Extended Multi-Phase Study

Human Factors
2026, Vol. 0(0) 1–34

DOI: 10.1177/00187208261480471
journals.sagepub.com/home/hfs
Sage

Elicitation and Assessment of Hand-Based Gestures for Information Processing in Immersive Environments

Fangyuan Cheng[1], Emily Fang[1], and Karen B. Chen[1]

Abstract

Objective: This study aims to investigate how information processing tasks can be performed by hand-based gestures in immersive environments and to identify usable gesture designs.
Background: With advances in display and interaction technologies, immersive environments have evolved to support productivity-oriented activities, in which information processing plays a critical role. However, performing such tasks in immersive environments remains challenging and calls for intuitive interaction techniques aligned with users' cognitive processes. Gestures offer a promising approach by allowing users to externalize and manipulate information through bodily actions.
Methods: Using 19 cognitive processes outlined in the revised Bloom's taxonomy as representative information processing tasks, a gesture elicitation study with 15 participants explored how these tasks could be mapped to hand-based gestures (Study I). The resulting candidate gestures were subsequently validated and evaluated with 20 participants using a combination of objective and subjective measures (Study II).
Results: Study I produced an initial set of 32 candidate gestures, which were validated and evaluated in Study II. Four recurring mapping strategies were identified: linguistic-symbolic, spatial-manipulative, metaphoric, and social-conventional mappings. A consolidation model was then introduced to identify a recommended set of 20 gestures based on their performance across multi-dimensional measures.
Conclusions: Information processing tasks can be translated into usable hand-based gestures. Future gesture design should treat information as an interactive entity while carefully considering ergonomics, semantic expressiveness, and gesture similarity to improve usability.
Applications: The recommended gestures, mapping strategies, and design considerations can inform the gesture design for immersive systems that involve information processing tasks.

Keywords

gestures, information processing, immersive environments, elicitation study, usability

[1]Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, USA

Corresponding Author:
Karen B. Chen, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, 915 Partners Way, Raleigh, NC 27606, USA.
Email: kbchen2@ncsu.edu

Introduction

With advances in display and interaction technologies, immersive environments (i.e., virtual reality and augmented reality [VR, AR]) are increasingly used to support productivity-oriented activities, such as work and learning (Luo et al., 2025; Yang et al., 2021, 2022), where information processing is a critical activity. Information processing underpins important human cognitive abilities such as knowledge construction, problem solving, and decision making (Eggen, 2020). Immersive environments have shown promise in supporting information processing tasks such as web browsing, sensemaking, and immersive analytics (Ichsan et al., 2025; Lisle et al., 2021; Skarbez et al., 2019; Toyama et al., 2018). More fundamentally, research has shown that immersive environments can facilitate the underlying cognitive processes of intellectual activities, including information structuring, synthesis, and memorization (Yang et al., 2021, 2022). Such benefits have been attributed to the spatial and embodied affordances of immersive environments, which facilitate a more engaging experience and enhance the ways information is perceived, organized, and acted upon (Arms et al., 1999; Krokos et al., 2019; Ragan et al., 2012; Yang et al., 2021). Overall, immersive environments show great potential for supporting information processing and represent a growing area of research.

However, immersive environments also introduce challenges for information processing. First, information can be more difficult to read in immersive environments due to a constrained field of view and limited resolution. Improved readability through spatially partitioned information layouts requires correspondingly evolved interaction techniques to match this segmented approach to presentation. Second, cognitive processing of information, such as comprehension and reasoning, can be more difficult in immersive environments. This is attributed to difficulty in text entry, selection, and editing. Many prior studies have sought to address this challenge by exploring interaction techniques that are more natural than predefined hand- or controller-based input (Dudley et al., 2024), leveraging modalities such as gaze, voice, and head motion (Lee et al., 2022; Monteiro et al., 2021). Yet, these solutions do not directly support the diverse cognitive tasks that users seek to perform on information. Instead, the present study approaches this problem from the perspective of users' cognitive tasks by treating cognitive processes as predefined functions and examining how they can be supported through natural interaction methods.

Information processing refers to the coordination of cognitive operations that individuals use to perceive, transform, encode, store, and retrieve information (Anderson, 1990). Components of the human information processing system (e.g., perception, working memory, and long-term memory) operate in a coordinated manner to support higher-level information processing tasks (Cowan, 2014; Ormrod, 2008; Çeliköz et al., 2019). These tasks can be classified according to Bloom's taxonomy (Bloom et al., 1956) and its revised version (Anderson & Krathwohl, 2001), which outlines 19 cognitive processes organized into six major levels (Figure 1). For example, tasks targeting the "Understand" level include finding examples (exemplifying), identifying patterns (comparing), and drawing inferences (inferring). Although the taxonomy originated in the educational domain, it has been widely applied across diverse contexts to structure cognitive tasks. For instance, it has been adopted to analyze human problem-solving strategies and to design task hierarchies that align with cognitive demands (Castillo-Barrera et al., 2017; Suffian et al., 2022; Yaacoub et al., 2025). It has also been used to examine cognitive engagement in data visualization and to structure evaluation tasks in user experience design (Burns et al., 2020; Rante et al., 2024). Together, Bloom's taxonomy can be considered a practical framework for indexing information processing tasks that may arise in immersive environments.

Levels	Index	Cognitive processes	Working definitions	Index	Cognitive processes	Working definitions
Remember	1	Recognizing	Identification of information in long-term memory that is consistent with presented material.	2	Recalling	Retrieving relevant information from long-term memory.
Understand	3	Interpreting	Making sense of of information by translating it from one representational form (e.g., visual) into another (e.g., conceptual).	4	Inferring	Drawing logical conclusions from available information, or predicting possible outcomes based on given facts or evidence.
	5	Exemplifying	Finding a specific example or illustration of a concept or principle.	6	Comparing	Identifying similarities and differences between two or more ideas, objects, or pieces of information.
	7	Classifying	Assigning concepts or objects to categories based on shared attributes or characteristics.	8	Explaining	Articulating the underlying mechanisms, logic or principles of a concept, process, or phenomenon.
	9	Summarizing	Condensing information into a concise representation that captures its central theme or main ideas.			
Apply	10	Executing	Directly applying data, information or procedures to solve problems.	11	Implementing	Applying knowledge in a new context to find solutions to new problems.
Analyze	12	Differentiating	Determine the item in a set of concepts is irrelevant or inconsistent with the others.	13	Organizing	Arranging relevant information into systematic and coherent structures according to rules or principles.
	14	Attributing	Determine a point of view, a bias, values, or intent underlying given material.			
Evaluate	15	Checking	Assessing whether a statement, outcome, or procedure conforms to established standards or theory.	16	Critiquing	Evaluating an argument or idea by identifying weaknesses, limitations, or alternative perspectives.
Create	17	Generating	Come up with original ideas, possibilities, or solutions in response to a given problem or prompt.	18	Planning	Devising a detailed procedure for accomplishing tasks.
	19	Producing	Creating a product by applying knowledge, skills, or resources to transform ideas or descriptions into concrete outcomes.			

Figure 1. Nineteen cognitive processes in Bloom's taxonomy, grouped by six levels. Adapted from Anderson and Krathwohl (2001)

Gestures, on the other hand, can serve as an external manifestation of cognitive processes, as illustrated by the spontaneous hand movements that accompany speech (Driskell & Radtke, 2003; Goldin-Meadow & Beilock, 2010; McNeill, 1992). In cognitive science, the theory of embodied cognition posits that human thought is not confined to internal symbolic computation but is fundamentally grounded in bodily actions and perceptual experiences (McNeill, 1992; Skulmowski & Rey, 2018). The interaction between the internal and external systems is considered bidirectional: On one hand, gestures function as reflections of sensorimotor representations that accompany conceptual processing (Hostetter & Alibali, 2008; Singer et al., 2008). For example, people naturally gesture by tracing invisible paths in the air when trying to remember the route to a location (Galati et al., 2018), or use gestures to represent their spatial reasoning during assembling (Rawat & Barrera Machuca, 2025). By externalizing internal representations into physical space, individuals can reduce working memory demands and improve performance on concurrent tasks (Cook et al., 2012; Goldin-Meadow et al., 2001). In this sense, gestures function as external cognitive aids that scaffold the mental organization of the information. Gestures not only reflect thoughts but also reshape them. Supported by embodied learning theory, intentionally engaging the body during learning can enhance knowledge comprehension, retention, and transfer. Gestures allow learners to embody abstract content through physical movement (Nathan & Martinez, 2015). When individuals perceive abstract concepts, such as time or causality, gestures can help concretize otherwise invisible properties (Alibali et al., 2011; McNeill, 1992). Encouraging individuals to use gestures to simulate spatial relations has been shown to improve spatial reasoning (Chu & Kita, 2011) and long-term memorization (Cherdieu et al., 2017). Furthermore, learners who gesture while solving math or language tasks demonstrate deeper understanding and better transfer of knowledge (Novack et al., 2014). These studies showed that gestures can provide multifaceted cognitive benefits in information processing tasks.

Three-dimensional (3D) immersive environments offer a natural medium for using gestures to perform cognitive tasks. First, their spatial affordances can externalize intermediate information representations through spatial visualization cues, thereby supporting spatial memory and reducing cognitive load

(Jian & Abu Bakar, 2024). Second, they facilitate embodied interaction by placing users and digital content within the same 3D interactive space, where information can be manipulated through meaningful bodily actions (Riva, 2008). This has been shown to support information comprehension and memorization (Bagher et al., 2021; Huang et al., 2023).

As a concrete form of embodied interaction, gestures can serve as vocabulary for systematically mapping bodily actions to specific cognitive tasks. The gesture elicitation method can help identify gestures for given functions (Cheng et al., 2024; Pereira et al., 2015; Xia et al., 2022). It is a widely applied, user-centered approach for generating usable gestures and revealing users' underlying mental models (Luo et al., 2025; Tian et al., 2024; Wobbrock et al., 2009). Specifically, end users are individually shown the desired effect of an action (a "referent") and asked to propose the symbol (i.e., gesture). Results are often reconciled into a canonical gesture set using agreement-based metrics such as agreement score (Wobbrock et al., 2009) or consensus score (Morris, 2012). These metrics quantify the degree to which participants converge on a particular gesture for a given referent. Prior research has effectively adopted the gesture elicitation method across a wide range of computing platforms, including tabletop interfaces, television, projection-based systems, and VR/AR systems (Morris, 2012; Tian et al., 2024; Vatavu & Zaiti, 2014; Wobbrock et al., 2009). It has also been applied in diverse immersive applications, such as 3D design, virtual shopping, and virtual museum (Liu et al., 2024; Sun et al., 2024; Wu et al., 2022). Moreover, this method has demonstrated effectiveness for abstract tasks and non-visual modalities, such as spatial reasoning, conveying emotions, or olfactory interactions (Li et al., 2025; Luo et al., 2024; Rawat & Barrera Machuca, 2025). Building on its robustness across diverse media and task types, gesture elicitation represents a promising approach for supporting the embodied articulation of information processing tasks.

Prior works reveal a gap in understanding how information processing tasks can be performed through gestures in immersive environments. Addressing this gap, this research adopted the 19 cognitive processes outlined in the revised Bloom's taxonomy as representative information processing tasks to explore user-defined gestures for articulating them. There were two main research questions (RQs): **RQ1.** What gestures did users commonly use to articulate these information processing tasks? **RQ2.** How do the proposed gestures fulfill their intended functions and perform in terms of usability? To address these questions, a gesture elicitation study was conducted. The elicited gestures were analyzed across five characteristic dimensions through descriptive statistics. Based on consensus scoring, 32 gestures formed a candidate set. A subsequent study assessed the candidate set and identified 20 recommended gestures for future immersive applications. Drawing on findings from both studies, this work further contributes strategies for mapping cognitive processes to physical actions, and a set of design considerations for designing and implementing usable gestures in immersive environments.

Study I: Gesture Elicitation Study

Methods

Referent Design. Referents serve as conceptual anchors that represent the intended meaning or function to be conveyed through gesture. They typically illustrate the "before" and "after" states of the intended function. To create appropriate referents, the authors synthesized information from multiple sources, including instructional design (Qamar et al., 2016; Villanueva, 2015) and the definition of lexical terms, to establish a working definition and representative example for each information processing task ("task"; Appendix A, Figure A1). For example, for Summarizing, the working definition was "Condensing information into a concise representation that captures its central theme or main ideas." Accordingly, the "before" state presented a detailed paragraph, and the "after" state displayed a concise outline of its main ideas (Figure 2).

The referent content was designed to avoid requiring specialized domain knowledge from the participants. It was based on common knowledge (e.g., fruits, animals), pre-college science concepts (e.g., photosynthesis), and online news articles (e.g., discussions of AI ethics). As each cognitive process in Bloom's taxonomy may correspond to various types of tasks in practice, the referent contents adopted the most representative and generalizable ones whenever possible to support broader applicability. They also accounted for the need to distinguish cognitive processes with potentially overlapping meanings. For example, Comparing emphasizes identifying distinguishing features between items, whereas Differentiating focuses on identifying an item that differs from others within a set (Figure 2). Most referents were implemented as static text-based information panels, except for Interpreting, which included a chart, and Producing, which included a 3D model.

Participants. Study I recruited 15 participants (8 males, 7 females) aged 22 to 30 years (M = 25.13, SD = 2.61; Appendix B, Table B1). The sample size aligns with earlier findings suggesting that 11–20 participants can support the effective generation of gesture sets (Vogiatzidakis & Koutsabasis, 2018). Since this study was conducted in VR, participants' familiarity with VR was recorded. Among all participants, 11 reported having used VR before, while four had not. Most participants described their familiarity as neutral (n = 9), followed by unfamiliar (n = 3), familiar (n = 2), and very unfamiliar (n = 1). Regarding gesture input device usage, six participants had prior experience, whereas nine had none. As for gesture input familiarity, responses were more varied: six participants reported being unfamiliar, five neutral, three familiar, and one very unfamiliar. This study complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board of North Carolina State University (IRB Protocol #27188). Informed consent was obtained from each participant.

Experiment Procedure. Upon informed consent, participants first completed a demographic questionnaire. They then proceeded to

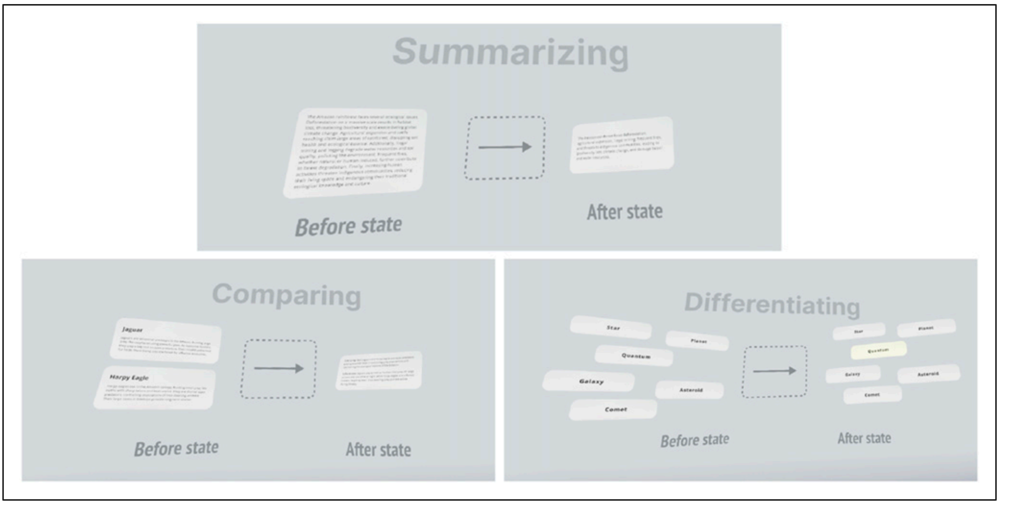


Figure 2. Referents used in the experiment for Summarizing (top), Comparing (bottom left), and Differentiating (bottom right), captured in the immersive environment

the gesture elicitation phase, which was conducted in a virtual environment developed using Meta XR Core SDK, C#, and Unity 3D (2018.4.28f1) and delivered through a head-mounted display (HMD; Oculus Quest 3, Meta). The experimenter verbally introduced an information processing task by providing its working definitions. The order of the task was randomized and counterbalanced across participants. The corresponding referent was then presented in the virtual environment. Participants could ask clarification questions as needed to ensure a clear understanding of each task. Once they indicated that they understood the task, they were asked to design at least three appropriate and distinct gestures to bridge the transition between the "before" and "after" states for each referent, and to perform each gesture three times. They were then asked to explain the rationale behind their designs. Participants' movements were video recorded using a webcam positioned on the desk in front of them, and their verbal expressions were captured through audio recording. Each gesture design was documented in real time by the experimenter through written notes. The experiment lasted 90 minutes.

Data Analysis

Each gesture's characteristics were extracted from video recordings, supported by participants' verbal explanations and the experimenter's notes. The classification scheme was adapted from prior gesture elicitation studies (Andrei et al., 2024; Cheng et al., 2025; Wobbrock et al., 2009), comprising five dimensions: hand usage, articulation level, form, nature, and binding (Figure 3). Two researchers independently coded the gesture characteristics based on the classification scheme. Before coding, both researchers practiced applying the coding criteria and reconciled their coding decisions using a small subset of gesture samples to establish a shared understanding of the categories. They then independently coded all remaining gestures while blinded to one another's decisions. Inter-rater reliability was calculated separately for each of the five coding dimensions, yielding a mean Cohen's κ of 0.977 (SD = 0.024), indicating almost perfect agreement (McHugh,

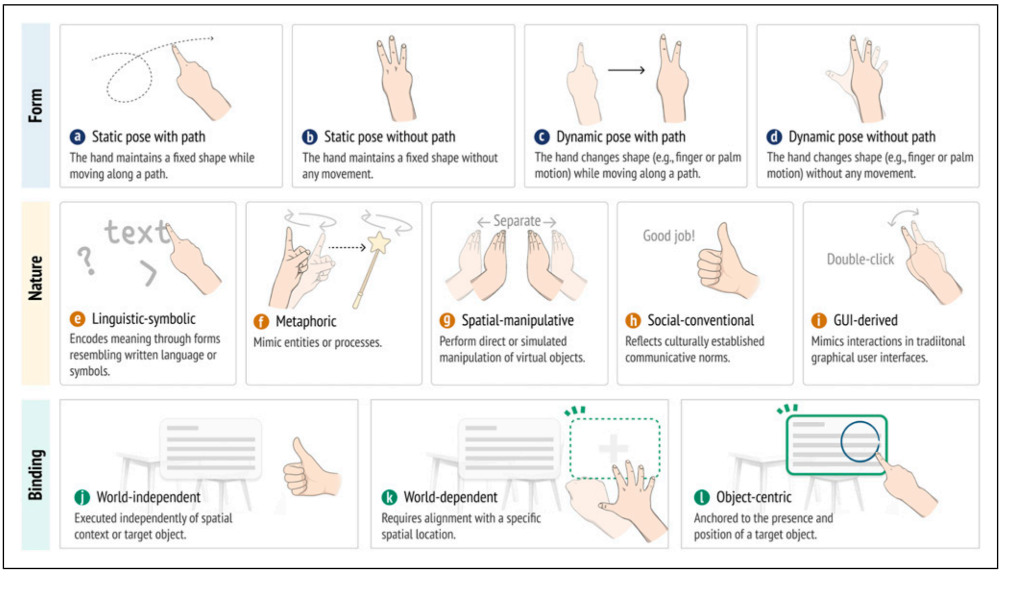


Figure 3. Illustrations of the categories used for classifying gesture characteristics in terms of form, nature, and binding

2012). Any discrepancies were subsequently resolved through discussion to reach consensus on a final code for each gesture.

To identify a set of gestures with practical applicability from the full pool of elicited gestures, this study adopted the max-consensus scoring metric proposed by Morris (2012) rather than the agreement score proposed by Wobbrock et al. (2009), as the latter typically assumes that participants provide a single proposal per referent. Max-consensus scores were specifically developed for multi-proposal gesture elicitation studies (Danielescu & Piorkowski, 2022). Although agreement scores have been widely adopted in VR/AR gesture elicitation research, max-consensus scores have so far been used primarily in gesture elicitation studies for 2D interfaces. Nevertheless, similar to agreement scores, max-consensus is calculated independently of the interaction context in which a gesture is performed. On this basis, the present study extends the use of the max-consensus metric for gesture elicitation to immersive environments. The metric specifies that, for each referent, the set of gesture proposals is denoted as $G = \{g_1, g_2, \ldots, g_K\}$, where gesture g_i is selected by c_i of the N participants. The max-consensus score is defined as:

$$\text{Max} - \text{consensus Score} = \max_{i \in G} \left(\frac{c_i}{N} \right) \quad (1)$$

The inclusion criterion for tasks in the candidate gesture set has been heuristically defined as a max-consensus score ≥0.20 (i.e., at least three participants proposing the same gesture). To ensure consistency in reporting the frequencies of gestures, two normalization procedures were applied: (1) Gestures that differed only in the hand used (left vs. right), in variations of tracing the same symbol, or in the number of repeated shapes used to convey quantity (e.g., drawing two or three circles) were treated as equivalent; (2) Gestures with the same path but different finger configurations were treated as equivalent, as users are generally insensitive to distinctions in finger configuration (Wobbrock et al., 2009). These decisions were made based on participants' self-reported design rationales.

Results of the Elicitation Study

Gesture Articulation Characteristics. Five articulation characteristics were analyzed: hand usage, articulation level, nature, form, and binding. Among all user-defined gestures, 72.41% were unimanual and 27.59% were bimanual for hand usage characteristic. Within the subset of bimanual gestures, 46.55% were classified as symmetric and 53.45% as asymmetric. Regarding the articulation level characteristic, wrist-level movements were the most frequently observed (39.29%), followed by elbow-level (27.11%), no motion (16.41%), and finger-level gestures (14.27%). There were very few shoulder-level movements (2.93%). As for nature characteristic, gestures of linguistic-symbolic nature were the most prevalent, accounting for 34.48% of all gestures, followed by spatial-manipulative (24.14%), metaphoric (23.07%), and social-conventional nature (14.03%). The GUI-derived nature was the least frequent, comprising 4.28% of the total. In terms of gesture form characteristic, static-poses-with-path form accounted for the largest proportion (55.89%) of all user-defined gestures, followed by static-poses-without-path (15.81%), dynamic-poses-with-path (13.32%), and dynamic-poses-without-path (14.98%). The distribution of gesture binding characteristic indicated a predominant use of world-independent bindings (46.97%) of all gestures, followed by object-centric (33.53%) and world-dependent (19.50%) bindings.

Consensus Score and Candidate Gesture Set. The consensus score indicates the extent to which participants elicited the same gesture for a given task. The two most commonly elicited gestures within the same task (i.e., the two with the highest consensus scores) were selected as "Gesture A" and "Gesture B" (Table 1). Tasks Summarizing, Inferring, and Comparing had the highest scores (0.80), followed by Explaining and Generating

Table 1. Two Gestures with the Highest Consensus Scores for each Information Processing Task

		Gesture A	Gesture B
Levels	Tasks	Score	Score
Remember	Recognizing	0.47	0.47
	Recalling	0.27	0.20
Understand	Interpreting	0.33	0.27
	Exemplifying	0.40	0.20
	Classifying	0.47	0.40
	Summarizing	0.80[a]	0.53
	Inferring	0.80	0.20
	Comparing	0.60	0.27
	Explaining	0.67	0.27
Apply	Executing	0.13	0.13
	Implementing	0.27	0.27
Analyze	Differentiating	0.40	0.33
	Organizing	0.33	0.20
	Attributing	0.13	0.07
Evaluate	Checking	0.53	0.33
	Critiquing	0.40	0.27
Create	Generating	0.67	0.33
	Planning	0.47	0.47
	Producing	0.53	0.27

Note.[a]Green cells indicate max-consensus scores ≥0.50, while orange cells indicate max-consensus scores < 0.15.

(0.67). We defined the inclusion criterion as a consensus score above 0.20, corresponding to at least three participants proposing the same gesture. Executing and Attributing fell below this criterion, with max consensus scores of 0.13 each. Both tasks were therefore considered to reflect insufficient consensus for further validation and evaluation and were excluded from the candidate gesture set. Moreover, the two gestures with the highest consensus scores for Producing and Generating were identical in form: one involved opening the fist upward and the other extending the index finger to draw upward circles. We interpret this as an indication that no discriminable gesture was identified for distinguishing the two tasks based on their conceptual distinction within our elicitation results. For the practical purpose of building a testable set in which each task corresponds to unique gestures, we collapsed the two tasks under the single label "Generating." In total, the candidate gesture set comprises 32 gestures representing 16 tasks (Figure 4).

Findings From the Elicitation Study

Mapping Cognitive Tasks to Hand-Based Gestures. The user-defined gestures differed considerably from the interaction forms of traditional 2D screen- or surface-based media, as indicated by the limited number of traditional GUI-derived gestures. Instead, participants tended to adopt diverse strategies to map abstract cognitive tasks onto embodied actions, as indicated by the prevalence of spatial-manipulative, metaphoric, and social-conventional gestures, which together accounted for 61.24% of the set.

Spatial manipulation of information was frequently employed to represent cognitive tasks at the "Understand" and "Apply" levels, which involve transformations in information structure or arrangement. These tasks can be readily transformed into mental visual representations and enacted through gestures, such as placing information side by side (Comparing_A), grouping similar information

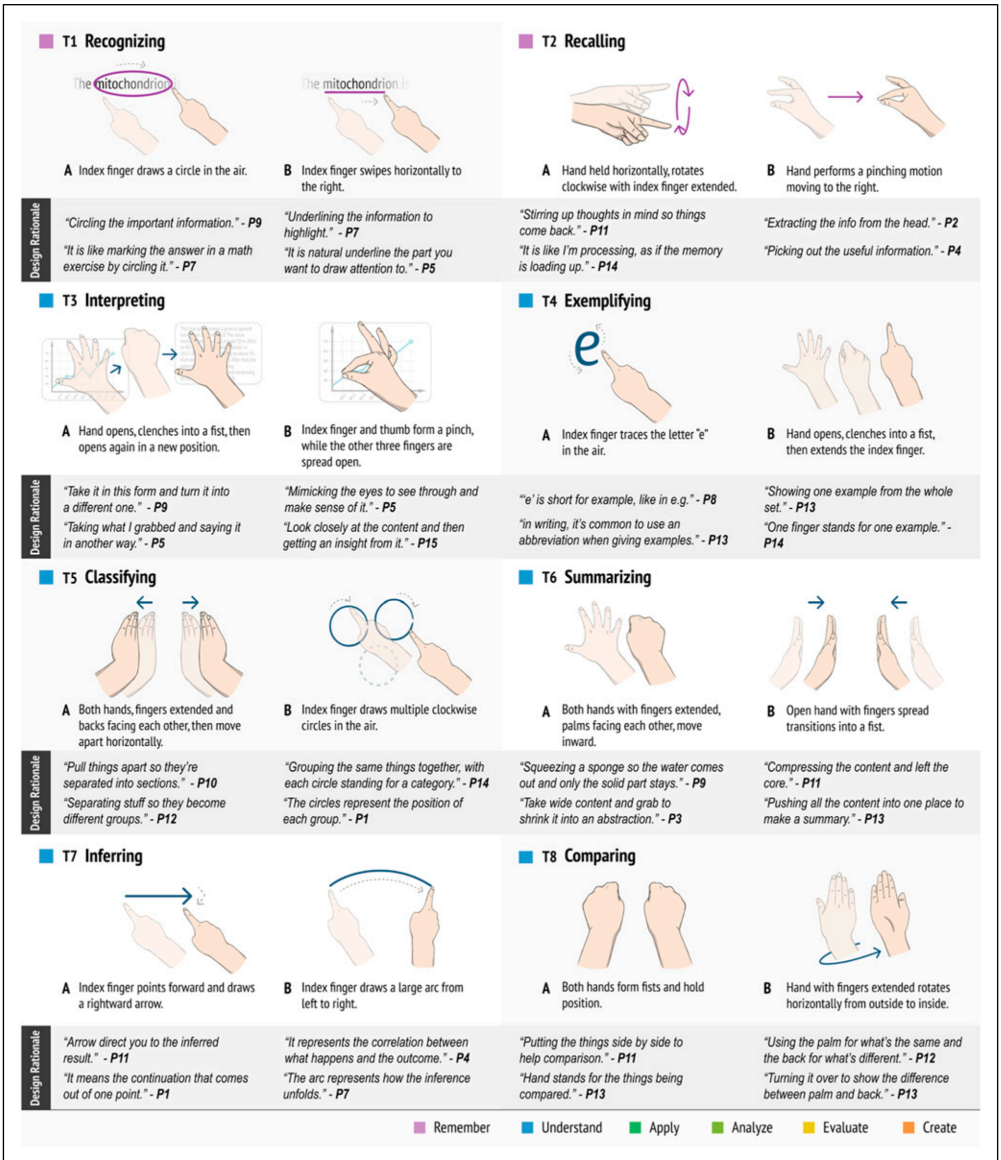


Figure 4. Candidate gesture set from the elicitation study, showing the two gestures with the highest consensus scores (i.e., Gestures A and B) for corresponding tasks (T), and representative design rationales provided by participants (P)

(Classifying_B), or placing information in a sorted structure (Organizing_B). The spatial affordances of immersive environments, including the placement of information panels in 3D space and the free movement of the upper limbs, have prompted participants to engage with information through spatial manipulation. These findings are consistent with prior work on spatial cognition (Rawat & Barrera Machuca, 2025), which shows that immersive environments can serve as a cognitive scaffold for information transformation.

Metaphorical representations of objects, actions, and tools can be used to express cognitive tasks. The flexibility of the fingers and the variety of their possible configurations allow them to mimic a wide range of entities. For example, participants pinched the thumb and index finger to symbolize an "eye" for Interpreting_B, opened the palm upward to represent "blooming" for Generating_A, or rotated the index finger to suggest a "magic wand" for Generating_B. One participant used one hand as a fist to represent the brain while tapping it with the other hand to express "extracting knowledge" for Recalling (Participant 7, hereafter P7). Another used one hand as a sheet of paper while the other mimicked a pen writing steps for Planning (P12). These metaphorical designs showed that users drew on embodied analogies to concretize cognitive tasks and also indicated that metaphors could serve as a bridge between mental concepts and gestural expression.

The presence of the social-conventional nature demonstrated the potential to simulate interpersonal communication as a natural interaction mode in immersive environments. For example, for Critiquing_A, participants waved their index finger side to side to signify "*Oh, I don't think so*." as if conversing with peers; for Inferring, participants spread both of hands open as if asking, "*So what?*" (P5) and for Explaining, a user shook the information panel back and forth to demand, "*Tell me why!*" (P2). Encouraging gesture-based interaction may strengthen the sense of embodiment (Hostetter & Alibali, 2008), thereby making the interaction feel more like an embodied experience than a mere interface operation. This effect may be further reinforced by the nature of the tasks, which involve forms of information processing that also occur in human communication. Together, these factors prompted participants to draw on real-world experiences of interpersonal interaction.

The Prevalence of Tracing Symbolic Forms. One of the most common gesture types was tracing letters or symbols, typically realized as linguistic-symbolic gestures in the static-pose-with-path form. These gestures resembled writing-related actions, using the index finger to trace letters or draw symbols in the air. This result differs from prior findings on tabletop gestures, where physical gestures (that mimicked handling physical objects) had dominated (Wobbrock et al., 2009), or metaphoric gestures dominated in olfactory interactions (Li et al., 2025). This tendency may be attributed to the nature of the information processing tasks, which were closely tied to learning experiences and prompted participants to recall typical writing actions. Moreover, the referents were primarily text-based, which may have reinforced this association with writing experience. These observations suggest that both the task context and the referents design may affect participants' mental models during gesture design. Notably, across 186 instances in which participants proposed linguistic–symbolic gestures, 55.3% occurred in their final (i.e., third) proposals. Such gestures may have served as a fallback strategy when participants ran out of alternative ideas.

Study II: Validation and Evaluation of the Elicited Gesture Set

Study II aimed to verify and evaluate the usability of the elicited gesture set. It assessed the proposed gestures using a combination of subjective and objective measures widely adopted in gesture evaluation research and collected users' comments through interviews.

Methods

Participants. Twenty participants (11 males, 9 females; Appendix B, Table B2) were recruited who were aged 18 to 34 years (M = 25.15, SD = 3.75). This sample size was determined based on prior gesture evaluation studies, which involve between 6 and 24 participants (Bailly et al., 2012; Markussen et al., 2014; Nancel et al., 2011; Sun et al., 2024). Among all participants, 13 reported having used immersive technologies before, while seven had not. In terms of familiarity with immersive technologies, most participants described themselves as unfamiliar (n = 8), followed by neutral (n = 5), familiar (n = 5), and very unfamiliar (n = 2). Regarding gesture input device usage, five participants had prior experience, whereas 15 had none. As for gesture input familiarity, seven participants reported being unfamiliar, six neutral, five very unfamiliar, and two familiar. This study complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board of North Carolina State University (IRB Protocol #28025). Studies I and II were approved under separate IRB protocols because they involved different experimental designs and data collection procedures. Informed consent was obtained from each participant.

Assessment Framework. The assessment framework was adapted from established assessment scales in prior research (Table 2; Chen et al., 2025; Cheng et al., 2024; Xia et al., 2022). Specifically, validation focused on the extent to which the elicited gestures fulfilled their intended functions, measured by objective matching, discoverability, and memorability. Evaluation focused on the usability of gestures, including learnability, ease of use, and user acceptance.

Discoverability, also called guessability or iconically logical toward functionality (Xia et al., 2022), reflects whether users can associate the gestures with their semantic meaning (Wobbrock et al., 2005). It focuses on users' recognition and interpretation of gestures as symbolic representations and is typically assessed through functionality-guessing methods proposed by Nielsen et al. (2003). Specifically, each elicited

Table 2. Assessment framework including objective and subjective measures for gesture evaluation and validation

Categories	Measures	Sub-measures	Metrics	Reference
Objective Measures	*Discoverability*	–	Accuracy of identifying the task name from gesture movement.	(Cafaro et al., 2018; Silpasuwanchai & Ren, 2015)
	Memorability	–	Accuracy of recalling the gesture movement based on task names.	(Céspedes-Hernández et al., 2020; Dong et al., 2015; Zaiţi et al., 2015)
Subjective Measures	*Objective matching*	–	Rate how the gesture aligns with the intended task.	(Dong et al., 2015; Silpasuwanchai & Ren, 2015; Wu et al., 2022)
	Learnability	–	Rate how easy the gesture is to learn.	(Li et al., 2017)
	Ease of use	Ease of execution	Rate how smooth the gesture feels to execute.	(Bhowmick et al., 2022; Ganapathi & Sorathia, 2019; Silpasuwanchai & Ren, 2015)
		Perceived effort	Rate perceived physical effort to perform the gesture.	(Dong et al., 2015; Felberbaum & Lanir, 2018; Ganapathi & Sorathia, 2019)
	User acceptance	Preference	Rate how much the gesture is liked.	(Dong et al., 2015; Ganapathi & Sorathia, 2019; Ortega et al., 2019)
		Social acceptability	Rate comfort using the gesture in the public settings.	(Dingler et al., 2018)
		Willingness to use	Rate likelihood of using the gesture in real applications.	(Li et al., 2017)

gesture was filmed as a video clip. Participants were asked to observe, guess, and report the corresponding task names associated with the gesture after watching the video. To reduce the impact of forced-choice bias (i.e., participants constrained to select a single option even if multiple alternatives seem appropriate or none seems suitable, which may lead to frustration and reduce data validity), each participant was allowed to provide 0–2 candidate names for each gesture. The scoring criteria differentiated levels of interpretation accuracy. A correct first guess reflects a direct mapping between the gesture and its intended function, whereas a correct second guess indicates a weaker but still meaningful association. Following symbol recognition literature that distinguishes correct, partially correct, and incorrect interpretations using a 0–1–2 scoring scheme (Deng et al., 2025), responses were scored as 2 points for a correct first guess, 1 point for a correct second guess, and 0 points if neither guess was correct.

Memorability captures the immediate retention of learned gesture–function mappings. It reflects the initial encoding of mappings, which plays an important role in transferring these mappings to long-term memory (Baddeley, 2012). Memorability was evaluated by asking participants to recall the two gestures of each task after a 15-minute delay. The use of 15-minute intervals has been shown to significantly reduce the impact of within-session manipulations on recall (Divis & Benjamin, 2014) and has also been adopted in prior evaluations of gesture memorability (Anderson & Bischof, 2013; Sun et al., 2024). A score of 1 was assigned to the response for correct recall and 0 for incorrect or no recall.

Objective matching measures the degree of semantic alignment between each gesture and its intended cognitive function on a 7-point Likert scale, with 7 referring to perfectly matched. **Learnability** was measured on a 7-point Likert scale, with 7 referring to very easy to learn. **Ease of use** dimension assessed the physical ease of performing each gesture in terms of *ease of execution* and *perceived effort* using a 7-point Likert scale, with 7 referring to very smooth to execute or minimal effort required. **User acceptance** was evaluated across three dimensions: *preference*, *social acceptability*, and *willingness to use*. All three items were measured on 7-point Likert scales, with higher scores indicating better outcomes.

Experiment Procedure. The experiment included five phases (Figure 5). Upon informed consent, participants completed demographic questionnaires. In the guessing phase, participants viewed video clips of each gesture three times on a laptop and were asked to

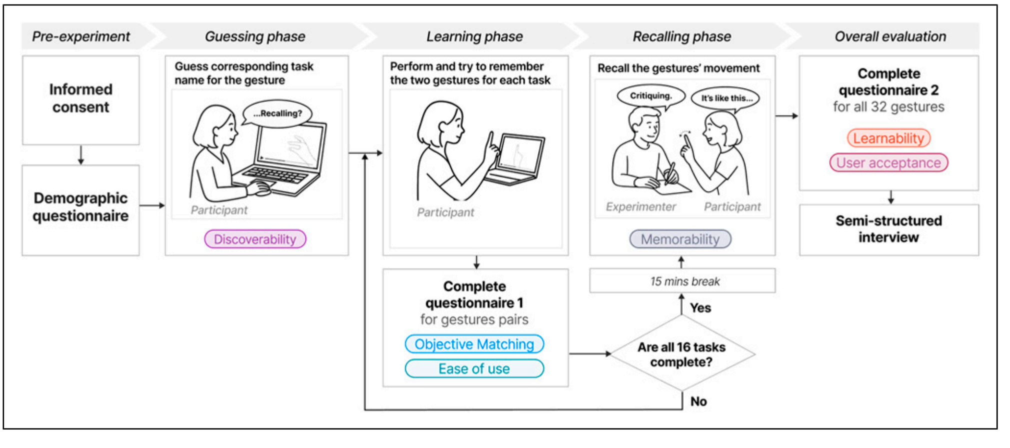


Figure 5. Experiment procedure for the validation and evaluation study

identify 0–2 candidate names for the task from the full list of 32 task names provided. In the learning phase, participants were presented with gestures along with their corresponding task names. They were asked to perform each gesture three times while memorizing both the gesture movements and their corresponding task mappings. After viewing the two gestures for a given task, participants filled out a brief questionnaire to assess them. Next, a 15-minute break was provided to reduce the influence of within-session manipulations, such as presentation order, on gesture recall. Participants were then asked to recall and perform the two gestures for each task announced by the experimenter. In the evaluation phase, participants completed a post-study questionnaire. Across all phases, the presentation of gesture videos or task names was randomized. At the end of the session, a semi-structured interview was conducted to gather participants' impressions and feedback (Appendix C).

Data Analysis. Objective measures, including discoverability and memorability, were averaged across participants for each gesture. For the subjective measures, the ease-of-use and user-acceptance subscales were first aggregated by calculating their mean scores. Statistical comparisons were then conducted for gesture pairs within each task across four measures: objective matching, learnability, ease of use, and user acceptance. For each comparison, normality was assessed using the Shapiro–Wilk test. If the normality assumption was violated, the nonparametric Wilcoxon signed-rank test was used; otherwise, paired t-tests were conducted. To control for inflated Type I error, p-values were adjusted within each task using Holm's method, with the four measures treated as a single comparison family. Qualitative data from the semi-structured interviews were transcribed and independently coded by two researchers who were blinded to each other's decisions, yielding strong inter-rater agreement (Cohen's κ = 0.81; McHugh, 2012).

Results

Given the breadth of results, the following report highlights gestures at the lower and upper ends of the distribution for each measure. Specifically, gestures at or below the 25th percentile of all gestures were classified as low-scoring, whereas those above the 75th percentile were classified as high-scoring.

Objective Measures. The **discoverability** mean scores varied considerably across gestures, ranging from 0.05 to 1.70 (Figure 6). The highest scores were observed for Critiquing_A, followed by Summarizing_B, Checking_A, Summarizing_A, and Generating_A. Lower scores were found for Interpreting_A, Organizing_A, Inferring_B, Exemplifying_B, Recognizing_B, and Interpreting_B. The **memorability** mean scores spanned a wide range from 0.15 to 0.90. The highest scores were observed for Comparing_A, followed by Organizing_A, Critiquing_B, Critiquing_A, and Checking_A. Lower scores were found for Interpreting_B, Implementing_B, Exemplifying_B, Differentiating_B, Explaining_B, Inferring_B, and Differentiating_A.

Subjective Measures

Questionnaire Responses. The comparison of subjective evaluation scores for gesture pairs (A and B) across four measures is summarized in Figure 7.

The **objective matching** mean scores ranged from 4.20 to 6.16 on a 7-point scale. The highest scores were observed for Generating_A, followed by Inferring_A, and Summarizing_A. In contrast, lower ratings were found for Recognizing_B, Interpreting_B, Implementing_B, and Planning_B. Paired-sample t-tests revealed significant differences between gesture pairs (i.e., the two gestures for the same task) for five tasks: Generating ($t = 2.97$, $p < .05$, Gesture A rated higher than B, and hereafter noted as A > B), Planning ($t = 3.80$, $p < .01$, A > B), Interpreting ($t = 3.09$, $p < .05$, A > B), Recalling ($t = 3.38$, $p < .01$, A > B), and Recognizing ($t = 2.85$, $p < .05$, A > B).

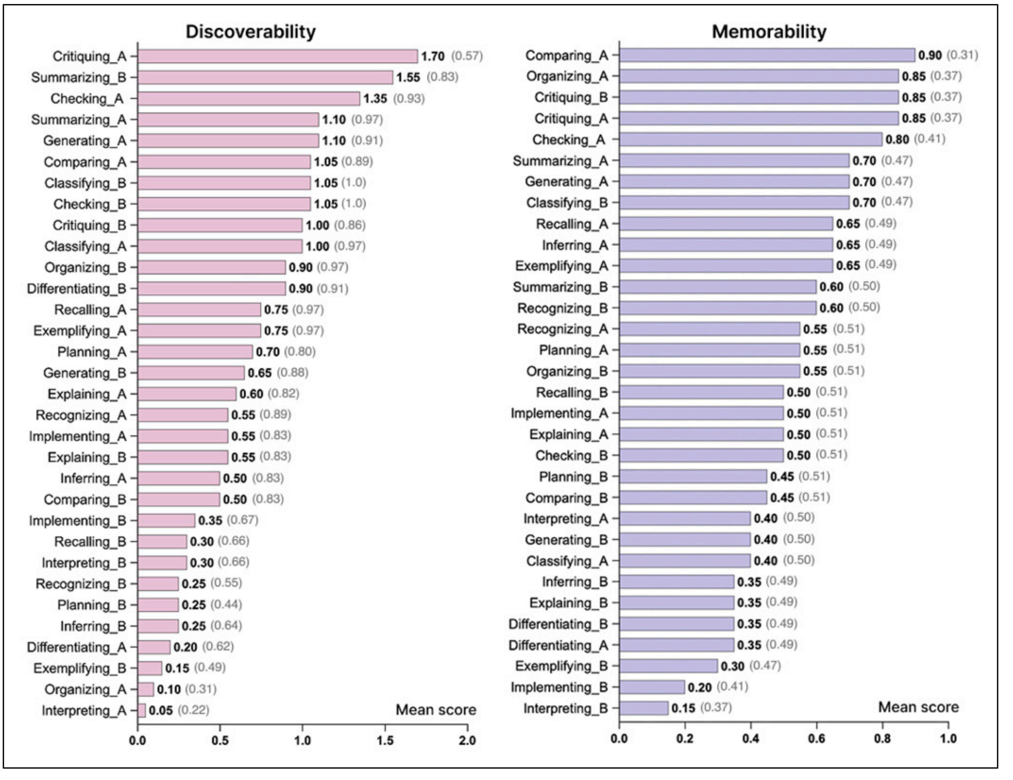


Figure 6. Mean discoverability and memorability scores for gesture variants (A and B) across 16 tasks, with standard deviations shown in parentheses

The **learnability** mean scores ranged from 4.76 to 6.24. The highest scores were observed for Generating_A, followed by Critiquing_B, Summarizing_A, and Critiquing_A. In contrast, lower learnability scores were found for Recalling_B, Interpreting_B, Differentiating_A, and Recognizing_B. Wilcoxon signed-rank tests revealed significant differences between the two gestures for Recalling ($W = 0$, $p < .01$, A > B) and Generating ($W = 3.0$, $p < .05$, A > B).

The **ease-of-use** mean scores were generally high across all gestures, with mean scores ranging from 4.63 to 6.70. The highest scores were observed for Recognizing_B, followed by Inferring_B, Summarizing_A, Critiquing_B, and Generating_A. Lower scores were found for Differentiating_A, Planning_A, Inferring_A, and Exemplifying_B. Significant differences were found for the gesture pairs in the following tasks: Differentiating ($t = -4.77$, $p < .01$, B > A), Inferring ($W = 2.0$, $p < .01$, B > A), Explaining ($t = 3.35$, $p < .05$, A > B), and Summarizing ($t = 3.10$, $p < .05$, A > B).

The **user acceptance** mean scores ranged from 4.14 to 6.24. The highest scores were observed for Generating_A, followed by Inferring_A, Organizing_B, and Planning_A. Lower scores were found for Interpreting_B, Implementing_B, Recalling_B, and Planning_B. Significant differences were found for the gesture pairs across the following tasks: Generating ($t = 3.74$, $p < .01$, A > B), Interpreting ($t = 2.61$, $p < .05$, A > B), and Recalling ($t = 2.59$, $p < .05$, A > B).

Interview Responses. To better understand participants' rating rationales, their comments and feedback in the interviews are summarized as follows:

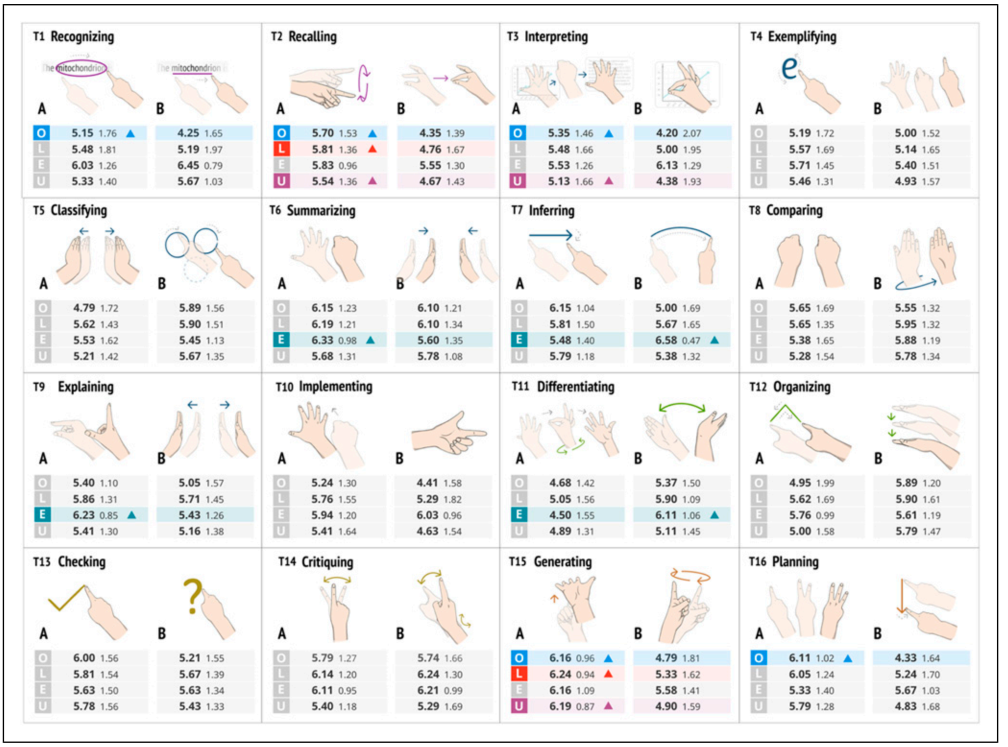


Figure 7. Comparison of subjective evaluation scores for gesture pairs (A and B) across 16 tasks (T), using the following measures: objective matching (O), learnability (L), ergonomics (E), and user acceptance (U). Measures with color-highlighted cells indicate statistically significant differences between gesture pairs, with an upward triangle marking the significantly higher score

Design Features Favored by Users. Gestures with a familiar metaphorical mapping to objects, symbols, or actions in everyday life are easy to understand (P6, P11-12, P15, P19). For example, Generating_A resembled "*a blooming flower to symbolize the emergence of new ideas*" (P6), Checking_A evoked "*the act of marking corrections on assignments*" (P12), and Organizing_B arranged content in a structured order, "*like placing books neatly on a shelf*" (P20). These mappings are commonly used outside of the immersive context (P15), making them "*feel intuitive to relate to*" (P11). Moreover, incorporating spatial relationships and anchors makes gestures easier to understand, learn, and remember (P7-12, P14-15, P17, P20). For example, Comparing_A involved positioning two hands "*side by side to represent contrasting items*" (P11), and Summarizing_A "*condensed extended passages into a smaller spatial area*" (P14). These mappings helped participants to infer meaning from spatial arrangement (P8, P12). Additionally, gestures that mimicked interpersonal communication tended to leave a strong impression on participants (P10, P15, P18), as in the cases of Critiquing_A and Critiquing_B. P15 remarked, "*It felt like I was communicating with a person rather than a machine, and it allowed me to release my emotions.*"

Establish Mappings Through Learning. Some gestures were initially difficult for participants to discover, but once a mapping was established, they were recognized and accepted by participants. Some gestures required additional contextual information to establish reliable associations with intended functions (P8, P15, P20). For example, regarding Explaining_A, P20 remarked: "*Once I knew this *gesture meant explaining, I could imagine it expanding a concept on a larger panel*." Second, generic gesture forms and trajectories could lead to multiple interpretations. For instance, the question mark gesture for Checking_B could also signify a request for explanation in Explaining (P17); the arrow gesture for Inferring_A could also be interpreted as moving to the next step for Implementing (P14); and tracing the letter "e" for Exemplifying_A could likewise be associated with other gestures that share the same initial letter (P6, P11-12). However, participants noted that once they learned the correct one-to-one mappings, these gestures can be easy to understand and acceptable to them (P7, P10, P17).

Physical Effort Concerns. First, gestures involving finger-level articulations and wrist rotations were often perceived as less comfortable, such as Differentiating_A, Planning_A, Recalling_B, and Exemplifying_B (P6, P10-12, P15, P19). Second, gestures requiring multi-step trajectories were considered cumbersome and effortful, such as Inferring_A, Interpreting_A, and Classifying_B (P7, P18, P20). Finally, two-handed gestures were perceived as requiring more effort than one-handed gestures, such as Explaining_B and Summarizing_B (P7-8, P18).

Gesture Form-Related Concerns. First, some gestures were overly simple in form or path, making them insufficient to establish mappings with the intended tasks, especially for Recalling_B, Recognizing_B, and Inferring_B (P8, P12, P15, P17) as they were often reduced to a simple straight or curved stroke. Second, gestures that shared similar forms or motion trajectories with others could confuse and make it difficult to recall. Examples include Exemplifying_B with Implementing_B and Summarizing_A, all of which featured a grabbing motion (P17); and Recalling_A with Generating_B, both involving circular movements of the index finger (P6, P11, P14); Recognizing_B, Inferring_B, and Recalling_B all included a horizontal line trajectory (P11-12, P15, P17); Planning_B and Inferring_A, both had an arrow (P17, P19); and Classifying_A and Explaining_B, both involved outward movements of two hands (P7).

Semantic Concerns. Metaphorical gestures can be interpreted differently across individuals. For example, Interpreting_B was considered difficult to understand and prone to misinterpretation, with participants noting that the gesture could resemble a peacock, an "OK" sign, or a camera lens (P11-12, P15, P17). Additionally, participants expressed concerns that gestures mimicking interpersonal communication may be less appropriate in some public settings, as they could be misinterpreted by passersby (P11, P19). As P19 noted, "*If I were in a café, performing this gesture might make others think I was acting strangely*."

Consolidation of Gestures Into a Recommended Set

While various measures have been widely adopted for gesture assessment, a consolidated framework that integrates these measures for systematic gesture selection had been largely absent from the literature. Driven by this need, this study introduces a consolidation model based on both objective and subjective assessment metrics to support gesture selection. Discoverability and memorability scores were first normalized to a 0–1 scale and then mapped onto the x- and y-axes, respectively. The mean scores of four subjective measures for each gesture were aggregated to provide an overall indicator of the subjective evaluation profile (Figure 8; Han et al., 2025). For clearer presentation of participants' comparative evaluation across gestures, the aggregated scores were min-max normalized and encoded as circles of varying sizes, with larger circles indicating higher scores. To facilitate gesture selection, this study adopted a heuristic thresholding approach. For discoverability and memorability, a cutoff of 0.50 was used.

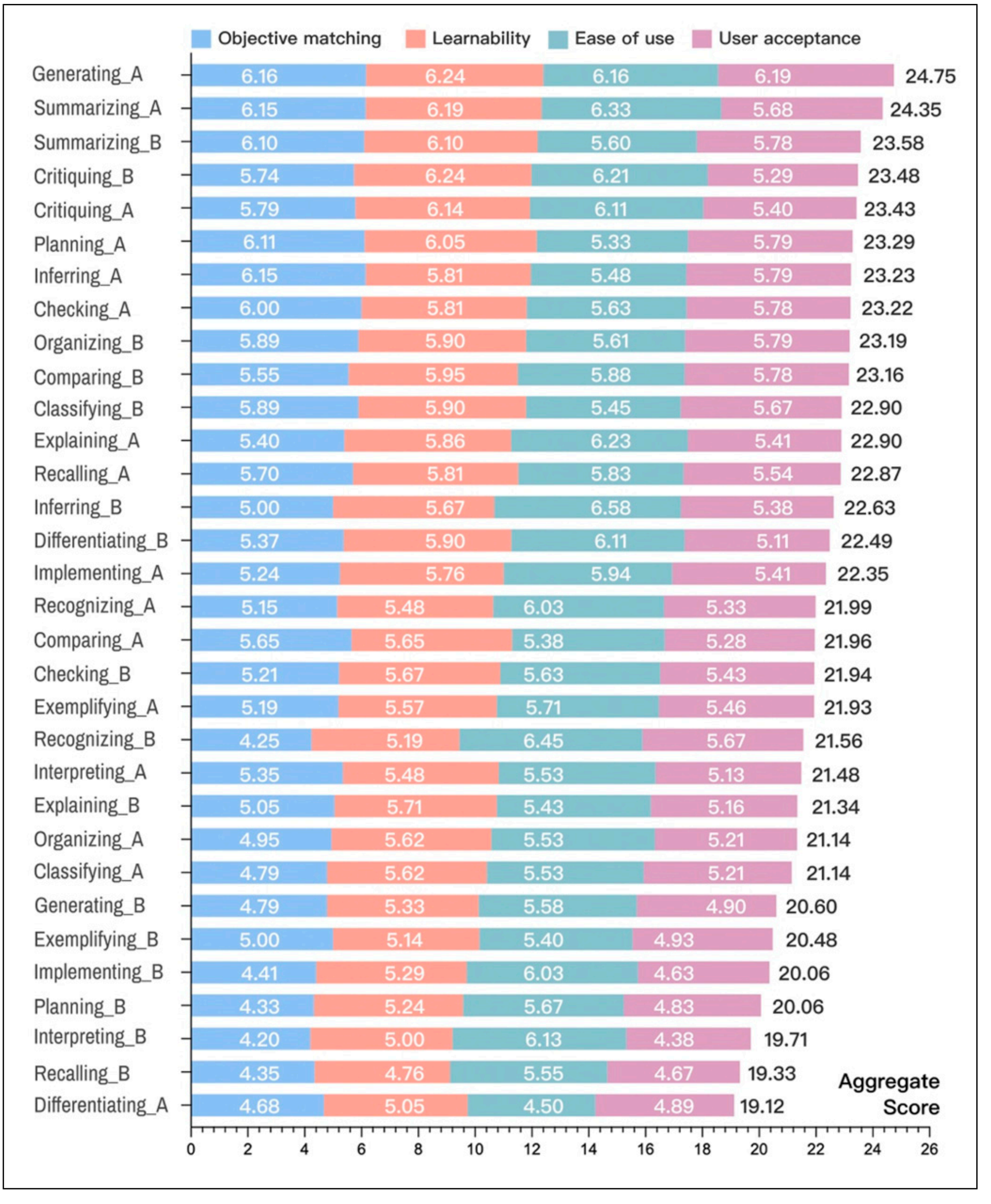


Figure 8. Aggregated scores across candidate gestures based on four subjective evaluation dimensions

For the aggregated subjective evaluations, quartile-based thresholds (25th, 50th, and 75th percentiles) were used to provide fine-grained differentiation. Based on these thresholds, the model organized the gesture set into three regions and represented each gesture as a circle of varying color to facilitate selection (Figure 9).

Gestures in quadrant showed high performance in both discoverability and memorability. Notably, this region contained all gestures whose aggregated subjective evaluations fell within the top 25th percentile, as indicated by the green circles. This suggests consistency between the objective and subjective measures. Therefore, Gestures in quadrant

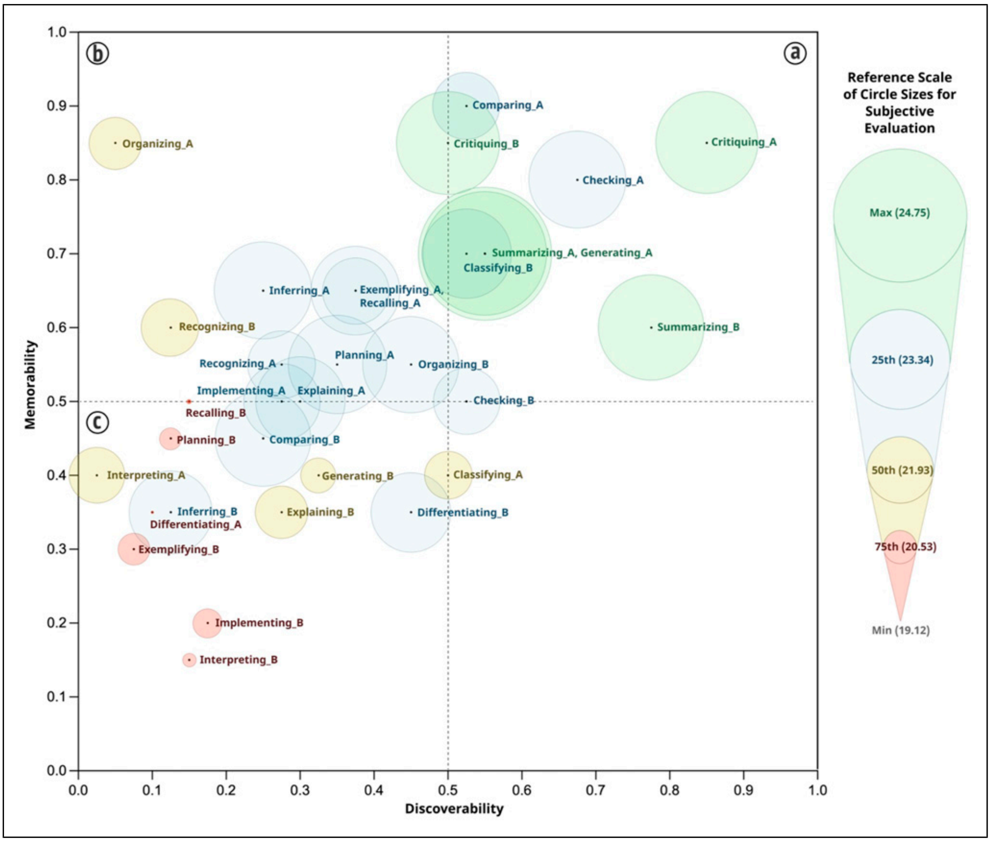


Figure 9. Consolidation model for gesture selection. Each circle represents one candidate gesture, positioned by discoverability (x-axis) and memorability (y-axis). Dashed lines at 0.50 divide the plot into four quadrants, three of which contain candidate gestures and are labeled as regions: (a) high–high, (b) low discoverability–high memorability, (c) low–low. Circle size indicates aggregated subjective evaluation, and circle color denotes the subjective-evaluation tier used for gesture selection

(including those on the boundary lines) were classified as Tier 1. Gestures in quadrant ⓑ showed high memorability but low discoverability. This region also included gestures with lower subjective evaluations (e.g., Organizing_A). Accordingly, gestures in quadrant ⓑ whose subjective evaluations fell within the top 50th percentile (including those on the boundary lines) were classified as Tier 2. Gestures in quadrant ⓒ performed poorly on both objective dimensions. Nevertheless, some of these gestures still received favorable subjective evaluations (e.g., Comparing_B), suggesting potential for practical use. Therefore, gestures in quadrant ⓒ whose subjective evaluations fell within the top 50th percentile were classified as Tier 3. In total, nine gestures were selected for Tier 1, eight gestures for Tier 2, and three gestures for Tier 3, resulting in 20 recommended gestures (Figure 10).

Discussion

Strategies for Mapping Cognitive Tasks to Physical Actions

To address **RQ1**, Study I identified 32 distinct gestures proposed for 16 information processing tasks and revealed the strategies users employed in mapping cognitive tasks to physical actions. Drawing on the results of

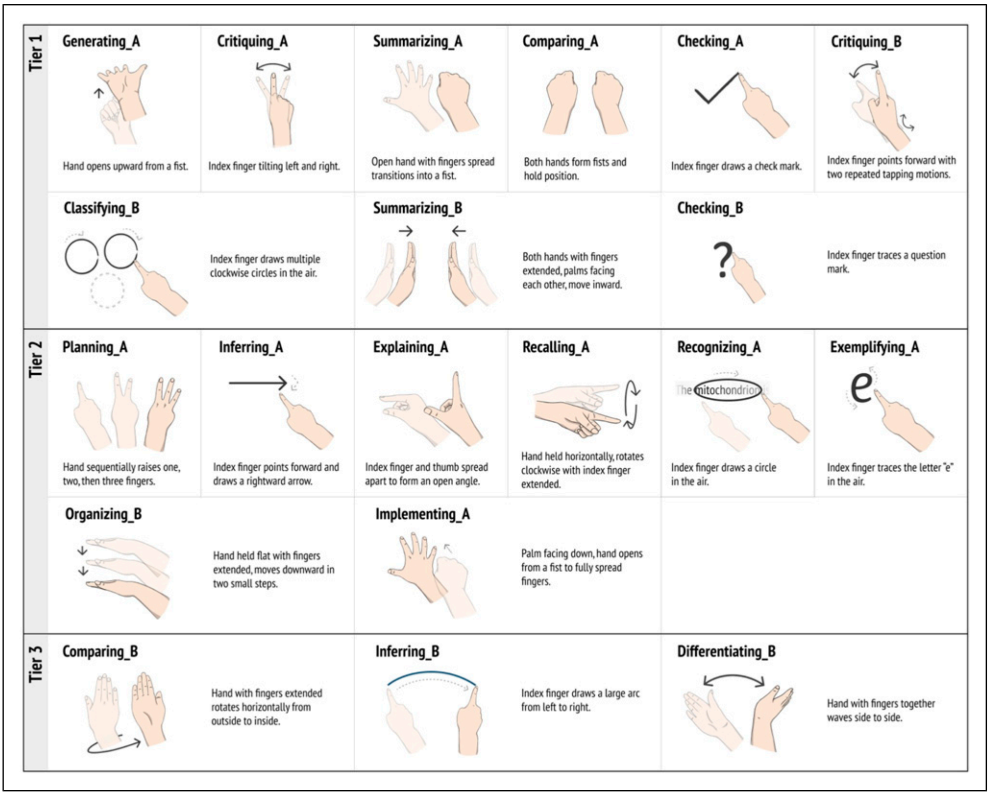


Figure 10. The recommended gesture set, arranged from top to bottom: Tier 1 gestures, Tier 2 gestures, and Tier 3 gestures

Study II, this section further discusses these strategies with respect to usability and in relation to prior research.

Tracing letters or symbols was the most frequently adopted strategy in elicitation. Familiar symbols, such as check marks (Checking_A), helped users associate gestures with cognitive tasks because their meanings were already well established and required little additional learning. However, this type of gesture may face challenges in consistency, especially for tracing letters. When multiple commands share the same initial letter, interference may arise in remembering the intended gesture–command mappings, as illustrated by Exemplifying and Explaining (P6, P11-12). In addition, tracing letter gestures may carry different meanings across gesture sets. For example, tracing an "M" denoted "Music" in office commands (Andrei et al., 2024), whereas in TV interaction (Vatavu & Zaiti, 2014), it represented "open Menu." Such inconsistencies within and across gesture sets can impair the memorability of gesture-command mappings. Therefore, these gestures should be used sparingly and supported with sufficient reminders or cues. Nevertheless, this study found that tracing letters and symbols may offer a promising fallback option for abstract tasks that are difficult to represent through other gestural inputs.

Mapping cognitive tasks to spatial manipulation was the second most frequently used strategy. Clear spatial mappings can help users understand and remember gestures (P7-12, P14-15, P17, P20). This type of gesture may offer additional benefits in cognitive tasks. As suggested in prior studies, simulating mental operations in space can enhance comprehension of content (Alibali, 2005; Fugate et al., 2019; Kirsh, 2010), when coupled with embodied actions, can reinforce the consolidation of information into long-term memory (Sullivan, 2018). Spatial manipulative gestures are also prevalent in other immersive tasks, such as assembling (Rawat & Barrera Machuca, 2025), 3D design (Sun et al., 2024), and document management (Luo et al., 2025). By comparing prior gesture sets, this study found that these spatial transformation gestures recurred in similar forms across different tasks. For example, Comparing_A is similar to the "Comparison of two blocks" gesture for assembling (Rawat & Barrera Machuca, 2025), and Classifying_B is similar to the "Split groups" gesture in organizing documents (Luo et al., 2025). These findings suggest the potential to establish a consistent set of "action primitives" that can serve as a shared gesture vocabulary for manipulating virtual information and virtual objects (Hosseini et al., 2024).

Using metaphors to create gestures was the third most frequently used strategy. This study confirmed previous findings that metaphors serve as a bridge between experiences in the physical world and abstract conceptual language (Cienki & Müller, 2008; Larsson & Stolpe, 2023; Roth & Lawless, 2002). However, consistent with prior findings (Chui, 2011; Cienki & Müller, 2008; Quinn, 2008), this study also found that the interpretation of metaphoric gestures could be diverse and shaped by individuals' mental models (P11-12, P15, P17). This presents challenges for their broader standardization and general use. Still, their value may become more evident as personalized systems continue to advance, particularly when such systems can support users in constructing customized metaphorical mappings and accurately recognizing these gestures (Oh & Findlater, 2013; Zou et al., 2025).

This study also revealed that gestures from interpersonal communication hold potential for use in computing systems. These gestures may elicit emotional responses by evoking users' episodic memories of their use in interpersonal settings (Levine & Edelstein, 2009; Tyng et al., 2017), thereby making the actions engaging and memorable (P10, P15, P18). This type of gesture enriches the expressive bandwidth of existing interaction paradigms by incorporating affective dimensions for conveying nuances (Luo et al., 2024). While interaction on conventional interfaces, such as mouse clicking or screen tapping, offers limited capacity to convey affective expressions, gestures may support this through socially grounded actions and variations in movement (e.g., force, speed, or openness). Additional interpersonal behaviors, such as head shaking and shrugging, may be drawn from social interaction research and incorporated into interaction design (Bavelas & Gerwing, 2007; McNeill, 2018). Such integration may cultivate a greater sense of intimacy in human-computer relationships, promoting more engaging experiences and facilitating trust. However, the use of interpersonal gestures can be context sensitive (P11, P19). It would be beneficial to delineate the contexts in which these gestures are appropriate or to provide users with contextual reminders regarding their suitability.

Design Considerations for Developing and Implementing Usable Gestures

To address **RQ2**, Study II examined how the elicited gestures represented their intended cognitive functions and how they performed across multiple usability dimensions. The recommended gesture set can support a range of immersive applications that involve information processing tasks, such as learning, sense-making, and document work. For instance, the Comparing gesture could let users juxtapose documents or visualizations side by side, the Summarizing gesture could condense key points from a lengthy article into a brief note for quick review, and the Explaining gesture could support annotating content to articulate understanding while reading or browsing. Given that the full gesture set comprises 20 gestures, adopting the entire set may impose considerable memory burden on users. We therefore recommend that immersive applications select an appropriate subset based on their specific task needs. Furthermore, synthesizing both objective and subjective results, this study proposes four design considerations to inform the gesture design and implementation in future practice.

Information as an Interactive Entity in Immersive Environments. Certain gestures demonstrated notably strong performance, as reflected in their placement within Tier 1 of the recommended gesture set. Drawing on interview feedback, the success of these gestures points to the potential of conceptualizing information as an interactive entity, suggesting that users may perceive information not merely as static content, but as entities with physical, metaphorical, and even human-like properties. First, information can be positioned, grouped, or scaled in 3D space (P6, P12, P20). This enables real-world physical manipulations to be transferred into gestures, such as compressing items to summarize them (Summarizing_A). Second, metaphoric mappings can be attached to information entities (P11, P14). This enriches the ways information can be represented and allows object-oriented gestures to be applied to information, such as arranging information like books on a shelf. Third, information entities can also incorporate human-like qualities to facilitate more natural communication with users (P10, P15, P18). They may convey emotions through animations or typography and respond in expressive ways that align with social norms. In practice, these properties provide a useful basis for exploring richer information representations through features such as shape, spatial layout, and skeuomorphic elements, which can in turn support more engaging and expressive interaction design.

Designing for Ergonomics and Semantic Expressiveness. The findings of this study highlight ergonomics and semantic expressiveness as two important factors shaping gesture usability. First, consistent with prior research, certain design features were found to reduce the ease of use, such as fine finger articulations (Rempel et al., 2015), multi-step trajectories (Rekik et al., 2014), and two-handed gestures (Morris et al., 2010). These features can make gestures effortful to perform and lead to fatigue during prolonged use. On the other hand, gestures can be easy to execute but lack sufficient distinctiveness to convey their intended meanings (P8, P12, P15, P17). A representative example is Inferring_B, which received high ease-of-use ratings but showed very low objective matching due to ambiguity in interpretation. Furthermore, when ergonomic limitations and insufficient semantic expressiveness co-occur in a single gesture, they can jointly contribute to gesture failure, as illustrated by Recalling_B, which received notably low user acceptance. In practice, designing usable gestures requires deliberate consideration of both ergonomic efficiency and semantic clarity. Based on the findings of this study, usable gestures tend to involve simple one-handed movements and clearly interpretable forms, while minimizing fine finger articulation and multi-step trajectories.

Similarity Across Gestures. Beyond the properties of individual gestures, this study also found that similarity among gestures can affect usability. Similarities can interfere with users' interpretation and memorization of gestures, as reflected in reduced performance on memorability (e.g., Exemplifying_B and Explaining_B) and discoverability (e.g., Inferring_B and Recognizing_B). Based on interview responses, gesture similarity can arise across three dimensions, including hand shape (e.g., grabbing in Exemplifying_B and Summarizing_A; P17), movement direction (e.g., outward motions in Classifying_A and Explaining_B; P7), and motion path (e.g., horizontal strokes in Inferring_A, Recognizing_B, and Recalling_B, P12). In some cases, the similarity may not involve the entire gesture but only a shared motion primitive (i.e., a basic movement unit, such as grabbing) that overlaps with another gesture. These similar motions remain distinct in their semantic meanings, which may lead to confusion and memory interference. For example, the outward motion in Classifying_A represented separating different groups, whereas in Explaining_B it represented expanding content. This issue may reflect a broader challenge of the end-user gesture elicitation method, as users typically do not intentionally differentiate each gesture during elicitation and are also unaware of the gesture design produced by others. This suggests the need for a set-level differentiation mechanism to identify and reduce gesture similarity during the gesture development process. In practice, the identified dimensions can help designers identify overlapping motion primitives and strategically select a gesture subset to avoid conflicts and accidental activation across gestures.

Consolidation Model and Strategic Gesture Implementation. This study proposes a consolidation model that integrates both objective and subjective assessment metrics to support gesture selection. This model allows designers to recognize potential usability issues and further inform the development of gesture tutorials and reminders. For instance, Tier 2 gestures, which are difficult to discover but easy to remember, may appear non-intuitive at first yet become intuitive once learned. This acquired intuitiveness is supported by both interview feedback (P7, P10, P20) and prior research (Xia et al., 2022). Acquired intuitiveness may be due to the need for more contextual information to be understood (e.g., Explaining_A) or are subject to multiple interpretations (e.g., Exemplifying_A). Therefore, the implementation of such gestures may benefit from contextual cues or additional explanations during a training phase (Sun et al., 2024). Tier 3 gestures were not only difficult to recognize but also difficult to remember (e.g., Differentiating_B). Beyond tutorials, such gestures may require repeated exposure, visual reminders, or progressive training to reinforce gesture–function associations. This model has the potential to be extended to other gesture evaluation studies and may inform the development of more comprehensive gesture assessment and selection frameworks.

Limitations

This research acknowledges several limitations. For Study I, first, a few cognitive processes in Bloom's taxonomy inherently involve subtle conceptual distinctions, such as executing versus implementing, which may not be optimally represented or differentiated through information-panel-based tasks. Second, certain cognitive processes may have been unfamiliar to some participants (e.g., attributing). The ambiguity and unfamiliarity can affect participants' interpretation of the tasks and result in lower consensus during gesture design. This also suggested that not all cognitive processes in Bloom's taxonomy can be directly translated into executable tasks in immersive environments. Some may require more specific contextual support and task design. Additionally, cultural background may also have shaped how participants interpreted tasks and produced gestures, particularly those of a social-conventional nature.

For Study II, first, the visual presentation of the referent could have influenced participants' understanding of the tasks. This was particularly evident in the Organizing task, where the tree-like structure of the referent shaped the symbolic form of participants' gestures. Second, the use of video clips to assess gesture discoverability may not fully capture how gestures are perceived and interpreted within immersive environments. Third, although participants were prompted to imagine using the gestures in public settings when rating social acceptability, these judgments may not fully reflect real-world situations because the study was conducted in a controlled laboratory environment. Lastly, the results of Study II may not reflect the usability of the gesture set in longitudinal usage.

Moreover, the gesture set's generalizability is likely bounded by referent–task alignment and population differences: gestures may not transfer well to substantially different presentation formats, and interpretation may vary across age and familiarity with immersive technology. Lastly, this study focused on gesture as the sole modality. It can avoid disrupting public spaces, support embodied cognition, and offer greater expressive diversity compared to using voice, controller, and gaze alone, respectively. However, it is worth further exploring how gesture could be combined with these modalities to further enhance interaction experience.

Conclusion

This study investigated the use of hand-based gestures for information processing tasks in immersive environments through both an elicitation study and a validation and evaluation study. A candidate set of 32 gestures and a recommended set of 20 gestures were identified. In capturing gestures, this study identified four prevalent gesture nature types for performing information processing tasks: linguistic-symbolic, spatial-manipulative, metaphoric, and social-conventional. In assessing these gestures, the study further revealed factors that influence gesture usability. Building on these findings, it discussed strategies for mapping cognitive tasks to physical actions and offered design considerations for creating and implementing usable gestures. Together, this work represents a valuable step toward more natural and efficient interaction techniques for performing cognitive tasks in immersive environments.

Key Points

- Sixteen of the 19 cognitive processes in the revised Bloom's taxonomy were mapped to hand-based gestures for information processing in immersive environments.
- Gesture design strategies in immersive 3D interfaces differ from those in traditional 2D GUIs, with linguistic-symbolic, metaphoric, spatial-manipulative, and social-conventional mappings emerging as prevalent strategies.
- The study identifies 32 candidate gestures and a recommended subset of 20 gestures for supporting information processing tasks in immersive applications.
- Four design considerations for developing and implementing usable gestures are proposed.

ORCID iDs

Fangyuan Cheng https://orcid.org/0009-0001-9306-8471
Emily Fang https://orcid.org/0009-0004-9898-4438
Karen B. Chen https://orcid.org/0000-0003-2904-1394

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, *5*(4), 307–331. https://doi.org/10.1207/s15427633scc0504_2

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144. https://doi.org/10.1177/0956797611417722

Anderson, F., & Bischof, W. F. (2013). Learning and performance with gesture guides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'13*, 1109–1118. ACM. https://doi.org/10.1145/2470654.2466143

Anderson, J. R. (1990). *The adaptive character of thought* (pp. xii, 276). Lawrence Erlbaum Associates, Inc.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. Longman.

Andrei, A.-T., Bilius, L.-B., & Vatavu, R.-D. (2024). Take a seat, make a gesture: Charting user preferences for On-Chair and from-chair gesture input. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3613904.3642028

Arms, L., Cook, D., & Cruz-Neira, C. (1999). The benefits of statistical visualization in an immersive environment. *Proceedings IEEE virtual reality* (Cat. No. 99CB36316), 88–95. IEEE. https://doi.org/10.1109/VR.1999.756938

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(63), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bagher, M. M., Sajjadi, P., Wallgrün, J. O., La Femina, P. C., & Klippel, A. (2021). Move the object or move the user: The role of interaction techniques on embodied learning in VR. *Frontiers in Virtual Reality*, *2*(1), 695312. https://doi.org/10.3389/frvir.2021.695312

Bailly, G., Müller, J., & Lecolinet, E. (2012). Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. *International Journal of Human-Computer Studies, Special Issue on Developing, Evaluating and Deploying Multi-Touch Systems*, *70*(10), 673–689. https://doi.org/10.1016/j.ijhcs.2012.05.006

Bavelas, J., & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In *Social communication*. Psychology Press.

Bhowmick, S., Sorathia, K., & Chandra Kalita, P. (2022). Understanding gesture performance for object selection in VR: Classification and taxonomy of gestures in HCI. In *Proceedings of the 12th Indian Conference on Human-Computer Interaction, India CHI', 21*, 90–93. ACM. https://doi.org/10.1145/3506469.3506483

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. In *Handbook 1: Cognitive domain*. Longman.

Burns, A., Xiong, C., Franconeri, S., Cairo, A., & Mahyar, N. (2020). How to evaluate data visualizations across different levels of understanding. In *2020 IEEE workshop on evaluation and beyond-methodological approaches to visualization (BELIV)* (pp. 19–28). IEEE.

Cafaro, F., Lyons, L., & Antle, A. N. (2018). Framed guessability: Improving the discoverability of gestures and body movements for full-body interaction. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, *18*(1), 1–12. https://doi.org/10.1145/3173574.3174167

Castillo-Barrera, F. E., Amador-García, M., Pérez-González, H., & Martínez-Pérez, F. E. (2017). Agile evaluation of the complexity of user stories using the bloom's taxonomy. In *2017 international conference on computational science and computational intelligence (CSCI)* (pp. 1047–1050). CSCI. https://doi.org/10.1109/CSCI.2017.182

Çeliköz, N., Erisen, Y., & Sahin, M. (2019). Cognitive learning theories with emphasis on latent learning, gestalt and information processing theories. In *Journal of Educational and Instructional Studies in the World 9*(3), (pp. 18–33). WJEIS. https://eric.ed.gov/?id=ED598366.

Céspedes-Hernández, D., González-Calleros, J. M., Guerrero-García, J., & Vanderdonckt, J. (2020). A grammar for specifying full-body gestures elicited for abstract tasks. *Journal of Intelligent & Fuzzy Systems*, *39*(2), 2433–2444. https://doi.org/10.3233/JIFS-179903

Chen, X., Zhang, J., Zhao, Y., Chen, Q., Chen, B., Xu, N., Jin, E., Shen, Y., Tian, Y., Shen, M., & Gao, Z. (2025). PATE model: A 30-Year review and analysis of gestural interaction research. *Human Factors*, *68*(1), 00187208251359299. https://doi.org/10.1177/00187208251359299

Cheng, F., Fang, E., & Chen, K. B. (2025). Mental models of gestural interaction for information processing in virtual reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *69*(1), 10711813251357931. https://doi.org/10.1177/10711813251357931

Cheng, Y., Wu, Z., & Xiao, R. (2024). Exploring methods to optimize gesture elicitation studies: A systematic literature review. *IEEE*

Access, *12*(1), 64958–64979. https://doi.org/10.1109/ACCESS.2024.3387269

Cherdieu, M., Palombi, O., Gerber, S., Troccaz, J., & Rochet-Capellan, A. (2017). Make gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. *Frontiers in Psychology*, *8*(1), 1689. https://doi.org/10.3389/fpsyg.2017.01689

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Chui, K. (2011). Conceptual metaphors in gesture. *Cognitive Linguistics*, *22*(3), 437–458. https://doi.org/10.1515/cogl.2011.017

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (1st ed., pp. 483–501). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.029

Cook, S. W., Yip, T. K. Y., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Danielescu, A., & Piorkowski, D. (2022). Iterative design of gestures during elicitation: Understanding the role of increased production. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems CHI 22*, 1–14. ACM. https://doi.org/10.1145/3491102.3501962

Deng, L., Wang, W., & Li, P. (2025). Effects of age and cognitive features on comprehension of healthcare symbols in hospitals in Guangzhou. *Scientific Reports*, *15*(1), Article 36355. https://doi.org/10.1038/s41598-025-20128-0

Dingler, T., Rzayev, R., Shirazi, A. S., & Henze, N. (2018). Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses. In*Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI 18*, 1–12. ACM. https://doi.org/10.1145/3173574.3173993

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*(7), 1049–1062. https://doi.org/10.3758/s13421-014-0425-y

Dong, H., Danesh, A., Figueroa, N., & Saddik, A. E. (2015). An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *IEEE Access*, *3*(1), 543–555. https://doi.org/10.1109/ACCESS.2015.2432679

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*(3), 445–454. https://doi.org/10.1518/hfes.45.3.445.27258

Dudley, J. J., Karlson, A., Todi, K., Benko, H., Longest, M., Wang, R., & Kristensson, P. O. (2024). Efficient mid-air text input correction in virtual reality. In *2024 IEEE international symposium on mixed and augmented reality (ISMAR)* (pp. 893–902). IEEE. https://doi.org/10.1109/ISMAR62088.2024.00105

Eggen, P. (2020). Information processing and human memory. https://doi.org/10.1093/acrefore/9780190264093.013.864

Felberbaum, Y., & Lanir, J. (2018). Better understanding of foot gestures: An elicitation study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,CHI 18*, 1–12. ACM. https://doi.org/10.1145/3173574.3173908

Fugate, J. M. B., Macrine, S. L., & Cipriano, C. (2019). The role of embodied cognition for transforming learning. *International Journal of School & Educational Psychology*, *7*(4), 274–288. https://doi.org/10.1080/21683603.2018.1443856

Galati, A., Weisberg, S. M., Newcombe, N. S., & Avraamides, M. N. (2018). When gestures show us the way: Co-thought gestures selectively facilitate navigation and spatial memory. *Spatial Cognition & Computation*, *18*(1), 1–30. https://doi.org/10.1080/13875868.2017.1332064

Ganapathi, P., & Sorathia, K. (2019). Elicitation study of body gestures for locomotion in HMD-VR interfaces in a sitting-position.

Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games, MIG. *19*, 1–10. https://doi.org/10.1145/3359566.3360059

Goldin-Meadow, S., & Beilock, S. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Han, S. H., Han, Y., Park, K., Lee, S., & Lee, W. (2025). SpatIO: Spatial physical computing toolkit based on extended reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems,CHI 25*, 1–22. ACM. https://doi.org/10.1145/3706598.3713747

Hosseini, M., Mueller, H., & Boll, S. (2024). Controlling the rooms: How people prefer using gestures to control their smart homes. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems,CHI 24*, 1–18. ACM. https://doi.org/10.1145/3613904.3642687

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Huang, H. H., Pfister, H., & Yang, Y. (2023). Is embodied interaction beneficial? A study on navigating network visualizations. *Information Visualization*, *22*(3), 169–185. https://doi.org/10.1177/14738716231157082

Ichsan, M. H. H., Sik-Lanyi, C., & Guzsvinecz, T. (2025). Multi-browser VE: Enhancing internet browsing experience through virtual reality. *Software Impacts*, *23*(1), Article 100733. https://doi.org/10.1016/j.simpa.2024.100733

Jian, Y., & Abu Bakar, J. A. (2024). Comparing cognitive load in learning spatial ability: Immersive learning environment vs. digital learning media. *Discover Sustainability*, *5*(1), 111. https://doi.org/10.1007/s43621-024-00310-6

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: Immersion aids recall. *Virtual Reality*, *23*(1), 1–15. https://doi.org/10.1007/s10055-018-0346-3

Larsson, A., & Stolpe, K. (2023). Hands on programming: Teachers' use of metaphors in gesture and speech make abstract concepts tangible. *International Journal of Technology and Design Education*, *33*(3), 901–919. https://doi.org/10.1007/s10798-022-09755-0

Lee, G., Healey, J., & Manocha, D. (2022). VRDoc: Gaze-based interactions for VR reading experience. In *2022 IEEE international symposium on mixed and augmented reality (ISMAR)* (pp. 787–796). IEEE. https://doi.org/10.1109/ISMAR55827.2022.00097

Levine, L. J., & Edelstein, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion*, *23*(5), 833–875. https://doi.org/10.1080/02699930902738863

Li, J., Wang, Y., Cui, Z., Brooks, J., Yan, Y., Lou, Z., & Li, Y. (2025). Mid-air gestures for proactive olfactory interactions in virtual reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems,CHI 25*, 1–18. ACM. https://doi.org/10.1145/3706598.3713964

Li, W. H. A., Zhu, K., & Fu, H. (2017). Exploring the design space of bezel-initiated gestures for Mobile interaction. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, *9*(1), 16–29. https://doi.org/10.4018/IJMHCI.2017010102

Lisle, L., Davidson, K., Gitre, E. J. K., North, C., & Bowman, D. A. (2021). Sensemaking strategies with immersive space to think. In *2021 IEEE virtual reality and 3D user interfaces (VR)* (pp. 529–537). IEEE. https://doi.org/10.1109/VR50410.2021.00077

Liu, Q., Li, Y., Chen, B., Wu, H., & Liang, H.-N. (2024). User-defined gesture interactions for VR museums: An elicitation study. In *2024 IEEE international symposium on mixed and augmented reality (ISMAR)* (pp. 633–642). IEEE. https://doi.org/10.1109/ISMAR62088.2024.00078

Luo, W., Ellenberg, M. O., Satkowski, M., & Dachselt, R. (2025). Documents in your hands: Exploring interaction techniques for spatial arrangement of augmented reality documents. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems,CHI 25*, 1–22. ACM. https://doi.org/10.1145/3706598.3713518

Luo, Y., Yu, J., Liang, M., Wan, Y., Zhu, K., & Santosa, S. S. (2024). Emotion embodied: Unveiling the expressive potential of single-hand gestures. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3613904.3642255

Markussen, A., Jakobsen, M. R., & Hornbæk, K. (2014). Vulture: A mid-air word-gesture keyboard. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *14*(1), 1073–1082. https://doi.org/10.1145/2556288.2556964

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* (pp. xi, 416). University of Chicago Press.

McNeill, D. (2018). Recurrent gestures: How the mental reflects the social. *Gesture*, *17*(2), 229–244. https://doi.org/10.1075/gest.18012.mcn

Monteiro, P., Gonçalves, G., Coelho, H., Melo, M., & Bessa, M. (2021). Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, *27*(5), 2702–2713. https://doi.org/10.1109/TVCG.2021.3067687

Morris, M. R. (2012). Web on the wall: Insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on interactive tabletops and surfaces* (pp. 95–104). IEEE. https://doi.org/10.1145/2396636.2396651

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding users' preferences for surface gestures. *Proceedings of Graphics Interface 2010*, *10*(1), 261–268. https://doi.org/10.5555/1839214.1839260

Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. (2011). Mid-air pan-and-zoom on wall-sized displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *11*(1), 177–186. https://doi.org/10.1145/1978942.1978969

Nathan, M. J., & Martinez, C. V. J. (2015). Gesture as model enactment: The role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice*, *1*(1), 4–37. https://doi.org/10.1080/23735082.2015.1006758

Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2003). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (pp. 409–420). Springer.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, *25*(4), 903–910. https://doi.org/10.1177/0956797613518351

Oh, U., & Findlater, L. (2013). The challenges and potential of end-user gesture customization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *13*(1), 1129–1138. https://doi.org/10.1145/2470654.2466145

Ormrod, J. E. (2008). *Human learning* (5th ed.). Pearson/Merrill Prentice Hall.

Ortega, F. R., Tarre, K., Kress, M., Williams, A. S., Barreto, A. B., & Rishe, N. D. (2019). Selection and manipulation whole-body gesture elicitation study in virtual reality. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 1723–1728). IEEE. https://doi.org/10.1109/VR.2019.8798105

Pereira, A., Wachs, J. P., Park, K., & Rempel, D. (2015). A user-developed 3-D hand gesture set for human–computer interaction. *Human Factors*, *57*(4), 607–621. https://doi.org/10.1177/0018720814559307

Qamar, S. Z., Kamanathan, A., & Al-Rawahi, N. Z. (2016). Teaching product design in line with Bloom's taxonomy and ABET student outcomes. In *2016 IEEE global engineering education conference (EDUCON)* (pp. 1017–1022). IEEE. https://doi.org/10.1109/EDUCON.2016.7474677

Quinn, N. (2008). Metaphoric gestures and cultural analysis. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture* (pp. 253–257). John

Benjamins Publishing Company. https://doi.org/10.1075/gs.3.15qui/html

Ragan, E. D., Bowman, D. A., & Huber, K. J. (2012). Supporting cognitive processing with spatial information presentations in virtual environments. *Virtual Reality*, *16*(4), 301–314. https://doi.org/10.1007/s10055-012-0211-8

Rante, H., Zainuddin, M. A., Achmad, Z. M., Ramadhanti, K. S., & Arissabarno, C. (2024). Analyzing bloom's taxonomy to interactions in virtual reality application for education. In *2024 international electronics symposium* (pp. 734–739). IES. https://doi.org/10.1109/IES63037.2024.10665778

Rawat, P., & Barrera Machuca, M. D. (2025). Spatial hand actions: Exploring the hand actions used to represent spatial thinking for 3D assembling tasks. In *Proceedings of the 2025 CHI conference on human factors in computing systems. CHI '25*. ACM. https://doi.org/10.1145/3706598.3713424

Rekik, Y., Vatavu, R.-D., & Grisoni, L. (2014). Understanding users' perceived difficulty of multi-touch gesture articulation. In *Proceedings of the 16th international conference on multimodal interaction* (pp. 232–239). ACM. https://doi.org/10.1145/2663204.2663273

Rempel, D., Camilleri, M. J., & Lee, D. L. (2015). The design of hand gestures for human-computer interaction: Lessons from sign language interpreters. *International Journal of Human-Computer Studies*, *72*(10–11), 728–735. https://doi.org/10.1016/j.ijhcs.2014.05.003

Riva, G. (2008). From virtual to real body: Virtual reality as embodied technology. *Journal of CyberTherapy and Rehabilitation*, *1*(1), 7–22.

Roth, W.-M., & Lawless, D. (2002). Scientific investigations, metaphorical gestures, and the emergence of abstract scientific concepts. *Learning and Instruction*, *12*(3), 285–304. https://doi.org/10.1016/S0959-4752(01)00023-8

Silpasuwanchai, C., & Ren, X. (2015). Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies*, *80*(1), 1–13. https://doi.org/10.1016/j.ijhcs.2015.02.010

Singer, M., Radinsky, J., & Goldman, S. R. (2008). The role of gesture in meaning construction. *Discourse Processes*, *45*(4–5), 365–386. https://doi.org/10.1080/01638530802145601

Skarbez, R., Polys, N. F., Ogle, J. T., North, C., & Bowman, D. A. (2019). Immersive analytics: Theory and research agenda. *Frontiers in Robotics and AI*, *6*(1), 82. https://doi.org/10.3389/frobt.2019.00082

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, *3*(1), 6. https://doi.org/10.1186/s41235-018-0092-9

Suffian, M., Khan, M. Y., & Bogliolo, A. (2022). Towards human cognition level-based experiment design for counterfactual explanations. In *2022 Mohammad Ali Jinnah University international conference on computing (MAJICC)* (pp. 1–5). MAJICC. https://doi.org/10.1109/MAJICC56935.2022.9994203

Sullivan, J. V. (2018). Learning and embodied cognition: A review and proposal. *Psychology Learning & Teaching*, *17*(2), 128–143. https://doi.org/10.1177/1475725717752550

Sun, L., Zhang, H., Chen, P., Jiang, Z., Xie, X., Zhou, Z., Liu, X., & Chen, X. (2024). Elicitation and evaluation of hand-based interaction language for 3D conceptual design in mixed reality. *International Journal of Human-Computer Studies*, *183*(1), Article 103198. https://doi.org/10.1016/j.ijhcs.2023.103198

Tian, J., Wang, Y., Yu, K., Xu, L., Xie, J., Li, F. M., Niu, Y., & Fan, M. (2024). Designing upper-body gesture interaction with and for people with spinal muscular atrophy in VR. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI 24*, 1–19. https://doi.org/10.1145/3613904.3642884

Toyama, S., Al Sada, M., & Nakajima, T. (2018). VRowser: A virtual reality parallel web browser. *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, 230–244. https://doi.org/10.1007/978-3-319-91581-4_17

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*(1), 1454. https://doi.org/10.3389/fpsyg.2017.01454

Vatavu, R.-D., & Zaiti, I.-A. (2014). Leap gestures for TV: Insights from an elicitation study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, TVK14*, 131–138. ACM. https://doi.org/10.1145/2602299.2602316

Villanueva, I. (2015). An exploration of Bloom's knowledge, skills, and affective-based goals in promoting development of freshmen engineering students' professional identities. In *2015 IEEE frontiers in education conference (FIE)* (pp. 1–5). IEEE. https://doi.org/10.1109/FIE.2015.7344057

Vogiatzidakis, P., & Koutsabasis, P. (2018). Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, *2*(4), 4. https://doi.org/10.3390/mti2040065

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005). Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on human factors in computing systems* (pp. 1869–1872). ACM. https://doi.org/10.1145/1056808.1057043

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1083–1092. ACM. https://doi.org/10.1145/1518701.1518866

Wu, H., Fu, S., Yang, L., & Zhang, X. (2022). Exploring frame-based gesture design for immersive VR shopping environments. *Behaviour & Information Technology*, *41*(1), 96–117. https://doi.org/10.1080/0144929X.2020.1795261

Xia, H., Glueck, M., Annett, M., Wang, M., & Wigdor, D. (2022). Iteratively Designing gesture vocabularies: A survey and analysis of best practices in the HCI literature. *ACM Trans. Comput.-Hum. Interact*, *29*(4), 37:1–37:54. https://doi.org/10.1145/3503537

Yaacoub, A., Da-Rugna, J., & Assaghir, Z. (2025). Assessing AI-Generated questions' alignment with cognitive frameworks in educational assessment (arXiv:2504.14232). arXiv. https://doi.org/10.48550/arXiv.2504.14232

Yang, F., Qian, J., Novotny, J., Badre, D., Jackson, C. D., & Laidlaw, D. H. (2021). A virtual reality memory palace variant aids knowledge retrieval from scholarly articles. *IEEE Transactions on Visualization and Computer Graphics*, *27*(12), 4359–4373. https://doi.org/10.1109/TVCG.2020.3009003

Yang, Y., Dwyer, T., Wybrow, M., Lee, B., Cordeil, M., Billinghurst, M., & Thomas, B. H. (2022). Towards immersive collaborative sensemaking. *Proceedings of the ACM on Human-Computer Interaction*, *6*(ISS), 722–746. https://doi.org/10.1145/3567741

Zaiţi, I.-A., Pentiuc, Ş.-G., & Vatavu, R.-D. (2015). On free-hand TV control: Experimental results on user-elicited gestures with leap motion. *Personal and Ubiquitous Computing*, *19*(5), 821–838. https://doi.org/10.1007/s00779-015-0863-y

Zou, Y., Wan, Y., Zhang, Y., Zhou, X., Cheng, J., Ma, C., Yu, C., Deng, X., & Wang, H. (2025). Gesture builder: Flexible gesture customization and efficient recognition on VR devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *9*(2), 67:1–67:37. https://doi.org/10.1145/3729484

Author Biographies

Fangyuan Cheng is a Ph.D. student in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. She earned her master's degree from Tongji University in 2023 and her bachelor's degree from Hunan University in 2020.

Emily Fang is a Ph.D. student in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. She earned her B.S. in Industrial and Systems Engineering from NC State in 2024.

Karen B. Chen is an Associate Professor in the Fitts Department of Industrial and Systems Engineering at North Carolina State University, Raleigh, NC, U.S.A. She received her PhD in Biomedical Engineering from University of Wisconsin-Madison in 2015.

Appendix A

Referent Design

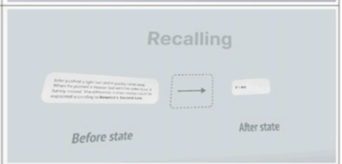
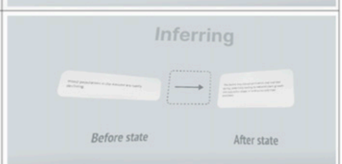
Tasks	Definitions and examples	Description of the before and after states	Screenshots in the immersive environments
Recognizing	**Definition:** Identification of knowledge in long-term memory that is consistent with presented material. **Example:** Identify all the organelles in a passage.	**Before state:** A biology passage describes the functions of several organelles, such as mitochondria generating energy, ribosomes synthesizing proteins. **After state:** All the organelles are identified and highlighted (e.g. mitochondria, ribosomes, nucleus.)	
Recalling	**Definition:** Retrieving relevant knowledge from long-term memory. **Example:** Recall the formula for kinetic energy.	**Before state:** "*John pushed a light ball and it quickly rolled away. When he pushed a heavier ball with the same force, it barely moved. The difference in their motion could be explained according to* ***Newton's Second Law.***" **After state:** F = ma	
Interpreting	**Definition:** Making sense of information by translating it from one representational form (e.g., visual) into another (e.g., verbal, conceptual). **Example:** Interpret the trend shown in a graph.	**Before state:** A graph labeled "Monthly Sales Growth in 2024" showing data points that steadily rise over time. **After state:** "*The chart shows steady sales growth in 2024. From January to March, sales increased gradually. After April, the growth accelerated...*"	
Exemplifying	**Definition:** Finding a specific example or illustration of a concept or principle. **Example:** Give an example of a mammal.	**Before state:** Carnivore **After state:** "*Seals – Seals are apex carnivores in the marine food web, feeding on animals like herring and cod, and helping balance the ecosystem by controlling prey populations.*"	
Classifying	**Definition:** Assigning concepts or objects to categories based on shared attributes or characteristics. **Example:** Categorize the given concepts or objects into their classes.	**Before state:** A set of unorganized items (Apple, Dog, Cat, Banana, Car, Orange). **After state:** Items grouped into meaningful categories: Fruits (Apple, Orange, Banana), Animals: (Dog, Cat), Vehicles (Car).	
Summarizing	**Definition:** Condensing information into a concise representation that captures its central theme or main ideas. **Example:** Summarize the main idea of a passage.	**Before state:** A long, detailed passage describing multiple ecological issues of the Amazon rainforest. **After state:** A concise version capturing the essential points.	
Inferring	**Definition:** Drawing logical conclusions from available information, or predicting possible outcomes based on given facts or evidence. **Example:** Infer the consequence of a phenomenon.	**Before state:** "*Insect populations in the Amazon are declining.*" **After state:** "*This decline may disrupt pollination and nutrient cycling, potentially leading to reduced plant growth and population drops in herbivores and their predators.*"	

Figure A1. Working definitions, examples, before and after states, and screenshots in the immersive environment for each referent

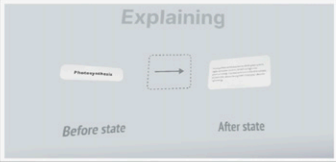
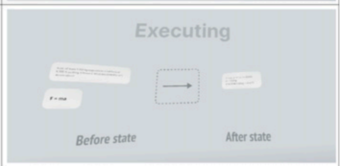
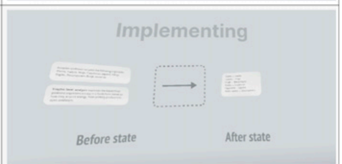
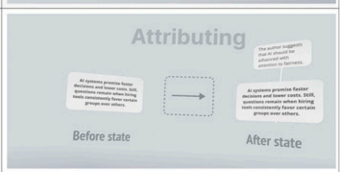
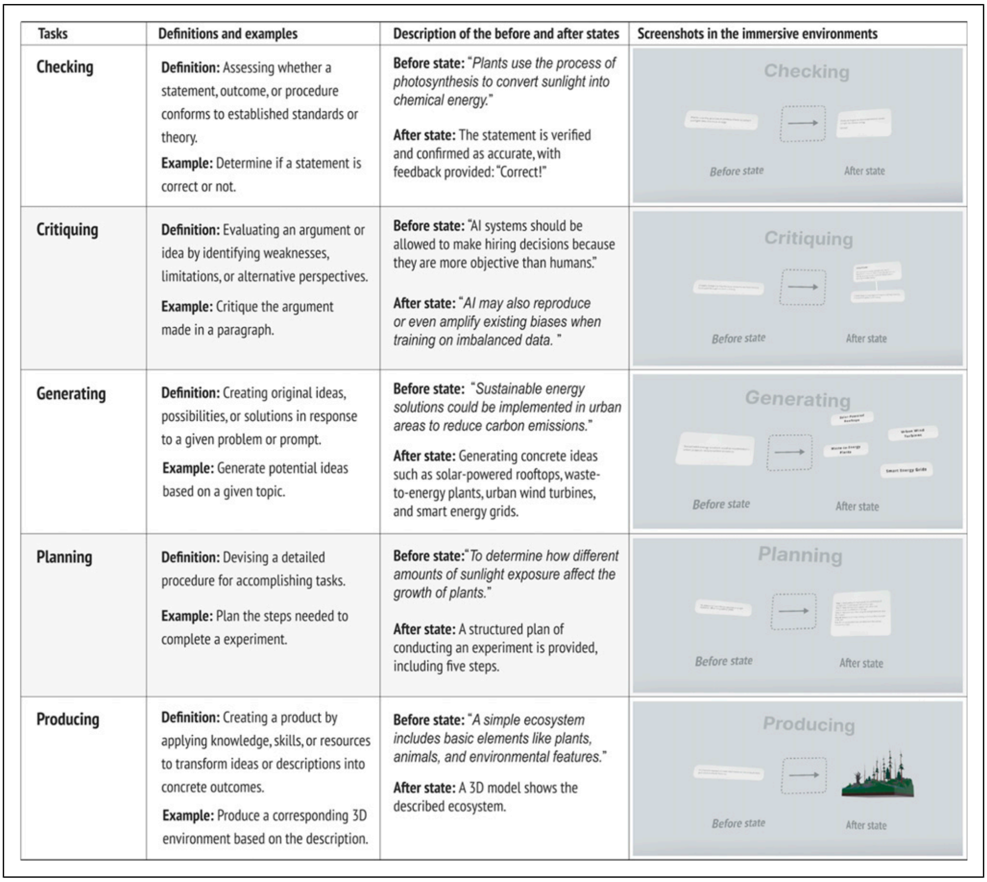
Tasks	Definitions and examples	Description of the before and after states	Screenshots in the immersive environments
Comparing	**Definition:** Identifying similarities and differences between two or more ideas, objects, or pieces of information. **Example:** Compare the similarities and differences between entities.	**Before state:** Descriptions of entities (Jaguar and Harpy Eagle). **After state:** Comparison of similarities and differences between the two.	
Explaining	**Definition:** Articulating the underlying mechanisms, logic or principles of a concept, process, or phenomenon. **Example:** Explain the mechanism of photosynthesis.	**Before state:** A concept: "Photosynthesis" **After state:** "*Photosynthesis is the process by which green plants, algae, and some bacteria convert sunlight into chemical energy...*"	
Executing	**Definition:** Directly applying data, information or procedures to solve problems. **Example:** Solve an equation based on the given data.	**Before state:** A problem statement along with a relevant formula. **After state:** Applies the formula to solve the problem and arrives at the result.	
Implementing	**Definition:** Applying knowledge in a new context to find solutions to new problems. **Example:** Apply trophic level analysis to determine the food web among a set of organisms.	**Before state:** A list of Amazon rainforest organisms is given on the top panel (plants, insects, frogs, capybaras, jaguars...) along with the concept of trophic level analysis in the bottom panel. **After state:** The organisms are organized into a food chain showing energy flow.	
Differentiating	**Definition:** Determine the item in a set of concepts is irrelevant or inconsistent with the others. **Example:** Find the one concept that does not belong in a given set of concepts.	**Before state:** A set of terms is presented together: Star, Planet, Quantum, Galaxy, Asteroid, Comet. **After state:** One item is highlighted: Quantum.	
Organizing	**Definition:** Arranging relevant information into systematic and coherent structures according to rules or principles. **Example:** Organize the given ideas logically.	**Before state:** Several concepts such as Discovery, Exploration, Observation, Patterns, Serendipity, and Insights are presented as separate items. **After state:** The concepts are arranged into an organized structure.	
Attributing	**Definition:** Determine a point of view, a bias, values, or intent underlying given material. **Example:** Identify the point view of the author in an article.	**Before state:** "AI systems promise faster decisions and lower costs. Still, questions remain when hiring tools consistently favor certain groups over others." **After state:** "The author suggests that AI should be advanced with attention to fairness."	

Figure A1. Continued.



Tasks	Definitions and examples	Description of the before and after states	Screenshots in the immersive environments
Checking	**Definition:** Assessing whether a statement, outcome, or procedure conforms to established standards or theory. **Example:** Determine if a statement is correct or not.	**Before state:** *"Plants use the process of photosynthesis to convert sunlight into chemical energy."* **After state:** The statement is verified and confirmed as accurate, with feedback provided: "Correct!"	Checking Before state After state
Critiquing	**Definition:** Evaluating an argument or idea by identifying weaknesses, limitations, or alternative perspectives. **Example:** Critique the argument made in a paragraph.	**Before state:** "AI systems should be allowed to make hiring decisions because they are more objective than humans." **After state:** *"AI may also reproduce or even amplify existing biases when training on imbalanced data."*	Critiquing Before state After state
Generating	**Definition:** Creating original ideas, possibilities, or solutions in response to a given problem or prompt. **Example:** Generate potential ideas based on a given topic.	**Before state:** *"Sustainable energy solutions could be implemented in urban areas to reduce carbon emissions."* **After state:** Generating concrete ideas such as solar-powered rooftops, waste-to-energy plants, urban wind turbines, and smart energy grids.	Generating Before state After state
Planning	**Definition:** Devising a detailed procedure for accomplishing tasks. **Example:** Plan the steps needed to complete a experiment.	**Before state:** *"To determine how different amounts of sunlight exposure affect the growth of plants."* **After state:** A structured plan of conducting an experiment is provided, including five steps.	Planning Before state After state
Producing	**Definition:** Creating a product by applying knowledge, skills, or resources to transform ideas or descriptions into concrete outcomes. **Example:** Produce a corresponding 3D environment based on the description.	**Before state:** *"A simple ecosystem includes basic elements like plants, animals, and environmental features."* **After state:** A 3D model shows the described ecosystem.	Producing Before state After state

Figure A1. Continued.

Appendix B

Demographic Information of Participants.

Table B1. Demographic Information of Participants in the Elicitation Study

ID	Age	Gender	Academic level	Major	VR usage	VR familiarity	Gesture device usage	Gesture familiarity
1	30	Male	PhD	ISE[a]	Yes	Unfamiliar	Yes	Unfamiliar
2	27	Male	PhD	ISE	Yes	Neutral	Yes	Neutral
3	24	Female	PhD	Computer science	No	Very unfamiliar	No	Very unfamiliar
4	26	Male	PhD	Computer science	Yes	Neutral	No	Unfamiliar
5	23	Female	Master's	Economics	Yes	Neutral	No	Unfamiliar
6	23	Female	Master's	Financial math	Yes	Neutral	Yes	Neutral
7	24	Female	PhD	ISE	Yes	Familiar	Yes	Familiar

(continued)

Table B1. (continued)

ID	Age	Gender	Academic level	Major	VR usage	VR familiarity	Gesture device usage	Gesture familiarity
8	22	Female	Junior	Textile engineering	No	Neutral	No	Neutral
9	23	Male	Master's	Computer engineering	No	Neutral	No	Neutral
10	25	Male	PhD	Computer science	Yes	Neutral	Yes	Neutral
11	22	Male	Master's	Textile engineering	Yes	Neutral	No	Unfamiliar
12	30	Male	PhD	ISE	Yes	Familiar	No	Familiar
13	25	Male	PhD	ISE	Yes	Unfamiliar	No	Unfamiliar
14	28	Female	PhD	ISE	Yes	Neutral	Yes	Familiar
15	25	Female	Master's	Physics	No	Unfamiliar	No	Unfamiliar

[a]Abbreviations: ISE, Industrial and Systems Engineering.

Table B2. Demographic Information of Participants in the Validation and Evaluation Study

ID	Age	Gender	Academic level	Major	Immersive tech usage	Immersive tech familiarity	Gesture device usage	Gesture input familiarity
1	18	Female	Sophomore	Biomedical engineering	Yes	Very unfamiliar	No	Very unfamiliar
2	34	Male	PhD	Industrial engineering	No	Unfamiliar	No	Unfamiliar
3	25	Male	PhD	ISE[a]	Yes	Unfamiliar	Yes	Familiar
4	24	Female	Master's	Economics	Yes	Neutral	No	Neutral
5	23	Female	PhD	ISE	Yes	Familiar	No	Unfamiliar
6	27	Male	PhD	ISE	Yes	Neutral	Yes	Neutral
7	28	Male	PhD	Mathematics	No	Neutral	No	Neutral
8	25	Female	Master's	Physics	No	Unfamiliar	No	Unfamiliar
9	23	Female	Sophomore	Computer science	Yes	Very unfamiliar	Yes	Very unfamiliar
10	23	Female	PhD	Computer science	Yes	Neutral	Yes	Neutral
11	25	Male	PhD	ECE[b]	Yes	Neutral	No	Unfamiliar
12	31	Male	PhD	ISE	No	Familiar	No	Neutral
13	26	Male	PhD	ISE	Yes	Unfamiliar	No	Very unfamiliar
14	26	Male	PhD	ISE	No	Unfamiliar	No	Unfamiliar
15	28	Female	PhD	ISE	Yes	Familiar	Yes	Familiar
16	25	Male	PhD	Civil engineering	No	Unfamiliar	No	Neutral
17	27	Male	PhD	Computer science	Yes	Familiar	No	Unfamiliar
18	19	Female	Sophomore	Genetics	Yes	Familiar	No	Unfamiliar

(continued)

Table B2. (continued)

ID	Age	Gender	Academic level	Major	Immersive tech usage	Immersive tech familiarity	Gesture device usage	Gesture input familiarity
19	20	Male	Sophomore	Genetics	Yes	Unfamiliar	No	Unfamiliar
20	26	Female	PhD	Chemical engineering	No	Unfamiliar	No	Very unfamiliar

[a]Abbreviations: ISE, Industrial and Systems Engineering.
[b]Abbreviations: ECE, Electrical and Computer Engineering.

Appendix C

Protocol for the Semi-structured Interview in the Validation and Evaluation Study

1. Were there any gestures that were particularly natural to you? Can you explain why?
2. Were there any gestures you found difficult to understand? Can you explain why?
3. Did you find any gestures to be easily confused with one another? If so, which ones, and in what ways?
4. Did any gestures feel physically tiring or uncomfortable to perform?
5. Were there any gestures that you felt required more effort than others?
6. Were there any gestures you would change or redesign? If so, how and why?
7. Are there any new gestures you wish to add to the current set? If so, what are they, and what would they be used for?
8. Do you have any additional comments or suggestions regarding the gesture set, the task, or the overall experience?