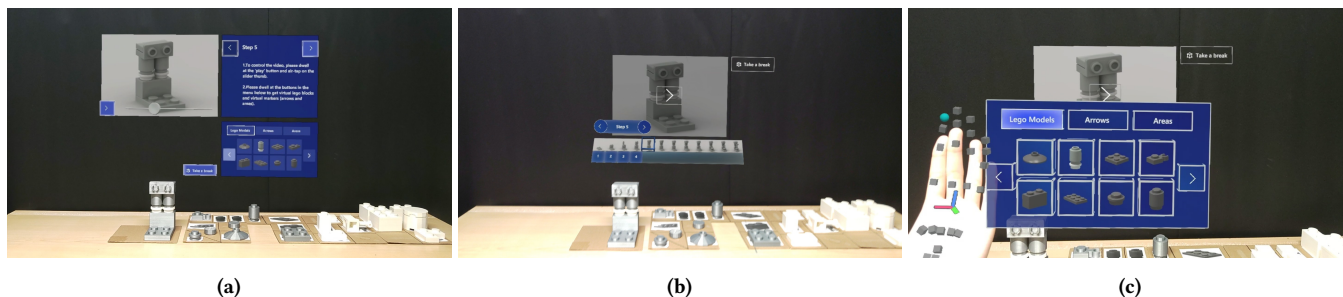




# On the Benefits of Image-Schematic Metaphors when Designing Mixed Reality Systems

Jingyi Li  
University of Cambridge  
Cambridge, United Kingdom  
jl2231@cam.ac.uk

Per Ola Kristensson  
University of Cambridge  
Cambridge, United Kingdom  
pok21@cam.ac.uk



**Figure 1:** (a) Baseline system. We designed the baseline based on Microsoft Dynamics 365 Guides, a Mixed-Reality (MR) solution that allows users to create and modify 3D instructional contents, and place them at real-world locations through an MR headset. (b) ISM-enhanced system. We enhanced the baseline system, in terms of its visual appearance, spatial placement, user interaction and interaction feedback, based on 12 image-schematic metaphors related to technology learning (see Table 1). (c) The redesigned toolkit menu in the ISM-enhanced system based on metaphor 2, 3, 4, 10, and 12 (see Table 1). The always-visible, fixed toolkit menu in the baseline was redesigned to be a palm-attached menu that can be shown by showing the palm to the face, or hidden by putting down the left hand.

## ABSTRACT

A Mixed Reality (MR) system encompasses various aspects, such as visualization and spatial registration of user interface elements, user interactions and interaction feedback. Image-schematic metaphors (ISMs) are universal knowledge structures shared by a wide range of users. They hold a theoretical promise of facilitating greater ease of learning and use for interactive systems without costly adaptations. This paper investigates whether image-schematic metaphors (ISMs) can improve user learning, by comparing an existing MR instruction authoring system with or without ISM enhancements. In a user study with 32 participants, we found that the ISM-enhanced system significantly improved task performance, learnability and mental efficiency compared to the baseline. Participants also rated the ISM-enhanced system significantly higher in terms of perspicuity, efficiency, and novelty. These results empirically demonstrate multiple benefits of ISMs when integrated into the design of this MR system and encourage further studies to explore the wider applicability of ISMs in user interface design.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Mixed / augmented reality**; **HCI theory, concepts and models**.

## KEYWORDS

Image Schema, Mixed Reality, Instruction Authoring

## ACM Reference Format:

Jingyi Li and Per Ola Kristensson. 2024. On the Benefits of Image-Schematic Metaphors when Designing Mixed Reality Systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3613904.3642925>

## 1 INTRODUCTION

Everyday end-users usually find it challenging to interact with Mixed Reality (MR) systems, given that the interaction demands high motor skills, spatial cognition, management of attention and conceptualization of unfamiliar computing concepts [49]. In order for an MR system to achieve commercial success, it needs to be easy to learn and use for users from experts to everyday end-users with lower levels of skill and expertise [13]. While extensive adaptations could be costly, there is clearly a need for design approaches that can make an MR system easy to learn and use for a wide range of users with varying skill levels.

One promising candidate for such design approaches is the concept of image-schematic metaphors (ISMs). These ISMs are almost universally understood, as they are grounded in basic sensorimotor experiences that are common to most people. Thus, users of ISMs



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0330-0/24/05  
<https://doi.org/10.1145/3613904.3642925>

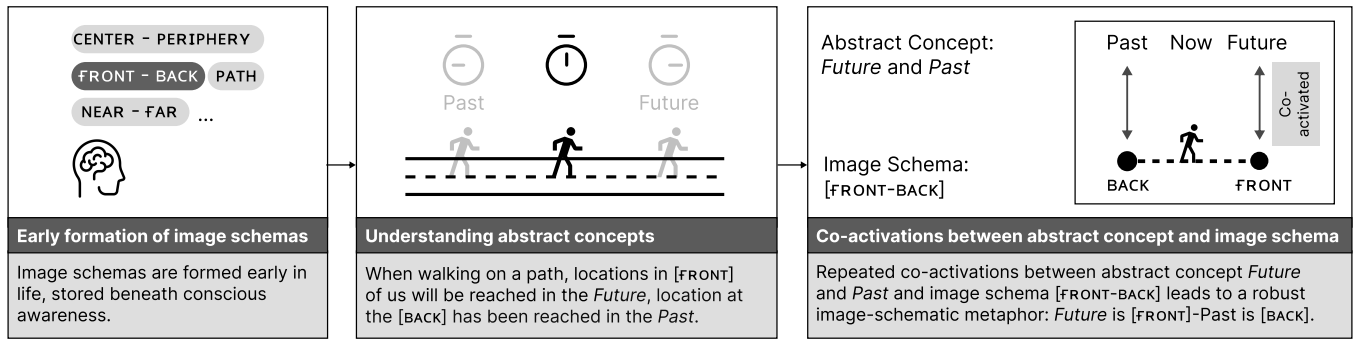


Figure 2: The formation of an image-schematic metaphor “*Future* is FRONT”.

should effortlessly navigate an MR system whose components are designed in alignment with these ISMs.

An image-schematic metaphor (ISM) is a metaphor that uses a set of basic cognitive structures derived from universal sensory experiences to understand abstract concepts. For example, a documented ISM “*Future* is FRONT-*Past* is BACK” uses a basic cognitive structure FRONT-BACK to understand abstract concepts *Future* and *Past* (see Figure 2). The set of basic cognitive structures we use as analogue references are image schemas [30, 34], which are multidimensional and multimodal gestalts derived from repeated sensory experience. As an infant, we develop these image schemas like FRONT-BACK, which is usually the side with the face in humans and the side opposite [42]. Then these image schemas become the foundation of the human conceptual system and are used to understand other new concepts [11]. When encountering abstract concepts like *Future* and *Past*, an image schema such as FRONT-BACK is subconsciously activated. This is because in the everyday experience of walking on a path, the locations in FRONT of us will be reached in the *future* and the locations at the BACK has been reached in the *past*. The repeated co-activations between an abstract concept and an image schema leads to a permanent and robust ISM [18]. There are a large number of such robust ISMs (e.g., “*Future* is FRONT-*Past* is BACK”, “*Similarity* is NEAR-*Difference* is FAR”) that are widely adopted by people. They operate subconsciously as a form of prior knowledge, and are easily accessible to the mind.

For users, prior knowledge is a robust indicator of performance and learnability. Using ISMs as a foundation when designing various aspects of an MR system could yield better task performance and learnability among everyday end-users as these ISMs can serve as a type of easily accessible and near-universal prior knowledge that users can apply when learning to use the system. For designers, ISMs can offer concrete design guidance by describing users’ mental model of a certain concept. For instance, “*Memory* is CONTAINER” describes users’ mental model of *Memory* as an image schema CONTAINER. In a design, *Memory* can then have the appearance, spatial placement, or behaviors of a CONTAINER.

ISMs could be leveraged as a valuable middle ground for designing MR systems that are easy to learn and use for everyday end-users without the need of extensive adaptations. However, the impacts of such an approach has never been examined in MR. Prior studies using screen-based GUIs [24, 28, 55] and embodied interaction [2, 41] have investigated the benefits of integrating ISMs.

Given the alignment between the bodily nature of ISMs and the inherent body-centered nature of MR interactions, MR naturally lends itself to effective implementations of ISMs. Such an alignment could potentially amplify the power of embodied metaphorical understandings. We hypothesize that integrating ISMs into MR system design will enhance task performance, learnability, mental efficiency, and perceived efficiency/learnability/novelty. To test these hypotheses in MR, we integrated 12 ISMs (see Table 1) related to technology learning that were elicited from contextual inquiry with 24 participants [40] into four aspects of an MR instruction authoring system, including the visual appearance and spatial registration of user interface elements, user interaction, and interaction feedback. Then we compared the ISM-enhanced system to a widely adopted baseline that offered the same instruction authoring functions but did not conform to the 12 metaphors. Our evaluation showed that in comparison to the baseline, the ISM-enhanced condition enabled participants to complete the same instruction authoring task significantly faster with a lower error rate. In addition, participants exhibited markedly enhanced learnability and mental efficiency, and rated the ISM-enhanced system significantly higher in terms of perceived efficiency, perspicuity, and novelty. For the first time, we have implemented ISMs in user interface elements and user interactions of an MR instruction authoring system and empirically demonstrated multiple benefits of such an approach. For these empirical benefits to be generalizable, future work needs to further investigate the impacts of ISMs as a design approach across different tasks, designers, and forms of user interfaces.

In summary, this paper makes the following contributions:

- (1) A demonstration of a novel ISM-based approach for designing user interface elements and user interactions of an MR system, and an empirical investigation on the benefits of this approach.
- (2) The presentation of an enhanced-version of an existing MR instruction authoring system using 12 ISMs related to technology learning.
- (3) The results of an empirical evaluation comparing the performance of the ISM-enhanced system against an existing baseline, in which we find that the ISM-enhanced system yielded a significant increase in speed, learnability and mental efficiency, a decrease in error rate, and a significant increase in perceived efficiency, perspicuity, and novelty.

## 2 RELATED WORK

### 2.1 Usability Guidelines for Mixed Reality

Research efforts have been made in establishing design guidelines that facilitate usability of MR systems. On one end of the spectrum, there is an abundance of guidelines and principles that are general in nature. On the other end, there exists some highly-specific design instructions and best practices, which go into details about aspects of interface and interaction design. What is noticeably missing is a middle ground that addresses the mental models of users, which sheds light on how users might understand and interact with different concepts in an MR system.

**2.1.1 General Design Objectives.** An MR system combines graphical user interfaces (GUIs), gesture interfaces and audio interfaces, overlaying them onto the physical reality [14]. Consequently, the design of MR systems can potentially draw insights from well-established GUI principles [3, 15, 48]. For instance, Apple’s Macintosh Human Interface Guidelines [3] suggest a set of principles, including Metaphors, Reflect the users’ mental model, Explicit and implied actions, Direct manipulation, User control, Feedback and communication, Consistency, What you see is what you get, Forgiveness, Perceived stability, Aesthetic integrity, Modelessness, and Managing complexity.

Prior researchers have also proposed a few MR-specific design guidelines [10, 13, 52] which derive principles that fit MR-specific requirements from literature on user interface design, tactile interface design and usability. For instance, Ciccarelli et al. [10] elicited a set of design principles from traditional user interface design guidelines that most apt for the development of AR/MR systems, and categorized them into three different classes: user interaction (e.g., usability and learnability), performance (e.g., efficiency of use), and application content (e.g., easy reversal of actions). While the existing MR design guidelines are good at describing a desirable outcome, they fall short at offering specific means to achieve it.

**2.1.2 Specific Best Practices.** Some well-known GUI guidelines provide detailed guidance on how to organize information and design fundamental system components, such as buttons, menus, text fields, and sliders [3, 15, 48]. For example, Galitz [15] suggests that the the most important elements should be placed at the top left corner.

There are also some specific best practices in MR system design provided by key industry players. The Microsoft Mixed Reality Toolkit provides various sample user interface elements, such as buttons, cursors, sliders, and bounding boxes, as well as usability guidance on how to use these user interface elements (for example, the appropriate size and color of a button). Apple’s Human Interface Guideline have evolved to include specific design recommendations for different interaction platforms [4]. It provides a set of MR-specific best practices for designing layouts, menus, navigation and search, content presentation, and selection and input methods in visionOS. For example, a horizontal slider is preferable in MR as it is generally easier for users to perform a side-to-side gesture. Meta Quest [43] has also established specific recommendations on multiple aspects of virtual and mixed reality design, such as information display, user experience, and hand gestures. For example, it is recommended to design user interfaces that rotate

to always face the user, regardless of their orientation—increasing readability and usability. While best practices like these are undoubtedly useful for guiding the implementation of MR system designs, they are so specific that they often do not lend themselves to more conceptual considerations in design.

As alluded to earlier, there is lack of design approaches that are neither as general as design objectives nor as specific as best practices, which help designers understand users’ mental models and guide them in creating designs that align with these mental models. The idea of image-schematic metaphors is one promising candidate for such design approaches, which can effectively translate users’ mental model into basic cognitive constructs.

### 2.2 Image Schemas and Image-Schematic Metaphors

**2.2.1 Image Schemas.** Image schemas, also known as embodied schemas, are highly abstract representations of recurring bodily experiences [30]. For example, the image schema *PATH* emerges when we observe an object moving along a trajectory in space; the image schema *BLOCKAGE* stems from our experiences with impeded motions; the image schema *CENTER-PERIPHERY* is formed when we experience bodily relations between the trunk and the extremities. The grounding in visual, haptic, acoustic, and kinaesthetic experiences makes image schemas multimodal. As a result, once formed, an image schemas can be instantiated in different modalities [29]. For example, a designer can instantiate the image schema *BLOCKAGE* visually in a greyed out button, haptically through vibration feedback, or aurally using a beep. The multimodal characteristic allows image schemas to be generally applicable to a broad range of interaction paradigms with different modalities and features.

**2.2.2 Image-Schematic Metaphors.** Once formed, image schemas are subconsciously used to understand other abstract concepts that lacks physical or conceptual referents [34]. The repeated co-activation between an abstract concept and an image schema leads to an image-schematic metaphor (ISM) [18]. An example of an ISM is “*Importance* is *CENTER-Unimportance* is *PERIPHERY*” [35], which conceptually associates the abstract concepts *Importance* and *Unimportance* with image schema *CENTER-PERIPHERY*. This ISM stems from the experiences that our central body parts are vital, while peripheral parts (e.g., hair and nails) can be removed without causing harm [25]. These experiences lead to a permanent co-activation between *Importance-Unimportance* and *CENTER-PERIPHERY*.

Triggered by recurrent experiences, the operation of ISMs become subconscious over time, rendering them effortlessly accessible to the mind [27]. These ISMs are subconsciously instantiated in different domains to understand new concepts [5, 37]. For example, when using the touchscreen on an ATM, we tend to look for important buttons in the central zone of the screen. Integrating such ISMs into system design can make the system operate in a way that aligns with users’ mental model.

An ISM differs from the conventional view of a *metaphor*, which considers a metaphor to be a novel or poetic linguistic expression [36]. ISMs concentrate on how individuals conceptualize a target domain (e.g., the level of *Importance*) in terms of an image schema (e.g., *CENTER-PERIPHERY*), rather than on language. Thus, from a language perspective, some ISMs do not appear to

be “metaphorical”, but they are good at translating users’ mental models of abstract concepts to specific image schemas in a concrete way. Additionally, these translated image schemas are multimodal and multidimensional gestalts, which provides designers with great flexibility in their implementations.

## 2.3 HCI Studies Involving ISMs

**2.3.1 Empirical Investigation on ISMs’ Benefits.** ISMs are instantiated in people’s spoken language and behaviors. It is therefore possible for designers to elicit ISMs for a specific design context. There is established guidance in the literature [51] on eliciting ISMs from transcribed contextual interviews, with a relatively small sample size of 4–12 participants [23].

Given their accessibility, ISMs have served as design inspiration for tangible user interface (TUI) [6], virtual reality interaction [12] and a mobile augmented reality system [16]. These studies, however, did not investigate the benefits or impacts of integrating ISMs in design; rather, they were more proof-of-concept demonstrations of the feasibility of doing so. Other studies in screen-based GUIs [24, 28, 55] and embodied interaction [2, 41] investigated the benefits of integrating ISMs through empirical comparisons. In these screen-based GUI studies, the ISM-congruent interfaces enhanced efficiency [24, 28], perceived suitability [24], perceived intuitiveness [28], perceived innovativeness [55] and reduced mental load [28]. In these embodied interaction studies, the metaphorical condition slightly outperformed (although not significantly) the non-metaphorical condition(s) in terms of effectiveness [41], efficiency [2, 41], perceived learnability [2] and perceived intuitiveness [2].

In the taxonomy for embodied learning put forward by Johnson-Glenberg et al. [31], MR is placed at the highest end of the embodiment taxonomy, as it incorporates a large amount of motoric engagement, gesture-action congruence, and strong perception of immersion. In contrast, interactive screen-based systems are positioned at the lower end of this taxonomy. ISMs aid cognitive processes through embodied metaphorical understanding—using fundamental embodied experience to understand novel concepts. Providing stronger forms of embodiment, MR could add to the power of embodied metaphorical understanding beyond what is known to be possible with screen-based systems. This motivates our work in exploring the benefits that ISMs might offer when implemented in MR systems.

**2.3.2 Hypothesized Benefits in MR systems.** We hypothesize that integrating ISMs in the components of an MR system could enhance *task performance, learnability, mental efficiency, perceived efficiency, perceived learnability, and perceived novelty*. An MR system can impose two types of cognitive load on its users: intrinsic load, caused by learning to perform the actual task (e.g., way-finding); and extraneous load, caused by learning to navigate the MR system (e.g., using an MR system to find ways) [50]. The comprehension of ISM-congruent system components can occur rapidly beneath conscious awareness [24, 37]. This optimizes the distribution of a user’s cognitive resource between intrinsic and extraneous loads, allowing a user to distribute more cognitive capacity to handling the task at hand, instead of wasting cognitive resource in figuring out how the system works. We hypothesize the additional cognitive

capacity made available by ISMs could enable users to learn the task faster and perform it better. Additionally, the optimized distribution of cognitive resources could enhance mental efficiency [1]—with the same amount of cognitive resources available, the ISM-based system can lead to better task performance. From the perspective of user experience, integrating ISMs, which underlie users’ mental models, could remove potential mental interruptions caused by mismatches between users’ mental models and how the system works, leading to a “fast” or “smooth” experience. We hypothesize of enhancement in perceived learnability because the integrated ISMs can serve as a type of accessible prior knowledge that facilitates the understanding of new concepts [37]. Finally, we hypothesize MR systems designed based on ISMs to be perceived more novel as we believe that the cross-modal nature of ISMs could encourage designers to consider implementing these ISMs creatively in different aspects of an MR system.

## 3 ISM-ENHANCED MIXED REALITY INSTRUCTION AUTHORIZING

### 3.1 Instruction Authoring in Head-Mounted Mixed Reality

To test the hypothesized benefits of ISMs in the design of MR systems, we chose MR instruction authoring as the design context in this research. In manufacturing, headset-based MR instructions has been developed to assist line workers in procedural tasks, such as manual assembly and maintenance [17, 20, 33, 53, 54]. Referring to traditional paper- or video-based instructions requires a constant switch of attention between the instructions and the context where users are performing the assembly task [32]. Headset-based MR instructions effectively address this challenge by directly superimposing virtual 3D instructions onto the real-world assembly context in a step-by-step manner. However, it can be challenging to translate an existing video tutorial for assembling a model to an MR tutorial that overlays the correct virtual part on the physical model step-by-step, as creating such MR instructions using current authoring tools largely demand programming and graphical expertise [46]. Only a limited number of tools (e.g., Microsoft Dynamics 365 Guides<sup>1</sup> and WAAT [7]) have attempted to allow users without these skills to directly create and modify virtual 3D assembly parts, and place these virtual parts at real-world locations for users wearing MR Head-Mounted Displays (HMDs). This can pose a challenge for MR novices since the problems they encounter when learning to use the MR system may impair their performance on instruction authoring. We hypothesize that integrating ISMs into the design of an head-mounted MR instruction authoring system can be a solution to this challenge.

### 3.2 Baseline System

The baseline system was designed based on Microsoft Dynamics 365 Guides<sup>2</sup>, a widely adopted MR instruction authoring tool. Using the baseline system, users (e.g., a line manager) can translate an existing video tutorial for assembling a model to an MR tutorial by creating and modifying virtual assembly parts and place the

<sup>1</sup><https://learn.microsoft.com/en-us/dynamics365/>

<sup>2</sup><https://learn.microsoft.com/en-us/dynamics365/>

**Table 1: We used the following 12 image-schematic metaphors that were identified by the authors in [40] to enhance the baseline system design. The interpretation for each ISM was elaborated based on the explanation of its image schema, and the explanation of each image schema was based on the ISCAT database [26].**

Image-Schematic Metaphor	Image Schema Explanation	ISM Interpretation
1. <i>Learning</i> is MATCHING	MATCHING: Corresponding in pattern, color, or design.	Learning is based on correspondence in pattern, appearance or design between learning objectives and examples.
2. Important <i>Information</i> is CENTER	CENTER: The midpoint of an object or a space.	Important information is located at the center of FOV (Field of view).
3. The position of <i>Video Tutorial</i> is PERIPHERY - Task is CENTER	CENTER: The midpoint of an object or a space. PERIPHERY: Parts that are on the outside or margins.	The ongoing task centers the FOV, while video tutorial locates at the periphery.
4. <i>Button</i> exerts COMPULSION to the system	COMPULSION: An external force that physically or metaphorically causes an entity to move or action.	Users compel a system to execute certain functions by pressing a button.
5. The form of <i>Step</i> is PART	PART: One individual segment of a larger organization.	The form of step is one individual segment within a larger system or structure.
6. The form of <i>Logic</i> is PATH 7. The form of <i>Operation</i> is PATH 8. The form of <i>Video Tutorial</i> is PATH 9. The form of <i>Step</i> is PATH	PATH: A PATH consists of a starting point, an end-point, and a sequence of contiguous locations connecting the two points.	The form of logic/operation/video tutorial/step is a path that consists of a starting point, an end-point, and a sequence of contiguous locations connecting the starting point with the end-point.
10. The position of <i>Video Tutorial</i> is FAR - Task is NEAR	NEAR-FAR: The spatial proximity or distance of entities.	The ongoing task locates closer to the user in their FOV, while video tutorial is further.
11. Extraneous <i>Information</i> is BLOCKAGE	BLOCKAGE: A blocking force that stops or redirects a force/movement.	Extraneous elements is a blocking force that stops or redirects an action attempted by a user.
12. The position of Hidden <i>Information</i> is DOWN	DOWN: Being vertically lower in space.	Information is hidden at (and can be retrieved from) the lower part of interaction space.

correct part at the correct real-world location at each step, through an MR head-mounted display. The baseline system consists of three components, a video tutorial that needs to be translated to a step-by-step MR tutorial, a toolkit menu where users can generate virtual assembly parts, and a step card that manages the steps of the MR tutorial currently being created. Users can interact with the baseline system through an MR headset using three gestures: Grab and Air-Tap for virtual object manipulation, and Hand-ray dwell for button selection.

### 3.3 Enhancing the Design of an Existing Baseline with ISMs

We integrated a set of 12 ISMs (see Table 1), which were not originally integrated in the baseline system, to enhance its design. This resulted in an ISM-enhanced system. The following sections discuss how the 12 ISMs were selected and how they were used to enhance the baseline system.

**3.3.1 Selection of Image-Schematic Metaphors.** In a prior ISM elicitation study with 12 younger and 12 older participants, the authors [40] elicited a set of 37 “universal” ISMs used by both younger and older participants related to technology learning (see Table 6). These ISMs were derived from interviews around participants’ experiences of navigating unfamiliar systems, as well as their interaction behaviors when using unfamiliar systems (three screen-based systems and one tangible user interface). Although the tasks from

which these ISMs were derived from were not MR-related, these ISMs are high-level enough to not carry any task- or context-specific attributes (e.g., The form of *Logic* is PATH), hence there are no constraints on what type of interactive systems they can be instantiated into.

We used the 37 ISMs as a checklist to analyze the baseline in terms of its *spatial placement*, *component appearance*, *interaction feedback* and *user input*, and then tagged each of the 37 ISMs as “incongruent” (Metaphor 1–12 in Table 6), “congruent” (Metaphor 13–19), or “irrelevant” (Metaphor 20–37). The baseline was congruent with seven of these ISMs, which were commonly seen in general interface design, for example, “General *Information* is UP - Detailed *Information* is DOWN”. On the other hand, the baseline was incongruent with 12 ISMs which are shown in Table 1. We used the 12 incongruent ISMs, which were not integrated in the baseline, to enhance its design, resulting in an ISM-enhanced system. We note that the integration of the 12 incongruent ISMs did not impair the instantiations of the congruent ISMs, which now exist in both the baseline and the ISM-enhanced system.

**3.3.2 Redesigning the Baseline using 12 ISMs.** The 12 ISMs in Table 1 were used to enhance the following components of the baseline system. Following a “System Component – ISM – Input/Output Instantiations” design pathway (see Figure 9), we first linked system components (e.g., video tutorial) to related ISMs, then considered whether each of these ISMs can be instantiated to the following

**Table 2: We redesigned the baseline following a “System Component – ISM – Instantiations” design pathway. Firstly, we listed six functional components that made up the MR instruction authoring system. Then, for each system component, we selected relevant ISMs from the 12 listed in Table 1. Lastly, we associated each selected ISM with one or more design aspects of the MR instruction authoring system where it could be instantiated. These aspects include visual appearance, spatial placement, audio feedback, and interaction. For instance, the pathway “Step Card – 5. The form of *Step* is PART – Visual Appearance” suggests that the ISM “The form of *Step* is PART” could be integrated in the appearance of the Step Card by visually presenting each step as one smaller segment of a larger graphical element. For explanations of each ISM and the image schema in this ISM, please refer to Table 1.**

System Components	Relevant ISMs	Visual Appearance	Spatial Placement	Audio Feed-back	Interaction
Video Player	8. The form of <i>Video Tutorial</i> is PATH 9. The form of <i>Step</i> is PATH 5. The form of <i>Step</i> is PART	✓ ✓ ✓			✓ ✓ ✓
Toolkit Menu	11. Extraneous <i>Information</i> is BLOCKAGE 2. Important <i>Information</i> is CENTER 10. The position of <i>Video Tutorial</i> is FAR - Task is NEAR 3. The position of <i>Video Tutorial</i> is PERIPHERY - Task is CENTER 12. The position of Hidden <i>Information</i> is DOWN		✓ ✓ ✓ ✓		✓ ✓  ✓
Step Card	7. The form of <i>Operation</i> is PATH 9. The form of <i>Step</i> is PATH 5. The form of <i>Step</i> is PART 1. <i>Learning</i> is MATCHING	✓ ✓ ✓ ✓	✓		
Instruction Style	1. <i>Learning</i> is MATCHING	✓			✓
Holograms Manipulation	7. The form of <i>Operation</i> is PATH			✓	
Button Selection	4. <i>Button</i> exerts COMPULSION to the system	✓			✓

aspects: *visual appearance, spatial placement, audio feedback, user interaction* (See Table 2). Table 3 compares the same system components in the baseline and the ISM-enhanced system, and lists the ISMs used for the redesign.

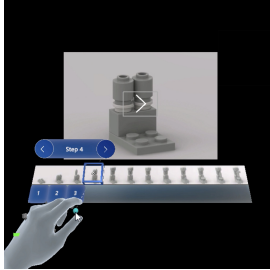
*Video Tutorial.* In the baseline, the video tutorial is a “window” that displays one video frame at a time in sequence, with a standard playback slider at the bottom. Users can drag and move the slider thumb to go back and forth in the video. We redesigned the visual appearance and interaction of the video tutorial using metaphor 5, 8, and 9 in Table 1. These ISMs indicate that the form of a video tutorial is a long path (metaphor 8) made up by a sequence of steps (metaphor 5), and each of these steps is a shorter path (metaphor 9) within this longer path. Therefore, the playback slider of the video tutorial was visually redesigned to be a path-like sprite sheet consisting of multiple smaller images that corresponded to each assembly step of the video tutorial. Users can click on any image on the sprite-sheet slider to pause the video at the corresponding assembly step.

*Toolkit Menu.* In the baseline, the toolkit menu is always visible and fixed at the bottom right of a user’s field of view. We redesigned the visibility and spatial placement of the toolkit menu using metaphor 2, 3, 4, 10, and 12. Metaphor 4 and 12 suggests that extraneous elements that are not in use will block users’ actions; hence, these unused elements could be hidden at the bottom of

the interaction space. Metaphor 2, 3, 10 suggest that important elements required by the current task should be at the center a user’s field of view and be closer to the user than other system elements. Taking into account these redesign requirements, we designed a toolkit menu attached to a user’s left palm. Whenever a user needs the toolkit menu, they hold up their left palm and the menu would appear close up and right in front of the user. When not in use, the menu would be hidden when the user put down their left hand. To rule out the impact of menu hierarchy on interaction time, the ISM-enhanced toolkit menu preserved the same hierarchy and tab structure as the baseline version.

*Step Card.* In the baseline, the Step card consists of a “go back” button, a “go next” button, and a text field in between showing the current step number of the MR tutorial in creation (e.g., step 1). We redesigned the visual appearance and spatial placement of the step card using metaphor 5, 7, and 9. These metaphors suggests that the whole operation should be a long path (metaphor 7) consisting of a sequence of steps (metaphor 5), and these steps are shorter paths (metaphor 9) within the longer path. The appearance of the step card was redesigned to be a visual path consisting of multiple grids, with each grid corresponding to one assembly step of the whole assembly operation. This is visually similar to the sprite-sheet playback slider, which is also a longer path consisting of the same number of shorter paths. Metaphor 1 suggests that the similarities and correspondence

**Table 3: A comparison between system components of the baseline and the ISM-enhanced system. For detailed explanations on how these image-schematic metaphors were instantiated in the ISM-enhanced system, please see section 3.3.2.**

	Baseline System	ISM-Enhanced System	Image-Schematic Metaphors
Video Tutorial			<p>The form of <i>Video Tutorial</i> is PATH</p> <p>The form of <i>Step</i> is PATH</p> <p>The form of <i>Step</i> is PART</p>
Toolkit Menu			<p>Extraneous <i>Information</i> is BLOCKAGE</p> <p>Important <i>Information</i> is CENTER</p> <p>The position of <i>Video Tutorial</i> is FAR - Task is NEAR</p> <p>The position of <i>Video Tutorial</i> is PERIPHERY - Task is CENTER</p> <p>The position of Hidden <i>Information</i> is DOWN</p>
Step Card			<p>The form of <i>Operation</i> is PATH</p> <p>The form of <i>Step</i> is PATH</p> <p>The form of <i>Step</i> is PART</p> <p><i>Learning</i> is MATCHING</p>
Instruction Style			<p><i>Learning</i> is MATCHING</p>
Holograms Manipulation			<p>The form of <i>Operation</i> is PATH</p>
Button			<p><i>Button</i> exerts COMPULSION to the system</p>

in patterns or appearance between two objects can enable people to learn one object based on another. Based on this metaphor, we aligned the path-like step card directly underneath the sprite-sheet playback slider to enable spatial correspondence. The redesign of the step card could allow users to match the current assembly step of the MR tutorial in creation to the assembly step currently shown in the video tutorial, making it easier to track progress.

*Instruction Style.* In the baseline, the Step card displays text-form instructions on how to interact with the system. We used metaphor 1 to redesign the form of instruction. Metaphor 1 suggests that learning is based on similarity between instructions and learner actions. The more similar they are, the easier learning is. Learner actions in this task are hand gestures. Instead of using text to instruct users how to perform a hand gesture, we redesigned the form of instruction to be semi-transparent 3D hand animations that demonstrate the interaction gestures. These animations only played twice: when the user initiated the task, and during the task if no visible hands were detected for more than ten seconds.

*Manipulation Feedback.* In the baseline, there is no audio feedback when users manipulate virtual objects. We added audio feedback for these manipulation behaviors based on metaphor 7, which suggests that a manipulation behavior is a path that consists of a starting point and an endpoint. This image schema can be implemented not only visually but also audibly. We implemented a starting sound and an ending sound signaling the start and end of every manipulation behavior.

*Button Selection Gesture.* To select a button in the baseline, users hold the hand ray on the button for a short period of time until it is selected. Strictly following Microsoft's Design Guidelines on dwell feedback<sup>3</sup>, the onset delay after users started a hand-ray dwell action was set at 250 ms, and the time to complete a dwell action was set at 850 ms.

Metaphor 4 suggests that users force the system to function by pressing a button. Inspired by this metaphor, we redesigned the appearance and selection method of all buttons. All buttons in the ISM-enhanced system were Pressable Buttons from Microsoft Mixed Reality Toolkit (MRTK)<sup>4</sup>, which can be triggered by both Touch and Air-tap. When the pressable buttons were triggered, the front plate of the button retracted in response to the interaction, offering a visual representation of the image schema COMPULSION.

## 4 EVALUATION: COMPARING THE ISM-ENHANCED SYSTEM AGAINST THE BASELINE

We carried out a between-subject user study to evaluate our proposed system enhanced with 12 ISMs by comparing it with a baseline that did not conform to these ISMs. This study employed a between-subject experimental design, where participants were exposed to one of two conditions: either the baseline system or the ISM-enhanced system. The motivation for using a between-subjects design stemmed from the proposition that the integration of ISMs

<sup>3</sup><https://learn.microsoft.com/en-us/windows/mixed-reality/design/gaze-and-dwell-eyes>

<sup>4</sup><https://learn.microsoft.com/en-us/windows/mixed-reality/mrkt-unity/mrkt2/features/ux-building-blocks/button?view=mrktunity-2022-05>

might better aid users in understanding the materials and the task. Employing a within-subject design would introduce a very high risk of an asymmetrical skill-transfer effect [47], which would unfairly penalize the performance of the ISM-enhanced condition if it follows the baseline condition and unfairly favor the baseline condition if it follows the ISM-enhanced condition. To mitigate this confounding factor, a between-subject design was used, resulting in an enlarged sample size for the experiment ( $n = 32$ ).

### 4.1 Participants

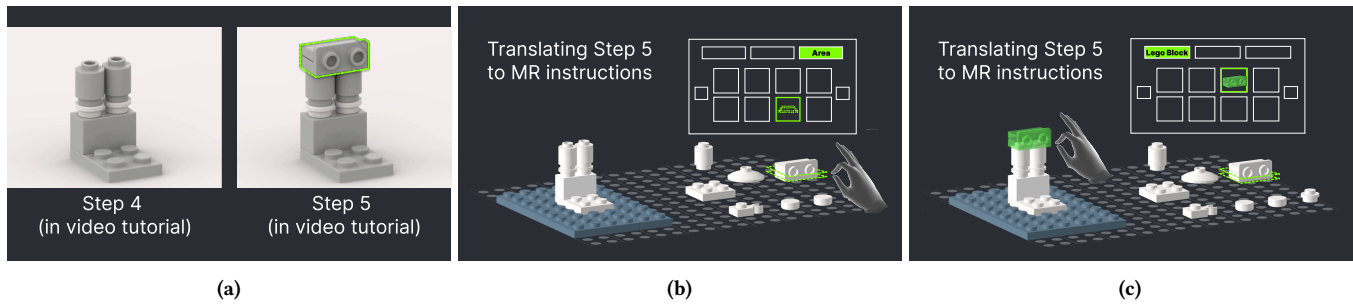
We recruited 32 participants via convenience sampling (average age = 25.78, standard deviation = 4.46, min = 18, max = 37; 11 females, 21 males). The baseline group and the ISM-enhanced group were gender balanced to the extent possible. A Mann-Whitney  $U$  Test with a significance level of  $\alpha = 0.05$  was no significant difference between the ages of participants in the two groups ( $U = 125$ ,  $p = 0.9283$ ). We measured participants' technology experience at two levels, exposure and competency, using a questionnaire. At the exposure (general technologies) level, participants reported how often they had used the 12 common technologies in everyday life, on a five-point Likert scale (from "Never" (1) to "Always" (5), range of total score: 12 – 60). At the competency (general technologies) level, we used the Computer Proficiency Questionnaire (CPQ) [9] to measure participants' competence in performing 33 common digital tasks on a five-point Likert scale (from "Never Tried" (1) to "Very Easily" (5), range of total score: 33 – 165). A Mann-Whitney  $U$  Test with a significance level of  $\alpha = 0.05$  indicates no significant difference between the technology experiences of the participants assigned to the two conditions ( $U = 109$ ,  $p = 0.4839$ ). A Levene's test indicates that the two participant groups have similar variances in technology experiences ( $F_{1,30} = 0.2310$ ,  $p = 0.6343$ ).

### 4.2 Apparatus

Both the baseline system and the ISM-enhanced system were developed and deployed on the Microsoft HoloLens 2 using Unity and the Microsoft Mixed Reality Toolkit. Throughout the experiment, the participants maintained a seated position while wearing the HoloLens 2 device. We used the Holographic Remoting application on the HoloLens device to create a connection with Unity 3D running on a Windows 11 workstation. The configuration facilitated a distant observation of the content exhibited on the HoloLens display, enabling the experimenter to deliver instructions throughout the training sections. The interface aesthetics of both systems followed the default design of the Microsoft Mixed Reality Toolkit (MRTK) examples.

### 4.3 Tasks

In both the baseline and ISM-enhanced conditions, participants were asked to translate an existing video tutorial for assembling a 12-part Lego model to a 12-step MR instruction, with each step showing the correct virtual Lego block at the correct real-world location. The participants were seated in front of a table, on which there was an assembly plate on the left and a pick-up area with 12 Lego blocks on the right. Each Lego block was placed on a grid bearing a printed image of itself.



**Figure 3: Visual demonstrations of how participants in this experiment translated step 5 of the video tutorial to MR instructions. (a) Refer to the video tutorial to identify the correct block and how it should be assembled. (b) Get a virtual maker in the toolkit menu and anchor it to the correct block. (c) Get a virtual Lego block and anchor it to the correct assembly position.**

To translate each assembly step in the video tutorial to MR (see Figure 3), participants were required to first refer to the video tutorial and identify the correct Lego block for the current step; use any virtual markers (arrows or areas) from the toolkit menu to indicate which block should be picked up from the real-life pick-up area; then, get the corresponding virtual Lego block from the toolkit menu and place it at the correct real-world location to indicate where (location) and how (orientation) this block should be assembled. After finishing each step, participants moved on to the next step using the step card function. The 12 steps were 12 trials of similar complexity. Participants were instructed to perform the authoring task as accurately and as quickly as possible. We measured performance as the completion time and error rate for each trial.

#### 4.4 Procedure

32 participants were divided into two gender-balanced groups, one group used the baseline system and the other used the ISM-enhanced system. The experiment consisted of a familiarization phase and a test phase.

**4.4.1 Familiarization Phase.** At the start of the experiment, we gave participants a brief introduction to the HoloLens 2 and an overview of the authoring gestures. Participants were given the opportunity to practice these gestures. Then participants were introduced to their experimental condition and then performed instruction authoring tasks for a simple 5-part Lego model. The Lego model used in the familiarization phase was different from the one used in the test phase.

**4.4.2 Test Phase.** After a short break, participants proceeded with the test phase, where they used the same system in the familiarization phase to complete the instruction authoring task for a 12-part Lego model. After the testing section, participants filled out the NASA Task Load Index (NASA-TLX), User Experience Questionnaire (UEQ) [39], and Technology Proficiency Questionnaire. The UEQ comprises a total of 26 items, with each item containing pairs of contrasting attributes that may apply to the system. For instance, one pair of attributes is “easy to learn” versus “difficult to learn”. We employed the UEQ to evaluate user experience that both systems provide at six scales, including attractiveness, perspicuity, efficiency,

dependability, stimulation, and novelty [21]. These scales can be further categorized into aggregated indicators, like pragmatic quality and hedonic quality.

## 5 RESULTS

All data was checked against assumptions for the statistical tests used. For normally distributed data, we used Welch’s  $t$ -test for two-level comparisons. For non-normally distributed data, we used Mann-Whitney  $U$  test for two-level comparisons.

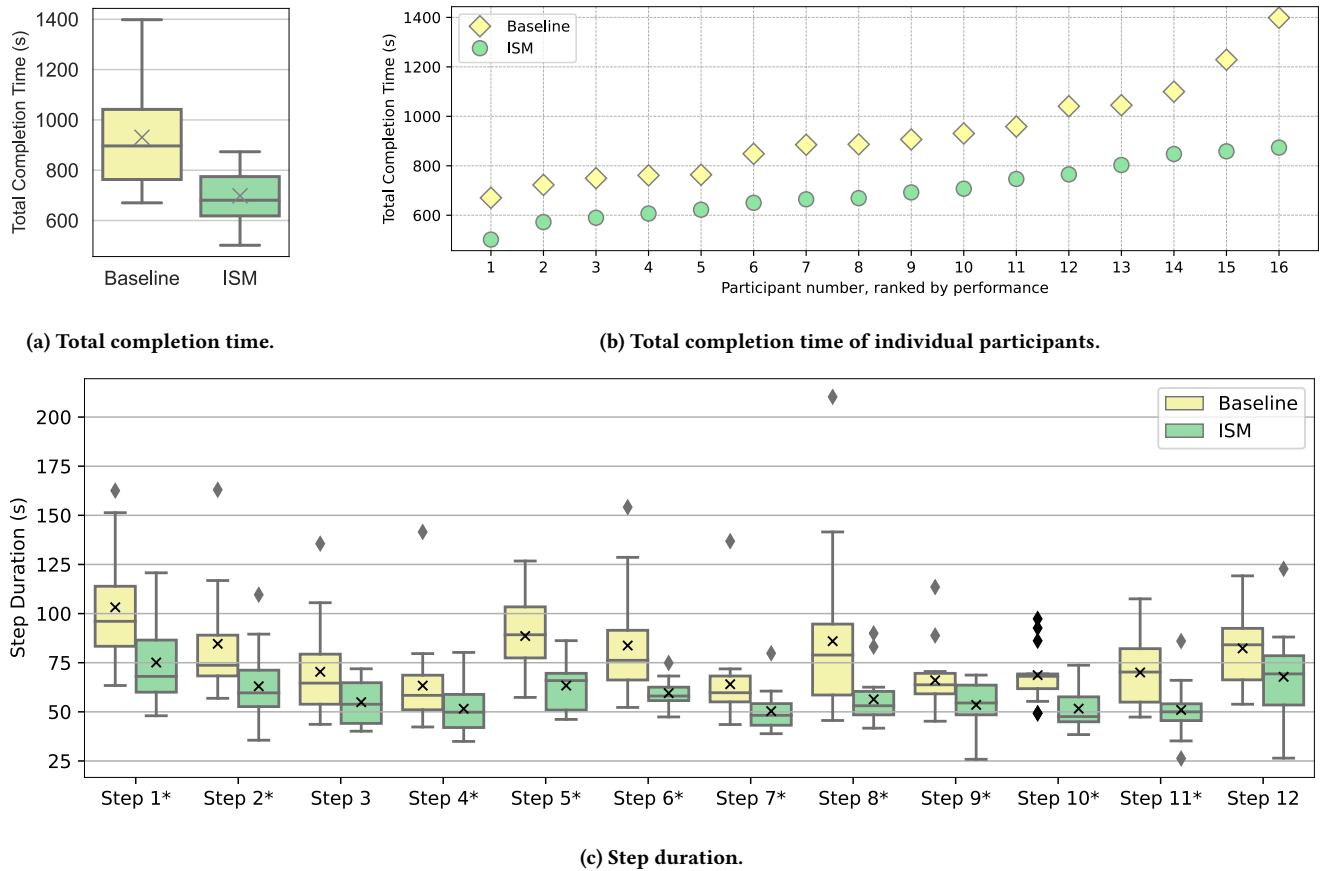
### 5.1 Completion Time

Figure 4a shows the distribution of participants’ total completion time in each condition. A Welch’s  $t$ -test at a significance level of  $\alpha = 0.05$  showed a statistically significant difference between both conditions ( $t(23.62) = 4.15, p = 0.0004, \text{Cohen’s } d = 1.47$ ). Participants using the ISM-enhanced system ( $M = 698.13, SD = 109.95$ ) completed the same task significantly faster than those using the baseline ( $M = 931.01, SD = 195.63$ ).

Figure 4b plots the total completion time for individual participants in each condition, ranked by performance (top performer using the baseline system versus top performer using the ISM-enhanced system, etc.). At every corresponding ranking position, the participant using the ISM-enhanced system consistently outperformed their counterpart using the baseline system, exhibiting shorter completion time. Additionally, the difference in completion times between the slowest performers in each condition is greater than that of the fastest performers.

Figure 4c plots participants’ completion times for each of the 12 trials under both conditions. We observed that the ISM-enhanced system enabled participants to complete the same task faster at all steps. Participants using the ISM-enhanced system completed the task significantly faster in step 2, 4, 6, 7, 8, 9, 10, 11, as demonstrated by a Mann-Whitney  $U$  test at a significance level of  $\alpha = 0.05$ ; and also in step 1 and 5, as demonstrated by a Welch’s  $t$ -test at a significance level of  $\alpha = 0.05$ . The results show that the ISM-enhanced system indeed enabled users to complete the same instruction authoring task faster, compared to a baseline system that did not conform to the ISMs implemented in the proposed system.

**5.1.1 Examining the Impact of Button Selection Methods on Completion Time.** As discussed in section 3.3.2, we incorporated 12



**Figure 4:** (a) shows the distributions of total completion time in each condition. (b) shows individual participants’ total completion times (ranked by performance), in each condition. (c) shows the distributions of step-wise completion times in each condition. The box plots show the median (the horizontal line), the mean (‘x’), the first and third quartile (the box) and the minimum and maximum (the whiskers). The diamond signs (‘◇’) indicate outliers. The asterisks (‘\*’) on the  $x$ -axis indicate that the sample groups have a statistically significant differences.

ISMs into the design of the ISM-enhanced system. One of these metaphors, “*Button* exerts *COMPULSION* to the system”, influenced our decision to implement pressable buttons in the ISM-enhanced system as opposed to dwell buttons in the baseline. This metaphor could possibly have a direct impact on the differences in completion time between the two conditions. In an ideal scenario without any failed attempts, the pressable buttons are activated immediately upon action while the dwell buttons necessitated a duration of 1.1 seconds from the initiation of the dwell action to its completion. Therefore, it was possible that the differences in completion time between the two conditions were solely attributed to the selection time difference between ISM-based buttons and non-ISM-based buttons. If this assumption holds true, it would suggest that apart from the “*Button* exerts *COMPULSION* to the system” metaphor, the integration of the remaining 11 ISMs may not have contributed to better user performance.

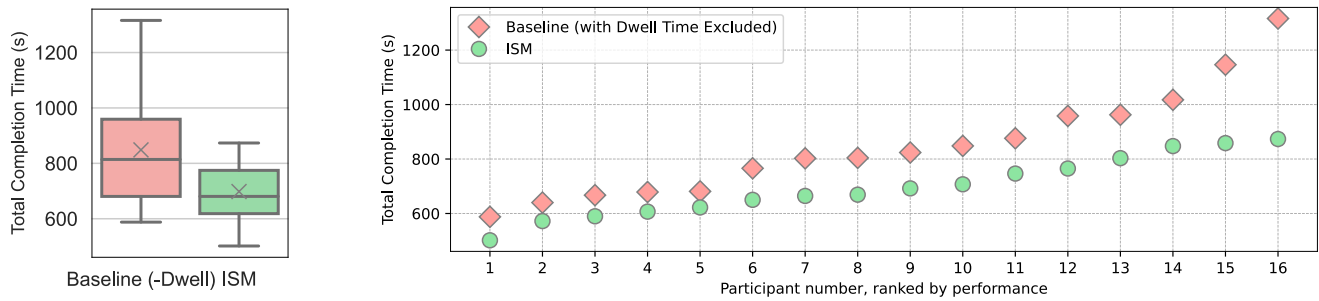
To test this assumption, we determined the number of necessary button clicks for every step and computed the button dwell time for each of these steps (see Table 4). Then we deducted the step-wise

dwell time from the step-wise completion time for the baseline condition. It was challenging to determine an accurate execution duration for a pressable button due to possible failed attempts before a successful activation. Hence, we did not make any time deductions for the ISM-enhanced condition. We then compared the baseline condition’s completion time, after deducting the dwell time, with that of the ISM-enhanced condition. It is important to note that this comparison was considerably harsh and unfavorable to the ISM-enhanced condition. Given that no button selection time was deducted from the ISM condition, it was assumed that all participants flawlessly activated all buttons immediately on their initial attempt.

Figure 5a plots the total completion time for the baseline condition (with button dwell time deducted) and the ISM-enhanced condition. A Welch’s  $t$ -test at a significance level of  $\alpha = 0.05$  showed a statistically significant difference between both conditions ( $t(23.62) = 2.68, p = 0.0132$ , Cohen’s  $d = 0.95$ ). Even with dwell time deducted from the baseline condition, participants using the ISM-enhanced systems ( $M = 698.13, SD = 109.95$ ) still completed the same task

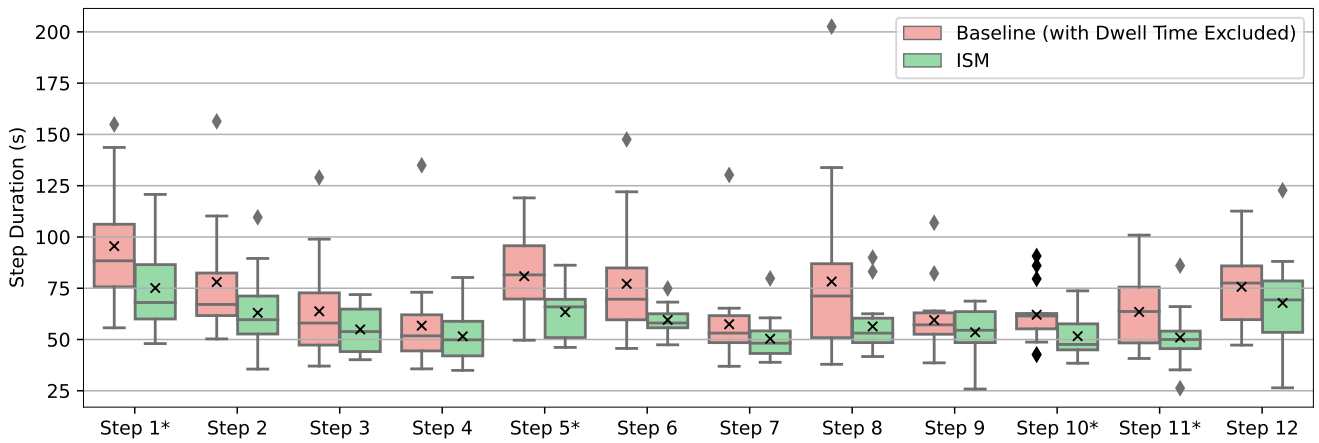
**Table 4: Necessary button clicks for each step. The total button clicks for each step were calculated by summing the counts of all necessary interactions that required a button click. A dwell button necessitated a duration of 1.1 seconds from the initiation of the dwell action to its completion**

Step	1	2	3	4	5	6	7	8	9	10	11	12
Hologram creation	2	2	2	2	2	2	2	2	2	2	2	2
Step navigation	1	1	1	1	1	1	1	1	1	1	1	1
Video control	1	0	0	0	0	0	0	0	0	0	0	0
Draw line	1	1	1	1	1	1	1	1	1	1	1	1
Menu tab switch	2	2	2	2	2	2	2	2	2	2	2	2
Menu page switch	0	0	0	0	1	0	0	1	0	0	0	0
Total button clicks	7	6	6	6	7	6	6	7	6	6	6	6
Total dwell time (s)	7.7	6.6	6.6	6.6	7.7	6.6	6.6	7.7	6.6	6.6	6.6	6.6



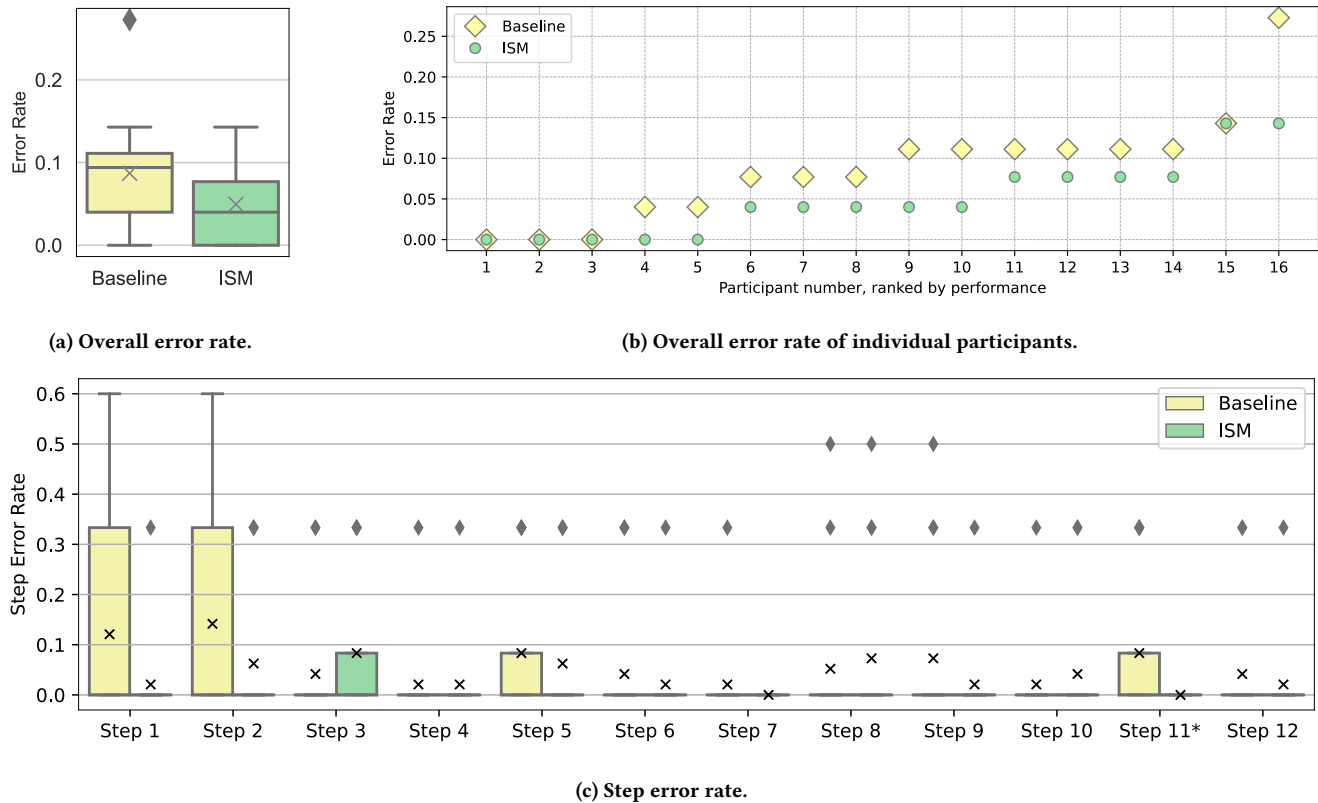
(a) Total completion time (-dwell).

(b) Total completion time of individual participants (-dwell).



(c) Step duration (-dwell).

**Figure 5: Button dwell time has been deducted from baseline’s total completion time and step-wise duration. (a) shows the distributions of total completion time in each condition. (b) shows individual participants’ total completion times (ranked by performance), in each condition. (c) shows the distributions of step-wise completion times in each condition. The box plots show the median (the horizontal line), the mean (‘x’), the first and third quartile (the box) and the minimum and maximum (the whiskers). The diamond signs (‘◊’) indicate outliers. The asterisks (‘\*’) on the x-axis indicate that the sample groups have a statistically significant differences.**



**Figure 6:** (a) shows the distributions of overall error rate in each condition. (b) shows individual participants’ overall error rates (ranked by performance), in each condition. (c) shows the distributions of step-wise error rates in each condition. The box plots show the median (the horizontal line), the mean (‘x’), the first and third quartile (the box) and the minimum and maximum (the whiskers). The diamond signs (‘◊’) indicate outliers. The asterisks (‘\*’) on x-axis indicate that the sample groups have a statistically significant difference.

significantly faster than those using the baseline ( $M = 848.51$ ,  $SD = 195.63$ ). Figure 5b plots the total completion time for individual participants for the baseline condition (with button dwell time deducted) and the ISM-enhanced condition, ranked by performance. Even with dwell time deducted from the baseline, at every corresponding ranking position, the participant using the ISM-enhanced system consistently outperformed their counterpart using the baseline. Figure 5c plots participants’ completion time for each of the 12 trials under both conditions, button dwell time was deducted from each step for the baseline condition. At all steps, the mean completion times for the ISM-enhanced system were still shorter than the baseline ones. A Welch’s  $t$ -test at a significance level of  $\alpha = 0.05$  revealed that participants using the ISM-enhanced system completed the task significantly faster in Step 1, 5, 10, 11.

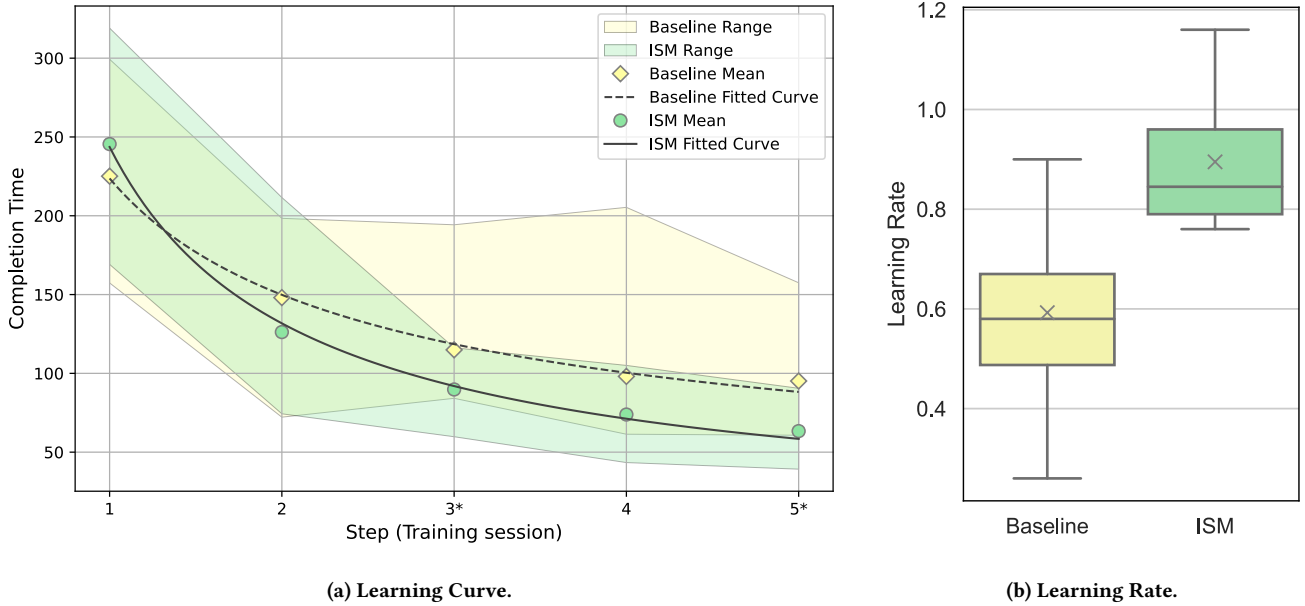
The results show that even after deducting button dwell time from the completion time of the baseline condition, participants using the ISM-enhanced system still exhibited both a significantly shorter total completion time and shorter step-wise completion times at all steps. This indicates that apart from the “Button exerts COMPULSION to the system” metaphor, the integration of the remaining 11 ISMs also contributed to the enhancement in performance.

## 5.2 Error Rate

The error rate was operationally defined as the quotient obtained by dividing the number of deleted holograms by the total number of holograms generated by the users. Figure 6a shows the distribution of participants’ overall error rate in each condition. We observed that the ISM-enhanced system ( $M = 0.05$ ,  $SD = 0.05$ ) had a lower overall error rate in comparison to the baseline ( $M = 0.09$ ,  $SD = 0.07$ ) in our sample. However, a Mann-Whitney  $U$  test with a significance level of  $\alpha = 0.05$  showed that the difference was not significant between both conditions ( $U = 173.5$ ,  $p = 0.0830$ ,  $r = 0.31$ ).

Figure 6b plots the overall error rate for individual participants in each condition, ranked by performance. At every corresponding ranking position, the participant using the ISM-enhanced system consistently had the same or a lower error rate than their counterpart using the baseline system. Additionally, the difference in error rates between the most error-prone performers in each condition is greater than that of the least error-prone performers.

Figure 6c plots participants’ error rate for each of the 12 trials under both conditions. We observed that the ISM-enhanced system had a lower error rate at eight steps (Step 1, 2, 5, 6, 7, 9, 11, 12). Among the eight steps, only at step 11 a Mann-Whitney  $U$  test



**Figure 7: (a) shows the comparison of the fitted power-law learning curves between the baseline and the ISM condition during the five-step training session. The diamond signs ( $\diamond$ ) indicate real data points for the baseline condition and the circle signs ( $\circ$ ) indicate real data points for the ISM condition. The asterisks ( $*$ ) on x-axis indicate significant differences between mean completion times of both conditions for that step. (b) shows the distributions of participants' learning rates in each condition. The box plots show the median (the horizontal line), the mean ( $\times$ ), the first and third quartile (the box) and the minimum and maximum (the whiskers).**

showed a statistically significant difference ( $U = 160.0$ ,  $p = 0.0383$ ,  $r = 0.37$ ).

The results demonstrate that, despite the ISM-enhanced system being significantly faster, it still enabled users to complete the same instruction authoring task with a comparable (or even slightly lower) error rate in comparison to the baseline.

### 5.3 Learnability

We computed the learning curves during the five-step training session for both conditions using Equation (1), which is a typical power law formula that is widely employed to examine learning effect in a variety of tasks [19, 56], for example, product assembly [22]:

$$C_x = C_1 \cdot x^{-b} \quad (1)$$

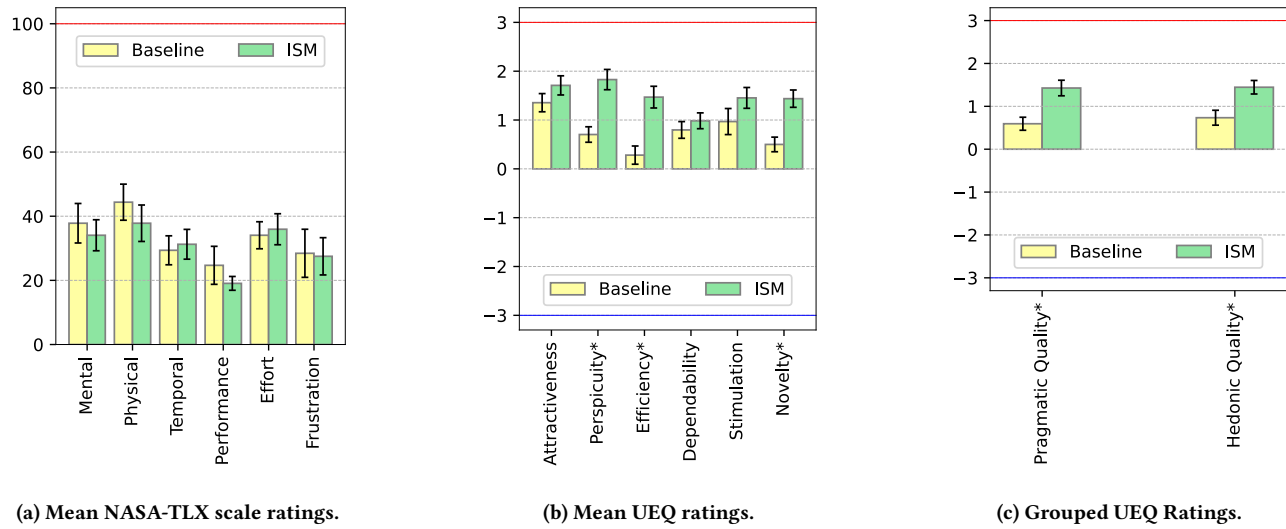
$C_x$  is the time taken to complete the  $x$ th same task,  $C_1$  is the time taken to complete the first task, and  $b$  is the learning rate. The greater the learning rate, the greater the learning effect.

**5.3.1 Learning Rate.** We performed non-linear least squares to find the best fitting curve to the completion time data during the five-step training and calculated the learning rate  $b$  based on the best fitting curve. Figure 7a shows fitted learning curves for each condition. The ISM condition had a higher learning rate compared to the baseline condition ( $b_{ISM} = 0.89$ ,  $b_{Baseline} = 0.58$ ), which was

also reflected in the visual representation of a steeper decline in completion times.

To examine the difference on learning rates between both conditions, we performed the same curve fitting method on each participant's completion time data in each condition to calculate their individual learning rates. Figure 7b plots the distribution of participants' learning rates in each condition. A Mann-Whitney  $U$  test at a significance level of  $\alpha = 0.05$  showed a statistically significant difference between both conditions ( $U = 25.0$ ,  $p = 0.0001$ ,  $r = 0.68$ ). Participants' learning rate in the ISM condition was significantly higher than that in the baseline condition.

**5.3.2 Learning Gain.** In both conditions, a Welch's  $t$ -test showed that participants spent comparable amount of time completing the first step ( $t(29.07) = 1.35$ ,  $p = 0.1880$ , Cohen's  $d = 0.48$ ). However, a Mann-Whitney  $U$  test at a significance level of  $\alpha = 0.05$  revealed that participants in the ISM-enhanced condition completed the task significantly faster in step 3 and 5 of the training session. The learning gain (measured by the difference between completion time of the first and fifth step) for participants in the ISM-enhanced system was significantly greater than for those in the baseline condition, as demonstrated by a Mann-Whitney  $U$  test ( $U = 38.0$ ,  $p = 0.0004$ ,  $r = 0.60$ ).



**Figure 8:** (a) Mean NASA-TLX scale ratings. Note that a lower ‘Performance’ rating indicates ‘better’ perceived performance. The ratings for each item could range from 0 to 100 (indicated by the red line). (b) Mean UEQ ratings at six scales. The ratings for each item could range from -3 (indicated by the blue line) to 3 (indicated by the red line). (c) Mean UEQ Rating with perspicuity, efficiency and dependability averaged into pragmatic quality; stimulation and novelty averaged into hedonic quality. The ratings for each item could range from -3 (indicated by the blue line) to 3 (indicated by the red line). Error bars show one standard error. The asterisks (“\*”) on the  $x$ -axis indicate the sample groups that have a statistically significant difference.

## 5.4 Subjective Ratings

Participants completed the User Experience Questionnaire (UEQ) and the NASA Task Load Index (NASA-TLX) at the end of the experiment. We summarized their responses in Figure 8. Figure 8a shows the mean NASA-TLX ratings across all participants for both conditions. The ISM-enhanced condition was considered by participants to be less physically and mentally demanding. Participants also perceived themselves as performing better and experiencing less frustration in the ISM-enhanced condition. The ISM-enhanced condition introduced a higher temporal load and more effort, likely due to the conditionally visible menu attached to the left palm. This design encouraged users to make hologram selections within a shorter timeframe (higher temporal load), and consistently involved the use of both hands (more effort) for menu navigation and button selection. A Mann-Whitney  $U$  Test with a significance level of  $\alpha = 0.05$  revealed that none of these measured differences were statistically significant.

Figure 8b plots the mean UEQ ratings across all participants at six user experience scales. The ISM-enhanced condition was rated higher by the participants at all six scales. A Mann-Whitney  $U$  Test with a significance level of  $\alpha = 0.05$  revealed statistically significant differences at perspicuity ( $U = 38.5$ ,  $p = 0.0007$ ,  $r = 0.60$ ), efficiency ( $U = 39.0$ ,  $p = 0.0008$ ,  $r = 0.59$ ), and novelty ( $U = 39.5$ ,  $p = 0.0008$ ,  $r = 0.59$ ). The results indicate that participants found the ISM-enhanced system significantly faster and more efficient; easier to learn and understand; more innovative, inventive and creatively designed [21].

We averaged ratings on perspicuity, efficiency and dependability for pragmatic quality, and averaged ratings on stimulation and

novelty for hedonic quality. A Mann-Whitney  $U$  Test with a significance level of  $\alpha = 0.05$  showed that the ISM-enhanced system was rated significantly higher for both pragmatic quality ( $U = 43.5$ ,  $p = 0.0015$ ,  $r = 0.56$ ) and hedonic quality ( $U = 57.0$ ,  $p = 0.0077$ ,  $r = 0.47$ ) (see Figure 8c). The results indicate that the integration of ISMs not only significantly improved pragmatic quality that includes conventional usability criteria, such as efficiency and learnability, but also markedly boosted non-goal-directed qualities, such as novelty.

## 5.5 Mental Efficiency

According to Paas and Van Merriënboer [45], higher mental efficiency is indicated when users exhibit better performance in one condition while experiencing comparable mental load in both conditions. In this study, both conditions were reported to impose comparable mental load. However, the ISM condition enabled significantly better performance, demonstrated by significantly faster speed at comparable accuracy. As a result, the ISM condition enabled higher mental efficiency compared to the baseline. We calculated the relative mental efficiency for both conditions using the method proposed by Paas and Van Merriënboer [45] and could observe that the difference in mental efficiency was statistically significant. For brevity, the calculation process and results are presented in the Appendix (see section A.1).

## 5.6 Participant Feedback

All participants were invited to complete a follow-up online questionnaire that asked them to describe aspects of the system that they liked or disliked. We collected 12 responses from participants in the ISM condition and 12 from those in the baseline condition.

**Table 5: Image-schematic metaphors found in participants' comments.**

Participant feedback	Image-schematic Metaphor(s)
"The <b>sequential breakdown of the task</b> made it very easy to keep track of the steps." (P4)	7. The form of <i>Operation</i> is PATH 9. The form of <i>Step</i> is PART
"It (the palm-menu) can be hidden when not in use so <b>it did not block my view</b> , and it brings buttons <b>closer</b> to me." (P1)	11. Extraneous <i>Information</i> is BLOCKAGE 10. The position of <i>Task</i> is NEAR
"It (the palm-menu) mirrors the experience of bringing something to the <b>center</b> of my view and <b>putting it down when it is no longer needed</b> ." (P3)	3. The position of <i>Task</i> is CENTER 12. The position of hidden <i>Information</i> is DOWN
"I like the boxes with numbers (progress indicator), it clearly shows which box <b>corresponds</b> to which step in the tutorial." (P11)	1. <i>Learning</i> is MATCHING
"The sprite sheet video slider is good, it is handy for seeing <b>the road ahead</b> ." (P8)	8. The form of <i>Video Tutorial</i> is PATH

Participants feedback was coded and categorized into four themes: *Usability*, *Novice-friendliness*, *Efficiency*, and *Accuracy*.

**5.6.1 Usability.** Both systems were perceived to have good usability. For the baseline condition, six participants commented that it was "easy to learn" and "easy to use". Five participants mentioned that it had a "clear" interface design. For the ISM condition, eight participants considered it "easy to learn" and seven mentioned "easy to use".

**5.6.2 Intuitive Use and Novice-friendliness.** The ISM system was considered intuitive to use and friendly for novice users, while these attributes were not mentioned by participants in the baseline condition. Six participants considered the ISM system "intuitive" to use. Five participant expressed that it was "novice-friendly" using phrases like "fresher friendly", "beginner friendly" and "friendly for AR novice like me". These effects could be attributed to the possibility that although these users possess very limited knowledge of MR interactions, the embedded ISMs offer a source of knowledge that could be subconsciously activated to understand unfamiliar task domains in the system.

**5.6.3 Efficiency.** The ISM system was considered "fast" or "efficient" by five participants, while the baseline was considered "slow" by five participants. Four participants expressed that it was "effortful" to complete the task with the baseline system, using the words "extra effort" or "demanding".

**5.6.4 Accuracy.** Although infrequent, two participants in the ISM condition and three in the baseline mentioned "mis-triggering", "inaccurate" or "insensitive hand tracking". These tracking issues were mostly observed to happen when the HoloLens 2 hand-tracking falsely reported the tip of the index finger as touching the thumb when there was in fact still a small gap in between, or vice versa.

Additionally, some ISMs integrated in the system were identified in the comments made by some participants when they described the aspects of the system that they liked (see Table 5). Although these ISMs were found from only a few participants' comments, it indicates that these ISMs have improved the system's information structure and presentation (P1, P3, P4) and have supported users in processing information and preparing actions (P4, P8, P11).

## 6 DISCUSSION

By analyzing the obtained data, we find that the incorporation of ISMs in the design of an MR instruction authoring system has significantly enhanced task performance, learnability, mental efficiency and subjective user experience including perspicuity, efficiency, and novelty. The ISM-enhanced condition and the baseline showed similar levels of perceived workload, perceived attractiveness, perceived dependability, and perceived stimulation. For the first time, these results empirically demonstrated the benefits of integrating ISMs into the design of an MR system in the context of instruction authoring.

### 6.1 Task Performance

Compared to the baseline, the ISM-enhanced condition enabled a substantial reduction in completion time. Generally speaking, users tend to be more error-prone when they complete a task faster. In contrast, the ISM-enhanced condition demonstrated a similar (or even lower) error rate than the baseline. The ISM-enhanced system allowed users to complete tasks significantly faster without sacrificing accuracy. This finding demonstrates the hypothesized benefit of enhancing task performance. Additionally, while all ISM-enhanced system users outperformed their counterparts with the same performance ranking in the baseline, our observations show that the difference in completion times between the slowest performers in each group is greater than that of the fastest performers. Similarly, the differences in error rate between the most error-prone performers in each group is greater than that of the least error-prone performers. This suggests that the integration of ISMs can be particularly useful in assisting MR-novice users with lower skill levels in system utilization. The "novice-friendliness" advantage was also reflected in participants' comments (see section 5.6.2). This finding is aligned with the theoretical nature of ISMs, a type of technology-independent prior knowledge that can be leveraged by everyone. Making MR systems easier to learn and use for everyday end-users with varying skill levels is a crucial prerequisite that remains unfulfilled for MR technology to achieve commercial success and greater impact [13]. Our findings suggest that the use of ISMs in MR design has therefore great potential to help address this challenge.

## 6.2 Learnability

In the five-step training session, participants in the ISM-enhanced condition learned significantly faster (higher learning rate) and better (greater learning gain) than their baseline counterparts. It is noteworthy that participants in both conditions began the training with similar performance levels, but those in the ISM-enhanced condition finished significantly faster. This indicates that the greater learning gain is not due to an unnecessarily complex design in the ISM-enhanced condition. The enhanced learnability can be attributed to the successful activation of ISMs when users encountered unfamiliar concepts during the initial steps, which facilitated the comprehension of system functions and behaviors. This finding supports Lakoff and Johnson's [37] claim that ISMs can facilitate the learning of new concepts. Prior comparative studies incorporating ISMs into screen-based GUIs did not reveal any significant enhancement in user learning. Prior empirical comparisons in embodied interaction only suggest that ISM-based design was perceived slightly (not significantly) easier to learn [2]. This is the first study that quantitatively demonstrate a significant enhancement in learning rate and learning gain enabled by the integration of ISMs in MR systems. The markedly enhancement in user learning could be attributed to the strong forms of embodiment provided by MR [31], which could potentially amplify the power of embodied metaphorical understanding, as discussed in section 2.3. The fast advancement of computing capabilities of MR systems will likely lead to an increasing number of novel concepts that novice users would have to learn. The use of ISMs shows great promise to address this challenge as it efficiently aids in the learning of unfamiliar computing concepts through embodied metaphorical understanding.

## 6.3 Mental Efficiency

Compared to the baseline, the ISM-enhanced system enabled higher mental efficiency as hypothesized. With comparable levels of mental effort invested in both conditions (indicated by the comparable ratings of "mental load" in NASA-TLX), the ISM-enhanced system enabled better task performance. This finding offers a different view from what Hurtienne [24] found, where users of ISM-congruent interfaces reported lower mental effort. This difference might stem from the different intrinsic complexities of the task (e.g., setting time/temperature on a heating control system [24] versus authoring assembly instructions for a model), which impose different levels of intrinsic load. For systems dedicated for complex tasks like instruction authoring, we argue that the integration of ISMs might not necessitate lower mental effort, which consists of intrinsic and extraneous load. Rather, its true benefit may lie in optimizing the distribution of invested mental effort between these two types of loads. By minimizing the extraneous load associated with learning to navigate the system, more cognitive resources become available for handling the task at hand.

## 6.4 User Experience

The ISM-enhanced system was rated significantly more efficient in UEQ. The enhanced perceived efficiency was also reflected in participants' feedback, where the ISM-enhanced system was described as "fast" or "efficient", and the baseline was described as

"slow". This finding is aligned with what Hurtienne [24] suggests: users of ISM-based interactions would not have to stop due to mismatches between the interaction flow with the system and the internal simulation. This finding highlights the potential benefits of integrating ISMs into systems where a perceptually fast and smooth user experience would be desirable (e.g., MR games).

The ISM-enhanced system was rated significantly higher in perceived perspicuity in UEQ. Moreover, participants described the ISM-enhanced system as "easy to learn", "novice-friendly" and "intuitive". One participant (P3) mentioned that the hand gesture that activates and hides the toolkit menu was "intuitive" as "*it mirrors the experience of bringing something to the center of my view and putting it down when it is no longer needed.*" This indicates that the integrated ISMs (metaphor 3 and 12 in Table 1) enabled users to make effective use of their everyday non-technological experience to understand system interactions and behaviors, which is an important prerequisite for making a system easier to learn for a novice.

The ISM-enhanced system was rated significantly higher in perceived novelty in UEQ. This is in alignment with Winkler et al.'s findings [55], where ISM-based GUIs were perceived more innovative. Intriguingly, participants found the ISM-enhanced system to be highly innovative while being easy to learn and use. There was a presumption that systems that are easy to learn and use could not meanwhile be innovative, as they typically adhere to the conventions of existing technologies, allowing users to leverage their prior technological experiences [8, 27, 44]. However, the integrated ISMs provided an alternative source of non-technological prior knowledge for users to leverage, this resolved the conflict between novelty and usability, as what Hurtienne et al. [27] predicted.

## 6.5 ISM-Inspired Design Pathway

This study has empirically demonstrated several benefits of incorporating ISMs into user interface elements and user interactions of one MR system. However, it is important to recognize that these benefits were observed in just one possible implementation of ISMs, and they are not generalizable without further comparative studies across different tasks and designers, while also considering different forms of user interfaces.

To facilitate such explorations in the future, this study uniquely demonstrated an ISM-inspired design pathway (see Figure 9), which can assist designers to systematically integrate a set of ISMs into various aspects of an MR system. Using a list of ISMs, we recommend that designers (1) first link different system components to appropriate ISMs; and (2) link these metaphors to possible output (e.g., visual (appearance and spatial relations), auditory, tactile) and input (e.g., gesture-based, speech-based, gaze-based interactions) aspects where they can be instantiated. This "System component—ISMs—Input/Output Instantiation" pathway is the initial attempt to operationalize ISMs as a potential design approach in MR system design, leveraging the compatibility between the possible instantiation modalities of ISMs and the input/output modalities of MR systems. Future studies can easily customize or augment this basic pathway based on specific requirements of various design tasks and different forms of user interfaces, exploring ISMs' wider applicability in the design of MR systems.

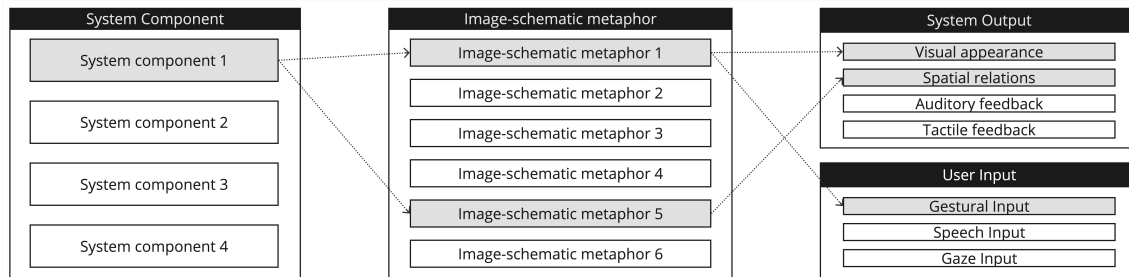


Figure 9: ISM-inspired design pathway.

## 7 CONCLUSION

This work empirically demonstrated the benefits of incorporating ISMs in user interface elements and user interactions of an MR system. This is supported by the evaluation between an established MR instruction authoring system and its redesigned version based on 12 ISMs that the baseline did not conform to. In a between-subject experiment with 32 participants, the ISM-enhanced system allows users to complete the same task significantly faster with the same level of accuracy. The ISM-enhanced system also demonstrated significantly higher mental efficiency: with comparable levels of mental effort invested, the ISM-enhanced system enabled significantly better task performance. During the five-step training sessions, users of the ISM-enhanced system learned significantly faster, as indicated by a higher learning rate, and better, as indicated by a greater learning gain. Additionally, the ISM-enhanced system was rated significantly higher in perceived efficiency, perspicuity and novelty. We observed usability and learnability benefits across participants of varying performance levels in the ISM-enhanced condition. This supports the general premise of this work in that the integration of ISMs can serve as an effective middle ground for designing MR systems that are usable to a wide range of users with varying skill levels, without requiring costly adaptations. However, it is important to note that achieving widespread usability with limited adaptations represents only one objective of MR system design. When designed appropriately, MR can deliver various additional benefits that were not explored in this work.

Further, the observed benefits of incorporating ISMs are not generalizable without further validations across various tasks and designers. To support such validations, the presented work demonstrated a “System Component—ISMs—Input/Output Instantiations” design pathway, which marks the first attempt to operationalize ISMs as a practical design method in MR.

Additionally, the benefits of incorporating ISMs in an MR system design is clearly demonstrated by the results obtained from a controlled user study. However, it is important to acknowledge that our experiment is inherently dependent on a specific task formulation. Further studies are desirable to evaluate both conditions in in-the-wild settings and examine the effectiveness of this design approach.

Finally, according to the findings of our study, implementing all twelve ISMs effectively improved performance, learnability,

and user experience ratings. Given that multiple ISMs were implemented in the design of some system components, their effectiveness was interdependent. The results of this study were unable to tell us how each of these 12 metaphors contributed to the improvement or which metaphors were more effective than others. Future research could investigate the efficacy of a set of ISMs as well as the effectiveness of individual metaphors within this set, as this could provide important information about what types of ISMs designers should prioritise as design guidelines.

We conclude that ISMs offer a promising method for designers to create widely usable user interfaces and interactions in MR systems that resonate with the mental models shared among a wide range of users. This could facilitate the widespread utilization of an MR system in a cost-effective manner, in comparison to adaptations and customization. Future studies should empirically validate if the integration of ISMs yields usability benefits consistently across different designers, tasks, and types of user interfaces.

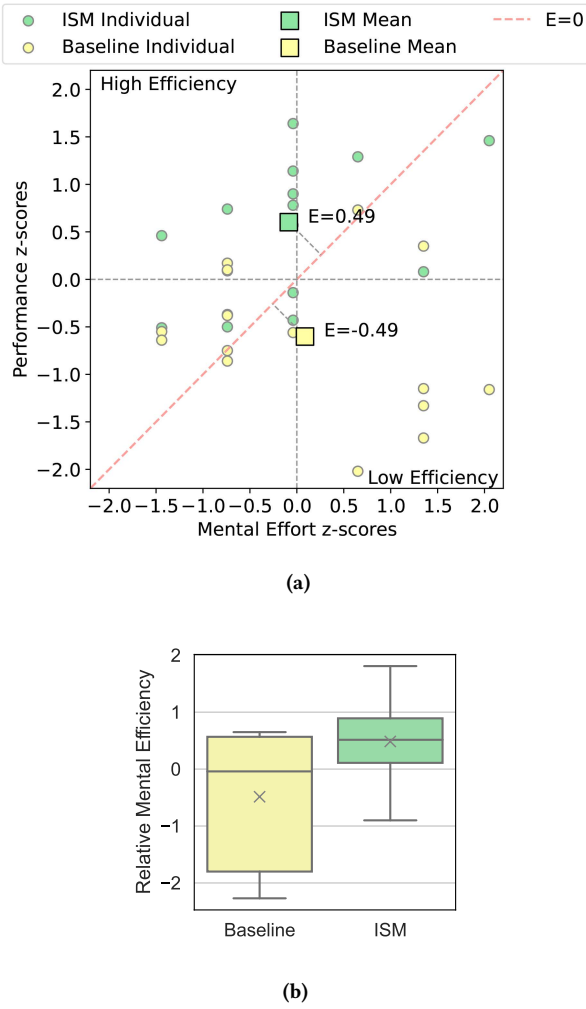
## 8 CODE AVAILABILITY STATEMENT

Source code for the two MR instruction authoring systems (baseline and ISM-enhanced) presented in this paper is available at <https://doi.org/10.17863/CAM.106314> and [https://github.com/Jingyi997/AR\\_Instruction\\_Authoring.git](https://github.com/Jingyi997/AR_Instruction_Authoring.git).

## REFERENCES

- [1] Sylvia Ahern and Jackson Beatty. 1979. Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science* 205, 4412 (1979), 1289–1292.
- [2] Alissa N. Antle, Milena Droumeva, and Greg Corness. 2008. Playing with the Sound Maker: Do Embodied Metaphors Help Children Learn?. In *Proceedings of the 7th International Conference on Interaction Design and Children*. Association for Computing Machinery, New York, NY, USA, 178–185.
- [3] Apple Inc. 1992. *Macintosh Human Interface Guidelines*. Addison-Wesley Publishing Company, USA.
- [4] Apple Inc. 2024. *Human Interface Guidelines*. Retrieved February 5, 2024 from <https://developer.apple.com/design/human-interface-guidelines>
- [5] Saskia Bakker, Alissa N. Antle, and Elise Van Den Hoven. 2009. Identifying embodied metaphors in children’s sound-action mappings. In *Proceedings of the 8th International Conference on Interaction Design and Children*. Association for Computing Machinery, Como, Italy, 140–149.
- [6] Saskia Bakker, Alissa N. Antle, and Elise Van Den Hoven. 2012. Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing* 16 (2012), 433–449.
- [7] P. Bégout, T. Duval, S. Kubicki, B. Charbonnier, and E. Bricard. 2020. WAAT: A Workstation AR Authoring Tool for Industry 4.0. In *Augmented Reality, Virtual Reality, and Computer Graphics*, Lucio De Paolis and Patrick Bourdot (Eds.). Lecture Notes in Computer Science, Vol. 12243. Springer, Cham. [https://doi.org/10.1007/978-3-030-58468-9\\_22](https://doi.org/10.1007/978-3-030-58468-9_22)
- [8] Alethea Blackler, Vesna Popovic, Doug Mahar, Gaurav Reddy, and Simon Lawry. 2012. Intuitive Interaction and Older People. In *Proceedings of the DRS 2012*

- International Conference: Research: Uncertainty Contradiction Value*, P. Prasena, J. Tangsantikul, and D. Durling (Eds.). Design Research Society, Bangkok, Thailand, 560–578.
- [9] Walter R Boot, Neil Charness, Sara J Czaja, Joseph Sharit, Wendy A Rogers, Arthur D Fisk, Tracy Mitzner, Chin Chin Lee, and Sankaran Nair. 2015. Computer proficiency questionnaire: assessing low and high computer proficient seniors. *The Gerontologist* 55, 3 (2015), 404–411.
  - [10] Marianna Ciccarelli, Agnese Brunzini, Alessandra Papetti, and Michele Germani. 2022. Interface and interaction design principles for Mixed Reality applications: The case of operator training in wire harness activities. *Procedia Computer Science* 204 (2022), 540–547.
  - [11] Maarten Coëgnarts. 2019. *Chapter 2. Embodying the Meaning: The Role of Image Schemas, Metaphors, and Metonymies*. Academic Studies Press, Boston, USA, 39–68.
  - [12] Vesna G. Djokic, Ekaterina Shutova, and Rebecca Fiebrink. 2021. MetaVR: Understanding metaphors in the mind and relation to emotion through immersive, spatial interaction. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 185, 4 pages. <https://doi.org/10.1145/3411763.3451565>
  - [13] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI principles to AR systems design. In *Proceedings of the 2nd International Workshop at the IEEE Virtual Reality 2007 Conference (MRUI'07)*. IEEE, Charlotte, NC, USA.
  - [14] Afshan Ejaz, Syed Asim Ali, Muhammad Yasir Ejaz, and Farhan Ahmed Siddiqui. 2019. Graphic user interface design principles for designing augmented reality applications. *International Journal of Advanced Computer Science and Applications (IJACSA)* 10, 2 (2019), 209–216.
  - [15] Wilbert O Galitz. 2007. *The essential guide to user interface design: an introduction to GUI design principles and techniques*. John Wiley & Sons, Hoboken, NJ, USA.
  - [16] Xuewang Geng and Masanori Yamada. 2020. An augmented reality learning system for Japanese compound verbs: study of learning performance and cognitive load. *Smart Learning Environments* 7 (2020), 1–19.
  - [17] M. González-Franco, J. Cermeron, K. Li, R. Pizarro, J. Thorn, P. Hannah, W. Hutabarat, A. Tiwari, and P. Bermell-Garcia. 2016. Immersive augmented reality training for complex manufacturing scenarios. *arXiv abs/1602.01944* (2016).
  - [18] Joseph E. Grady. 1997. *Foundations of meaning: Primary metaphors and primary scenes*. Ph. D. Dissertation. University of California, Berkeley.
  - [19] Andrew Heathcote, Scott Brown, and Douglas JK Mewhort. 2000. The power law revealed: The case for an exponential law of practice. *Psychonomic bulletin & review* 7, 2 (2000), 185–207.
  - [20] Steven J Henderson and Steven K Feiner. 2011. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE international symposium on mixed and augmented reality*. IEEE, Basel, Switzerland, 191–200.
  - [21] Andreas Hinderks, Martin Schrepp, Francisco José Domínguez Mayo, María José Escalona, and Jörg Thomaschewski. 2019. Developing a UX KPI based on the user experience questionnaire. *Computer Standards & Interfaces* 65 (2019), 38–44.
  - [22] Steven Hoedt, Arno Claeys, Hendrik Van Landeghem, and Johannes Cottyn. 2017. The evaluation of an elementary virtual training system for manual assembly. *International Journal of Production Research* 55, 24 (2017), 7496–7508.
  - [23] Karen Holtzblatt and Hugh Beyer. 2016. *Contextual Design, Second Edition: Design for Life* (2nd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
  - [24] Jörn Hurtienne. 2011. *Image schemas and design for intuitive use: Exploring new guidance for user interface design*. Ph. D. Dissertation. Technische Universität Berlin.
  - [25] Jörn Hurtienne. 2023. ISCAT - CENTER-PERIPHERY. <https://iscat.psyergo.uni-wuerzburg.de/image-schemas/space/center-periphery> Accessed: 2023-11-28.
  - [26] Jörn Hurtienne, Stephan Huber, and Cordula Baur. 2022. Supporting User Interface Design with Image Schemas: The ISCAT Database as a Research Tool. In *CEUR Workshop Proceedings (EUR-WS. org)*. Jönköping Sweden.
  - [27] Jörn Hurtienne, Kerstin Klöckner, Sarah Diefenbach, Claudia Nass, and Andreas Maier. 2015. Designing with image schemas: resolving the tension between innovation, inclusion and intuitive use. *Interacting with Computers* 27, 3 (2015), 235–255.
  - [28] Jörn Hurtienne and Patrick M. Langdon. 2010. Keeping warm in winter: Image-schematic metaphors and their role in design of central heating controls. In *Fourth International Conference of the German Cognitive Linguistics Association*. German Cognitive Linguistics Association, Bremen, 53–54.
  - [29] Jörn Hurtienne, Christian Stöfel, Christine Sturm, Alexander Maus, Matthias Rötting, Patrick Langdon, and John Clarkson. 2010. Physical gestures for abstract concepts: Inclusive design with primary metaphors. *Interacting with Computers* 22, 6 (2010), 475–484.
  - [30] Mark Johnson. 1987. *The Body in the Mind. The Bodily Basis of Meaning, Reason and Imagination*. Chicago University Press, Chicago.
  - [31] Mina C Johnson-Glenberg, David A Birchfield, Lisa Tolentino, and Tatyana Kozziupa. 2014. Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of educational psychology* 106, 1 (2014), 86.
  - [32] Junhan Kong, Dena Sabha, Jeffrey P Bigam, Amy Pavel, and Anhong Guo. 2021. TutorialLens: authoring Interactive augmented reality tutorials through narration and demonstration. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*. Association for Computing Machinery, New York, NY, USA, 1–11.
  - [33] Konstantinos Koumaditis, Sarune Venckute, Frederik S Jensen, and Francesco Chinello. 2019. Immersive training: outcomes from small scale AR/VR pilot-studies. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, Osaka, Japan, 1–5.
  - [34] George Lakoff. 1987. *Women, fire, and dangerous things*. Vol. 10. University of Chicago Press, Chicago.
  - [35] G. Lakoff. 1991. *Master Metaphor List*. University of California, Berkeley, CA.
  - [36] George Lakoff. 1993. The Contemporary Theory of Metaphor. In *Metaphor and Thought*, Andrew Ortony (Ed.). Cambridge University Press, Cambridge, UK, 202–251.
  - [37] George Lakoff and Mark Johnson. 2004. *Metaphors we Live by*. University of Chicago Press, Chicago.
  - [38] Li Lan, Pawel Wargocki, and Zhiwei Lian. 2014. Thermal effects on human performance in office environment measured by integrating task speed and accuracy. *Applied ergonomics* 45, 3 (2014), 490–495.
  - [39] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Vol. 5298. Springer, Berlin, Heidelberg, 63–76. [https://doi.org/10.1007/978-3-540-89350-9\\_6](https://doi.org/10.1007/978-3-540-89350-9_6)
  - [40] Jingyi Li, Nathan Crilly, and Per Ola Kristensson. 2024. Guiding the Design of Inclusive Interactive Systems: Do Younger and Older Adults Use the Same Image-Schematic Metaphors? *ACM Trans. Comput.-Hum. Interact.* (feb 2024). <https://doi.org/10.1145/3648618> Just Accepted.
  - [41] Anna Macaranas, Alissa N Antle, and Bernhard E Riecke. 2015. What is intuitive interaction? Balancing users' performance and satisfaction with natural user interfaces. *Interacting with Computers* 27, 3 (2015), 357–370.
  - [42] Jean M. Mandler. 2005. How to build a baby: iii. Image schemas and the transition to verbal thought. In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. De Gruyter Mouton, Berlin, 137–163.
  - [43] Meta. 2024. Informative Guides to Help You Design, Develop, and Distribute Your VR App. Retrieved February 5, 2024 from <https://developer.oculus.com/resources/learn/>
  - [44] Marita A O'Brien. 2010. *Understanding human-technology interactions: The role of prior experience and age*. Ph. D. Dissertation. Georgia Institute of Technology.
  - [45] Fred GWC Paas and Jeroen JG Van Merriënboer. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors* 35, 4 (1993), 737–743.
  - [46] Riccardo Palmari, Iñigo Fernández Del Amo, Dedy Ariansyah, Samir Khan, John Ahmet Erkoyuncu, and Rajkumar Roy. 2023. Fast Augmented Reality Authoring: Fast Creation of AR Step-by-Step Procedures for Maintenance Operations. *IEEE Access* 11 (2023), 8407–8421.
  - [47] EC Poulton and PR Freeman. 1966. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin* 66, 1 (1966), 1.
  - [48] Microsoft Press. 1995. *The Windows Interface Guidelines for Software Design*. Microsoft Press, USA. <https://books.google.co.uk/books?id=G8peHj787EUC>
  - [49] Iulian Radu. 2017. *Exploring the Usability of Augmented Reality Interaction Techniques During Children's Early Elementary-School Years*. Ph. D. Dissertation. Georgia Institute of Technology, Atlanta, GA, USA.
  - [50] John Sweller, Jeroen JG van Merriënboer, and Fred Paas. 2019. Cognitive architecture and instructional design: 20 years later. *Educational psychology review* 31 (2019), 261–292.
  - [51] Robert Tscharn. 2018. *Innovative and age-inclusive interaction design with image-schematic metaphors*. Ph. D. Dissertation. Bayerische Julius-Maximilians-Universität Würzburg (Germany).
  - [52] Neha Tuli and Archana Mantri. 2020. Usability principles for augmented reality based kindergarten applications. *Procedia Computer Science* 172 (2020), 679–687.
  - [53] Stefan Werrlich, Carolin Lorber, Phuc-Anh Nguyen, Carlos Emilio Franco Yanez, and Gunther Notni. 2018. Assembly training: comparing the effects of head-mounted displays and face-to-face training. In *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I 10*. Springer, Berlin, Heidelberg, 462–476.
  - [54] Stefan Werrlich, Phuc-Anh Nguyen, and Gunther Notni. 2018. Evaluating the training transfer of Head-Mounted Display based training for assembly tasks. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. Association for Computing Machinery, New York, NY, USA, 297–302.
  - [55] Armin Winkler, Kristian Baumann, Stephan Huber, Robert Tscharn, and Jörn Hurtienne. 2016. Evaluation of an application based on conceptual metaphors for social interaction between vehicles. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. Association for Computing Machinery, Brisbane, Australia, 1148–1159.
  - [56] Willard I Zangwill and Paul B Kantor. 1998. Toward a theory of continuous improvement and the learning curve. *Management Science* 44, 7 (1998), 910–920.



**Figure 10:** (a) is the visual representation of relative mental efficiency for the baseline and ISM conditions. (b) shows the distributions of participants’ relative mental efficiency in each condition. The box plots show the median (the horizontal line), the mean (‘x’), the first and third quartile (the box) and the minimum and maximum (the whiskers).

## A APPENDIX

### A.1 Relative Mental Efficiency Calculation

Paas and Van Merriënboer [45] propose a calculational method to estimate the relative mental efficiency of each condition by combining mental load and performance. This method computes the relative mental efficiency score ( $E$ ) for each participant in each condition as the perpendicular distance between a point (determined by the  $z$  score for mental effort and the  $z$  score for performance) and the diagonal where  $E = 0$  (see Figure 10a for a visual presentation), using the following equation.

$$E = \frac{z_{Performance} - z_{Mental\ Effort}}{\sqrt{2}} \quad (2)$$

The  $\sqrt{2}$  in this equation comes from the general formula that calculates the distance from a point ( $p(x, y)$ ) to a line ( $ax + by + c = 0$ ). However, it is important to note that the measures of relative mental efficiency should be considered as estimates [45].

In this study, we measured performance ( $P$ ) using the following equation proposed by Lan et al. [38], which considered both speed and accuracy when participants’ completion time included the time spent for the correction of errors.

$$P = \frac{1}{\text{Mean Step Completion Time (seconds)}} \quad (3)$$

After obtaining the Performance ( $P$ ) value for each participant, we transformed the  $P$  values to  $z$  scores to obtain  $z_{Performance}$ . We used participants’ self-reported mental load score in the NASA-TLX as a proxy for the mental effort value. Then we transformed the mental load data into  $z$  scores to obtain  $z_{Mental\ Effort}$  for each participant.

Figure 10b plots the distribution of participants’ relative mental efficiency in each condition. A Mann-Whitney  $U$  test with a significance level of  $\alpha = 0.05$  revealed a statistically significant difference between both conditions ( $U = 64.0, p = 0.0167, r = 0.42$ ). Figure 10a shows that the relative mental efficiency of the ISM condition ( $E_{ISM} = 0.4862$ ) is higher than the baseline condition ( $E_{Baseline} = -0.4856$ ). Additionally, the majority of individual data points (12 out of 16) in the ISM condition fell into the high efficiency zone, while the majority of individual data points (10 out of 16) in the baseline condition fell into the low efficiency zone. The calculation result provides an estimation of each condition’s relative mental efficiency, indicating that the ISM condition enabled higher mental efficiency compared to the baseline.

### A.2 Image-Schematic Metaphors Checklist

See Table 6.

**Table 6: A set of 37 “universal” image-schematic metaphors used by both younger and older adults in the context of technology learning, identified in [40]. The interpretation of each ISMs is elaborated based on the explanation of the image schema in this ISM, and the meaning of each image schema is defined based on the ISCAT database [26].**

Image-schematic metaphors	Interpretations
1. <i>Learning</i> is MATCHING	Learning is based on correspondence in pattern, appearance or design between learning objectives and examples.
2. Important <i>Information</i> is CENTER	Important information is located at the center of FOV (Field of view).
3. The position of <i>Video Tutorial</i> is PERIPHERY - Task is CENTER	The ongoing task centers the FOV, while video tutorial locates at the periphery.
4. <i>Button</i> exerts COMPULSION to the system	Users compel a system to execute certain functions by pressing a button.
5. The form of <i>Step</i> is PART	The form of step is one individual segment within a larger system or structure.
6. The form of <i>Logic</i> is PATH 7. The form of <i>Operation</i> is PATH 8. The form of <i>Video Tutorial</i> is PATH 9. The form of <i>Step</i> is PATH	The form of logic/operation/video tutorial/step is a path that consists of a starting point, an end-point, and a sequence of contiguous locations connecting the starting point with the end-point.
10. The position of <i>Video Tutorial</i> is FAR - Task is NEAR	The ongoing task locates closer to the user in their FOV, while video tutorial is further.
11. Extraneous <i>Information</i> is BLOCKAGE	Extraneous elements is a blocking force that stops or redirects an action attempted by a user.
12. The position of Hidden <i>Information</i> is DOWN	Information is hidden at (and can be retrieved from) the lower part of interaction space.
13. Upward movement on <i>Button</i> is BIGGER	An upward movement within the responsive area of a button results in an increase in the size of specific system elements.
14. The structure of <i>System</i> is COLLECTION	A system consists of several objects that are similar (e.g. in form, function, color), autonomous and mostly near each other.
15. The position of <i>Video Tutorial</i> is LEFT - Task is RIGHT	The ongoing task takes up the right side of users' FOV, while video tutorial takes up the left.
16. The function of <i>Interface</i> is enabling CONTACT	An interface can enable the physical coming together of user input and the system.
17. The structure of <i>System</i> is MERGING (different parts)	The fitting together of multiple objects forms a system.
18. General <i>Information</i> is UP - Detailed <i>Information</i> is DOWN	General information locates at the upper part of user's FOV, while detailed information locates the lower part.
19. The form of <i>Interface</i> is LINKAGE	An interface is a linking device between the user and the system.
20. The form of <i>Barrier</i> is BLOCKAGE 21. The form of <i>Problem</i> is BLOCKAGE 22. Missing <i>Step</i> is BLOCKAGE	The form of a barrier/problem/missing step is a blocking force that stops an action attempted by a user.
23. The position of <i>Manual</i> is CENTER - Task is PERIPHERY	Manual instructions locate at the center of users' FOV, while the ongoing task locates at the periphery.
24. <i>Assistance</i> is RESTRAINT REMOVAL for users	Assistance is the removal of a blocking force that stops an action attempted by the user.
25. <i>Learning</i> is taking IN (new things)	Learning is to take new knowledge into the container of memory.
26. <i>Assistance</i> is ENABLEMENT to users 27. <i>Software</i> is ENABLEMENT to its users 28. Providing <i>Option</i> is ENABLEMENT to users	Assistance/Software/Option provides users a felt sense of power to perform some action.
29. The position of <i>Video Tutorial</i> is RIGHT - Task is LEFT	The ongoing task takes up the left side of users' FOV, while video tutorial takes up the right.
30. The position of <i>Manual</i> is LEFT - Task is RIGHT	The ongoing task takes up the right side of users' FOV, while manual instructions take up the left.
31. The position of <i>Manual</i> is PERIPHERY - Task is CENTER	Manual instructions locate at the periphery of users' FOV, while the ongoing task locates at the center.
32. The form of <i>Search</i> is PATH	The form of search is a path that consists of a starting point, an end-point, and a sequence of contiguous locations connecting the starting point with the end-point.
33. The structure of <i>Manual</i> is COLLECTION 34. The structure of <i>Control Panel</i> is COLLECTION	An(A) instruction manual/control panel consists of several objects that are similar (e.g. in form, function, color), autonomous and mostly near each other.
35. Different <i>Interfaces</i> is BLOCKAGE to task completion	Using different interfaces to present the same functions is a blocking force that stops users from completing tasks.
36. Mismatch between <i>Button</i> and its assumed <i>Functionality</i> is BLOCKAGE	The mismatch between a button and its assumed function is a blocking force that stops an action attempted by a user.
37. The function of <i>Screen</i> is enabling CONTACT	An screen can enable the physical coming together of user input and the system.