieyun Ding. 2014. A Framework for BIM-Enabled Life-Cycle Information Management of Construction Project. *International Journal of Advanced Robotic Systems* 11, 8 (Aug. 2014), 126. doi:10.5772/58445
- [71] Huan Yang, Peng Wang, Yue Wang, Mark Billinghurst, Rong Luo, and Yi Zhang. 2025. Product assembly quality inspection of gap spacing using machine vision and human-centered AR visualization. *Computers & Industrial Engineering* (2025), 111531. doi:10.1016/j.cie.2025.111531
- [72] Guanlong Zhao, Sinem Sonsaat, Alif O. Silpachai, Ivana Lucic, Evgeny Chukharev-Hudilainen, John Levis, and Ricardo Gutierrez-Osuna. 2018. *L2-ARCTIC: A Non-Native English Speech Corpus*. Technical Report. University of Texas at Dallas, Perception and Sensing Lab. doi:10.21437/Interspeech.2018-1110
- [73] Yiran Zhao, Wenxuan Zhang, Guizhen Chen, Kenji Kawaguchi, and Lidong Bing. 2024. How do Large Language Models Handle Multilingualism?. In *Advances in Neural Information Processing Systems 37*. Neural Information Processing Systems Foundation, Inc., 15296–15319. doi:10.52202/079017-0489