Proceedings of the Human Factors and Ergonomics Society Annual Meeting
1–7

DOI: 10.1177/10711813251357931
journals.sagepub.com/home/pro
Sage

Mental Models of Gestural Interaction for Information Processing in Virtual Reality

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

Abstract

Virtual reality (VR) provides an immersive medium for information processing. Gestures offer a natural and intuitive means of translating cognitive processes into physical actions, making them a promising interaction method to support information processing in VR. However, how gestures reflect and map abstract cognitive tasks to physical interactions remains underexplored. To address this gap, this study explored the mental models by which end-users design gestural interactions to support information processing in VR. Using a gesture elicitation method, 8 participants created 445 gestures representing 19 cognitive processes in Bloom's taxonomy. Five categories of mental models were identified: linguistic-symbolic, spatial-manipulative, metaphoric, social-conventional, and traditional graphical user interface-derived. Furthermore, the study found that the cognitive process categories affected the mental model adoption, with higher-order cognitive processes prompting more human-like interaction. These findings suggest the potential for developing more natural interactions for cognitive tasks and offer guidance for designing gestural interactions in VR.

Keywords

mental models, gesture, virtual reality, information processing

Introduction

Virtual reality (VR) has demonstrated potential in supporting information processing by facilitating the visualization and synthesis of information in tasks such as visual analytics (Moran et al., 2015) and sense-making (Lisle et al., 2021). Gestures are a common interaction modality in VR, complementing controllers and other haptic input devices. Gestural interaction offers an intuitive means of expressing commands through movements aligned with mental concepts (Vuletic et al., 2019; Wang et al., 2023). Prior research has demonstrated that gestures can facilitate embodied cognition by appropriately mapping cognitive tasks to physical actions (Alibali et al., 2014; Dijkstra & Post, 2015; Skulmowski & Rey, 2018). Given these characteristics, gestural interaction is considered a suitable interaction method to translate cognitive tasks, such as information processing, to physical actions in VR. While previous research has explored various gestural interactions in VR, it has primarily focused on daily tasks (Pereira et al., 2015) or specific applications such as design (Vuletic et al., 2021) or immersive shopping (Wu, Wang, et al., 2019). How users map abstract cognitive processes in information processing tasks to physical gestural interactions remains underexplored. This study aimed to address this gap by exploring the mental models influencing gesture design for information processing tasks in VR.

Human information processing refers to the sequence of cognitive operations that individuals use to perceive, encode, store, retrieve, and manipulate information (J. R. Anderson, 1990). Further advancing the understanding of cognitive activities in information processing, Bloom's taxonomy (Bloom et al., 1956) provides a practical framework for cognitive processes involved in learning. The revised Bloom's taxonomy (L. W. Anderson & Krathwohl, 2001) classified 19 cognitive activities into a 6-level hierarchy, ranging from lower-order processes of remembering and understanding to higher-order processes of applying, analyzing, evaluating, and creating (Lutz & Huitt, 2003). In instructional design, this taxonomy provides practical guidance for aligning learning objectives with teaching materials. For example, to target the "analyzing" level, instructional materials might include case studies or data interpretation exercises that prompt students to break down information, identify patterns, and draw inferences. Although originally developed for educational purposes, it has provided a structured framework that aids in

[1]North Carolina State University, Raleigh, USA

Corresponding Author:
Fangyuan Cheng, North Carolina State University, 4101 Fitts-Woolard Hall, Raleigh, NC 27695-7001, USA.
Email: fcheng5@ncsu.edu

various information processing domains, such as artificial intelligence (AI) model comprehension ability training (Sahu et al., 2021) and data visualization evaluation (Burns et al., 2020).

Gestures are considered a natural mode of interaction with the surroundings, typically involving movements of the fingers, hands, arms, and upper body (Vuletic et al., 2019). Gestures and cognitive processes share an intrinsic connection, as evidenced by research in embodied cognition that human cognitive processes are fundamentally linked to physical movements (Varela et al., 2017). In other words, information processing is not merely an abstract process confined to the brain but is influenced by bodily interactions with the environment. These findings support the rationale and potential added value of using gestural interaction for information processing. Virtual reality further facilitates this by enabling hand-based interaction within a 3D environment. To explore the design of gestural interaction in VR, previous studies have widely used the gesture elicitation method, which was initially applied to surface-based digital systems like touchscreens (Morris et al., 2010; Wobbrock et al., 2009) and later adapted to immersive environments (Wu, Luo, et al., 2019). Gesture elicitation methods involve collecting spontaneous, user-generated gestures in response to predefined tasks or functions, then identifying those that are highly guessable and widely accepted for the task (Vatavu & Wobbrock, 2015). They effectively capture users' mental models because they naturally reflect how users conceptualize interactions based on their experiences, background, and embodied cognition (Hostetter & Alibali, 2008). This preliminary study adopted the cognitive processes outlined in the revised Bloom's taxonomy as representative information processing tasks in VR to investigate users' mental models for translating these tasks into gestural interactions using elicitation methods. The study was guided by the following research questions (RQs): **RQ1**. What types of mental models do users adopt when designing gestural interaction to perform different cognitive processes in virtual reality? **RQ2**. What are the effects of different cognitive processes on the adoption of mental model types in gestural interaction design?

Methods

Participants

Eight participants (5 females, 3 males) aged 22 to 30 years (M=24.88, SD=2.64) were enrolled in this preliminary study with informed consent, which was approved by the North Carolina State University Institutional Review Board. Their academic backgrounds include Computer Science (n=2), Industrial and Systems Engineering (n=2), Economics (n=1), Financial Mathematics (n=1), and Textile Engineering (n=1). Six participants have used VR devices, with familiarity rated as neutral (n=5), familiar (n=1), unfamiliar (n=1), and very unfamiliar (n=1). Four participants had prior experience with gesture-based input, and familiarity was rated as neutral (n=3), unfamiliar (n=3), familiar (n=1), and very unfamiliar (n=1).

Procedure and Task

After being introduced to the study, participants completed the gesture elicitation task in a virtual environment, which was created and rendered using Unity Engine (2018.4.28f1, https://unity3d.com/) and delivered via a head-mounted display (HMD; Oculus Quest 3, Meta). The environment presented referents that depicted the "before" and "after" states for each of the 19 cognitive processes listed in the revised Bloom's taxonomy, thereby helping participants contextualize the task. Referents were primarily text-based and static, consisting of information panels that visually represented changes in content or structure. For example, in the cognitive process of "summarizing," the before state was an information panel displaying a detailed paragraph, while the after state was another panel with a concise outline of its content (Figure 1). Participants were asked to design at least three appropriate and distinct gestures to bridge the transition between the "before" and "after" states, and then explain the rationale behind their designs. To facilitate data collection, participants were asked to perform each gesture three times. The presentation order of the cognitive processes was counterbalanced to control order effects. Participants' movements were video recorded, and any verbal expressions were audio recorded. Each designed gesture was documented in real time by researchers through written notes.

Data Analysis

Video and audio recording and notes were used for inductive thematic coding of participants' mental models during gesture design by two researchers independently. The form of the gestures, movement paths, and participants' explanations served as the reference materials. Any discrepancies were resolved through discussion. The resulting coding categories of mental models were then used as dependent variables, while the six categories of cognitive processes from Bloom's taxonomy were used as independent variables. Instead of a traditional null hypothesis significance testing, a Bayesian generalized linear mixed model (GLMM) was employed. In brief, a Bayesian GLMM uses Bayesian methods (i.e., updates prior beliefs with new data or posterior beliefs in the form of probabilities) and extends linear regression to account for the multinomial nature of variables, repeated measures, and random effects of non-normal data. Weakly informative priors provided by the *brms* package in R were adopted since there was not a lot of strong prior evidence in the literature for this type of study (Bürkner, 2017). The model was estimated using Markov chain Monte Carlo sampling with four chains (4,000 iterations per chain, 2,000 warm-up iterations; Bürkner, 2017). All parameters demonstrated sufficient

Figure 1. Virtual environment and the referent for the task.

convergence ($\hat{R}$ =1.00), and effective sample sizes exceeded 2,000 for all estimates, indicating high sampling efficiency and stable posterior inference. Contrast estimates (Est.) of the differences in predicted log-odds of adopting specific mental models between cognitive process categories were computed, along with standard errors (*SE*) and 95% credible intervals (CI). The posterior probability of direction (*pd*) was also reported, where a *pd* of .99 indicates 99% certainty in the effect's direction.

Results

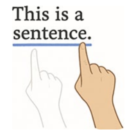
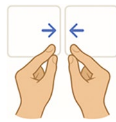
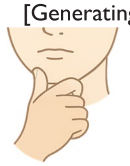
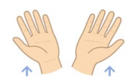
The elicitation task yielded 445 gesture designs for all 19 cognitive processes combined. Five distinct types of user mental models in gestures design were identified: linguistic-symbolic (n=154), spatial-manipulative (n=136), metaphoric (n=55), social-conventional (n=23), and graphical user interface (GUI)-derived (n=21) mental model (Table 1).

The linguistic-symbolic mental model is based on language, writing conventions, and symbolic notation, where users conceptualize gestures by referencing familiar written symbols and characters. For example, participants traced letters in the air to represent corresponding cognitive functions, such as drawing a letter "P" to indicate "planning," or drew checkmarks and underlines to signify "checking" and "recognizing." The spatial-manipulative mental model is grounded in the repositioning or transformation of virtual information panels. This was evident in gestures such as pinching two virtual panels together to merge them or flipping a panel to access information on its reverse side.

The third category relies on metaphorical associations, where gestures map cognitive processes onto physical objects, processes, or phenomena. For example, users simulated a "magic wand" by twirling their index finger in the air to represent "generating," or enacted a blooming flower to symbolize "producing." The social-conventional mental model is informed by socially recognized gestures and everyday bodily actions, as well as culturally conventional movements used in communication. Examples include placing a hand on the chin to indicate "thinking of a plan" for "planning" or spreading open the palms to express "what's next?" for "inferring." The GUI-derived mental model refers to interaction habits from traditional GUI-based computing devices, such as computers or tablets, where users transfer familiar input methods into VR. For example, participants performed gestures such as double-tapping in the air to simulate a mouse click for "recognizing" or long-pressing for "explaining."

A GLMM was fitted to examine the effects of cognitive process categories on mental model categories. The Bayes Factor (*BF*) analysis ($log_{10}(BF)>20$) revealed that the full model, compared to the reduced model that excluded cognitive processes as predictor variables, significantly better explained the data. A $log_{10}(BF)>2$ is considered decisive evidence (Kass & Raftery, 1995). The predicted probabilities of adopting mental models across cognitive process

Table 1. Five types of mental models and representative examples.

Types	Representative examples	
Linguistic-symbolic	 drawing a letter "P" [Planning]	 drawing underlines on the contents [Recognizing]
Spatial-manipulative	 pinching two panels together to merge them [Executing]	 flipping a panel to access information on its reverse side [Checking]
Metaphoric	 simulating a "magic wand" using the index finger [Generating]	 enacting an opening flower [Producing]
Social-conventional	 placing a hand on the chin [Planning]	 open the palms to express "what's next?" [Inferring]
GUI-derived	 double-tapping to simulate mouse clicks [Recognizing]	 long-pressing [Explaining]

Note. The labels in square brackets indicate the cognitive process represented by each gesture.

categories are shown in Figure 2, which illustrates that the full model, compared to the reduced model, captures more variations across cognitive processes. These results revealed the notable effect of cognitive process categories on the mental model adoption.

Figure 3 shows the proportional distribution of five types of mental models within each cognitive process category. The results indicated that the **Evaluate** category increased the likelihood of employing the linguistic-symbolic mental model than Remember (Est. = 3.07, *SE* = 1.29, CI [0.89, 5.97], *pd* = .998) and Understand (Est. = 2.28, *SE* = 1.25, CI [0.23, 5.17], *pd* = .986), and increased the likelihood of the social-conventional mental model than Remember (Est. = 2.47, *SE* = 1.41, CI [0.02, 5.6], *pd* = .978), Understand (Est. = 3.88, *SE* = 1.55, CI [1.16, 7.31], *pd* = .998), Apply (Est. = 3.09, *SE* = 1.96, CI [0.37, 7.34], *pd* = .96), and Analyze (Est. = 3.29, *SE* = 1.59, CI [0.5, 6.87], *pd* = .989). The **Create** category increased the likelihood of the metaphoric mental model than Remember (Est. = 3.1, *SE* = 1.32, CI [0.89, 6.12], *pd* = .998), Understand (Est. = 2.61, *SE* = 1.28, CI [0.51, 5.58], *pd* = .995), Apply (Est. = 4.36, *SE* = 1.93, CI [1.03, 8.65], *pd* = .995), and Analyze (Est. = 3.24, *SE* = 1.39, CI [0.86, 6.32], *pd* = .997). It increased the likelihood of the social-conventional mental model than Remember (Est. = 2.52, *SE* = 1.4, CI [0.13, 5.65], *pd* = .976), Understand (Est. = 3.94, *SE* = 1.53, CI [1.3, 7.39], *pd* = .995), and Analyze (Est. = 3.33, *SE* = 1.59, CI [0.55, 6.79], *pd* = .991). It also increased the likelihood of the linguistic-symbolic mental model over the Remember (Est. = 2.38, *SE* = 1.32, CI [0.19, 5.37], *pd* = .985). The **Remember** category decreased the likelihood of the spatial-manipulative mental model than Understand (Est. = −2.51, *SE* = 0.75, CI [−3.98, −1.07], *pd* = .999), Apply (Est. = −2.71, *SE* = 1.03, CI [−4.92, −0.85], *pd* = .998), Analyze (Est. = −1.97, *SE* = 0.85, CI [−3.68, −0.34], *pd* = .990), Evaluate (Est. = −2.49, *SE* = 1.37, CI [−5.5, −0.1], *pd* = .978), and Create (Est. = −3.34, *SE* = 1.37, CI [−1.05, −6.43], *pd* = .998).

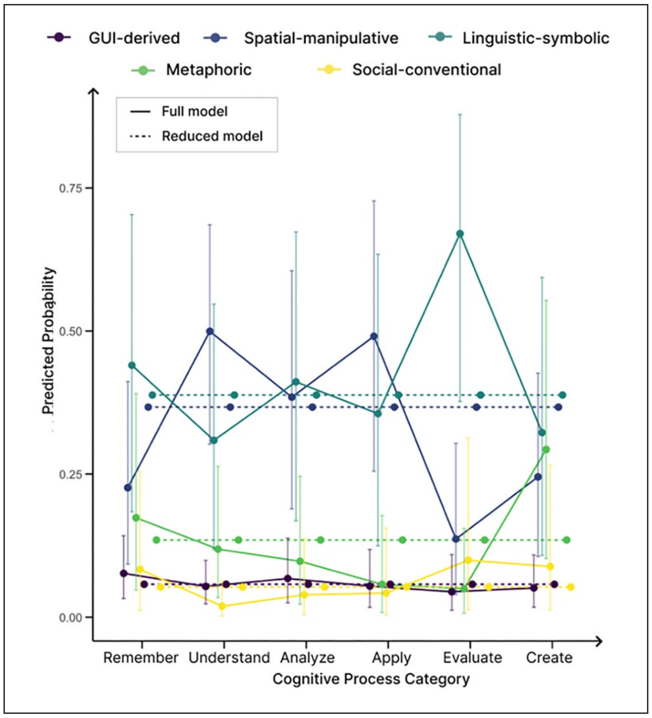


Figure 2. Predicted probabilities of adopting five types of mental models across six cognitive process categories.
Note. Error bars indicate 95% credible intervals for each predicted probability.

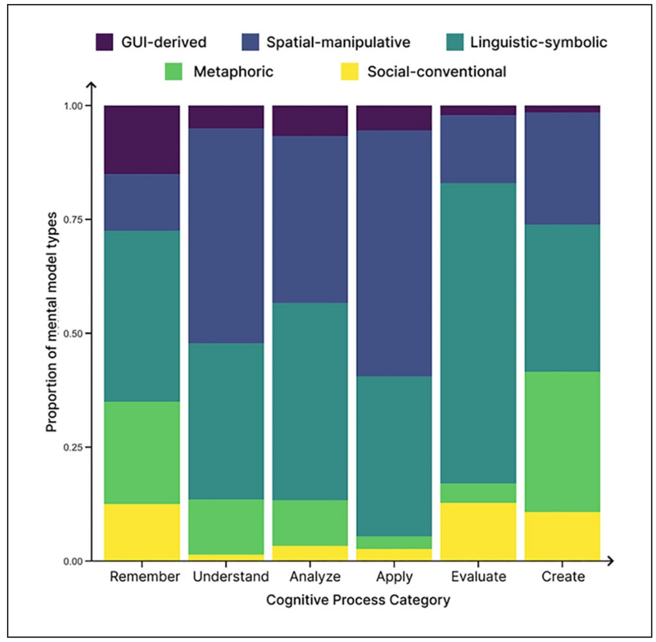


Figure 3. Distribution of five mental model types across cognitive process categories.

Discussion

To address RQ1, this study identified five types of mental models adopted in gesture designs for cognitive processing in VR: linguistic-symbolic, spatial-manipulative, metaphoric, social-conventional, and GUI-derived mental models. Among these, the linguistic-symbolic mental model was the most prevalent, which may be attributed to their intuitive alignment with verbal or conceptual representations of cognitive tasks. Specifically, many cognitive processes in learning and problem-solving, such as summarizing, classifying, or explaining, inherently involve the manipulation of abstract concepts and linguistic structures. As a result, participants may have naturally gravitated toward linguistic or symbolic gestures, such as writing abbreviations or using shorthand symbols like checkmarks, that correspond to familiar verbal or textual conventions. Furthermore, the findings revealed that VR affordances tend to elicit distinct mental models, particularly spatial-manipulative, metaphoric, and social-conventional types, which are less commonly associated with traditional interfaces. Unlike conventional media such as computers and tablets, VR offers a 3D interaction space that allows users to interact with digital content through embodied and spatial actions. The spatial-manipulative mental models leverage this affordance by enabling users to directly manipulate virtual information through gestures such as grabbing or flipping. The emergence of metaphoric mental models further highlights how VR enhances users' ability to integrate real-world interaction patterns into virtual experiences. It enables users to draw on familiar physical-world analogies and naturally enact spatial metaphors, thereby intuitively mapping real-world experiences to virtual actions. For example, opening up a closed fist, reminiscent of a flower blooming, can metaphorically represent meanings such as "expand" or "create." Moreover, social-conventional mental models illustrate VR's potential to support human-computer interactions that more closely resemble interpersonal communication. In these cases, gestures often reflect socially understood conventions to facilitate and bridge social interaction within virtual environments. Under such circumstances, the VR system may no longer be perceived merely as a machine, but rather as an entity capable of engaging in human-like exchanges.

In contrast, the GUI-derived mental models were the least frequently employed. This pattern suggests that although users may still be influenced by traditional interaction paradigms shaped by screen-based media, such influence was relatively limited. Instead, participants tended to explore new forms of interaction that are better suited to support cognitive tasks in immersive VR. Collectively, these findings underscore VR's potential to support a broader and more natural range of interactions that go beyond the abstract operations typical of traditional graphical user interfaces in cognitive tasks.

To address RQ2, this study revealed a notable effect of cognitive process categories on mental model adoption during gesture design. It also compared the predictive probabilities of gesture categories across different cognitive processes. The observed differences may be attributed to the distinct cognitive functions associated with each gesture category.

Specifically, the Evaluate category could prompt participants to relate experiences of assessing or being assessed using paper-based materials, such as exam answer sheets, thereby naturally evoking written symbols like checkmarks or question marks. This, in turn, increased the occurrence of linguistic-symbolic mental models. The Create category involves generating or producing novel information or models, prompting participants to use gestures to represent or metaphorically depict objects being created or the process of creation, thereby increasing the prevalence of metaphoric mental models. The Remember category primarily focuses on retrieving information that is not immediately available in the current context, rather than organizing existing information. This led to fewer gestures related to information manipulation. Instead, participants may adopt alternative mental models, such as the linguistic-symbolic model, where underlying words represent the need to recall more specific details, or the metaphoric model, where pointing to a clenched fist symbolizes the brain that reflects an attempt to seek the desired information with the help of the VR system. Furthermore, social-conventional mental models were more frequently employed in both the Create and Evaluate categories. This suggests that in higher-order cognitive processes, participants are more likely to perceive information processing support as a social entity rather than a passive tool, leading them to use socially conventional gestures for more interactive, human-like communication.

Limitations were acknowledged in this study. Despite the 445 generated gestures serving as an analytical starting point, the small sample size may affect classification and the observed effects of cognitive processes on mental models. The individual differences related to cultural background and educational level may influence gesture design and mental model adoption. The coding approach may introduce researcher bias, and alternative interpretations are possible. Future research could further explore gesture sets corresponding to specific cognitive tasks and refine gesture characteristics at a finer granularity.

Conclusion

To explore users' mental models for translating information processing tasks into gestural interaction in VR, this study conducted an elicitation study with eight participants. Five types of mental models were identified for designing gestural interactions across nineteen cognitive processes. The results also demonstrated that cognitive process categories affected the adoption of these mental models. These findings can assist researchers and developers in strategically designing gestural interaction for cognitive tasks in VR.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

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

References

Alibali, M. W., Boncoddo, R., & Hostetter, A. B. (2014). Gesture in reasoning: An embodied perspective. In L. Shapiro & S. Spaulding (Eds.), *The Routledge handbook of embodied cognition* (pp. 150–159). Routledge.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press. https://doi.org/10.4324/9780203771730

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

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. Handbook 1: Cognitive domain* (pp. 1103–1133). Longman.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Burns, A., Xiong, C., Franconeri, S., Cairo, A., & Mahyar, N. (2020). How to evaluate data visualizations across different levels of understanding. *arXiv*. https://doi.org/10.48550/arXiv.2009.01747

Dijkstra, K., & Post, L. (2015). Mechanisms of embodiment. *Frontiers in Psychology*, *6*, Article 1525. https://doi.org/10.3389/fpsyg.2015.01525

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

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.https://doi.org/10.1080/01621459.1995.10476572

Lisle, L., Davidson, K., Gitre, E. J. K., North, C., & Bowman, D. A. (2021). *Sensemaking strategies with immersive space to think* [Conference session]. IEEE Virtual Reality and 3D User Interfaces (VR), Lisboa, Portugal (pp. 529–537). IEEE. https://doi.org/10.1109/VR50410.2021.00077

Lutz, S. T., & Huitt, W. G. (2003). Information processing and memory: Theory and applications. *Educational Psychology Interactive*, *1*, 17.

Moran, A., Gadepally, V., Hubbell, M., & Kepner, J. (2015). *Improving big data visual analytics with interactive virtual reality* [Conference session]. IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA (pp. 1–6). IEEE. https://doi.org/10.1109/HPEC.2015.7322473

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). *Understanding users' preferences for surface gestures* [Conference session]. Proceedings of Graphics Interface 2010, Ottawa, ON (pp. 261–268). Canadian Information Processing Society.

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

Sahu, P., Cogswell, M., Rutherford-Quach, S., & Divakaran, A. (2021). Comprehension based question answering using bloom's taxonomy. *arXiv*. https://doi.org/10.48550/arXiv.2106.04653

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

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind: Cognitive science and human experience* (revised ed.). MIT Press.

Vatavu, R. D., & Wobbrock, J. O. (2015). Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1325–1334).

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, *129*, 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011

Vuletic, T., Duffy, A., McTeague, C., Hay, L., Brisco, R., Campbell, G., & Grealy, M. (2021). A novel user-based gesture vocabulary for conceptual design. *International Journal of Human-Computer Studies*, *150*, Article 102609. https://doi.org/10.1016/j.ijhcs.2021.102609

Wang, Y., Wang, Y., Li, X., Zhao, C., Ma, N., & Guo, Z. (2023). A comparative study of the typing performance of two mid-air text input methods in virtual environments. *Sensors*, *23*(15), Article 6988.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). *User-defined gestures for surface computing* [Conference session]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA (pp. 1083–1092). ACM Press.

Wu, H., Luo, W., Pan, N., Nan, S., Deng, Y., Fu, S., & Yang, L. (2019). Understanding freehand gestures: A study of freehand gestural interaction for immersive VR shopping applications. *Human-Centric Computing and Information Sciences*, *9*(1), Article 43. https://doi.org/10.1186/s13673-019-0204-7

Wu, H., Wang, Y., Qiu, J., Liu, J., & Zhang, X. (2019). User-defined gesture interaction for immersive VR shopping applications. *Behaviour & Information Technology*, *38*(7), 726–741. https://doi.org/10.1080/0144929X.2018.1552313