RESEARCH | Article in Press

Engage or observe? Eye-tracking and experimental evidence on integrating VR generative practice into short instructional videos

Xiao Liu · Yun Zhou · Bingyu Dong · Hui Zhang · Lingzhi Cai · Bin Li · Tao Xu

Received: 11 November 2025 / Revised: 21 August 2026 / Accepted: 2 September 2026
© The Author(s) 2026

Abstract
Instructional videos are widely used, yet their limited interactivity often leads to passive viewing. Embedding Virtual Reality (VR) as a form of generative practice, in which learners actively create content such as drawings or visual representations, offers a promising solution to foster deeper cognitive processing. This study examines the impact of integrating generative VR practice into instructional videos through an experimental design involving 100 university students. Learners engaged with vocabulary tasks under four spatial layouts that varied in the position (left vs. right) and size (small vs. large) of the interactive area, compared with a video-only control. Outcomes included retention, transfer, attention, cognitive load, and perceived enjoyment. Transfer benefits over video-only learning were observed only in the two large-area VR conditions. Among the VR conditions, right-side positioning increased attention allocation and reduced cognitive load, whereas larger interaction areas improved retention, transfer, and usability while reducing cognitive load. Eye-tracking further showed that VR-based generative practice redirected attention toward the interactive area. No significant interaction between position and size was found. These findings extend generative learning theory into VR-supported video contexts, propose the Attentional Alignment Principle and Workspace Expansion Principle, and provide design implications for interactive instructional videos.

Keywords Generative practice · VR-based interaction · Spatial affordances · Cognitive load · Eye tracking · Instructional videos

Introduction

Interaction is widely recognized as a critical factor for deeper understanding, sustained engagement, and effective knowledge application, making it an increasingly central concern in contemporary instructional practice (Rahimi & Sevilla-Pavón, 2025). Accordingly, instructional design has shifted toward embedding interactivity to encourage learners to actively explore and engage with content and environments (Deng and Gao 2024; Hung and Chen 2018; Yu et al. 2024; Zhang et al. 2026a, b). In instructional videos, such efforts have included learner navigation, embedded questions, and cues, which can improve engagement and comprehension (Deng & Gao, 2024; Geri et

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Published online: 11 September 2026

Springer Open

al., 2017; Hung & Chen, 2018; Kim et al., 2015; Yu et al., 2024; Yürüm et al., 2023). However, these forms of interactivity rarely require learners to generate new content. Generative learning theory, focusing on active strategies, highlights the importance of learners creating new content to organize and integrate knowledge, thereby deepening comprehension (Fiorella & Mayer, 2021).

Despite their wide adoption (C.-M. Chen et al. 2022a, b; Deng and Gao 2024; Hsieh 2020; Shen et al. 2023; Zhang et al. 2026a, b; Zheng et al. 2024), instructional videos often constrain learners to passive viewing, limiting opportunities for interaction that supports memory consolidation and skill transfer (Boateng et al., 2016; Marchionini, 2003; Sablić et al., 2021; Shih, 2010). Generative practice, where learners actively construct representations to integrate knowledge (Fiorella & Mayer, 2021), may enable deeper learning beyond passive viewing or lightly interactive video activities.

Virtual Reality (VR) offers a promising solution by enabling active participation. Its capacity for embodied and multisensory enactment provides potential for supporting generative practice (Karnchanapayap, 2023; Van Gemert et al., 2023). Prior studies have demonstrated positive effects of VR on engagement and learning outcomes (Al Mamun & Lawrie, 2023; Lee et al., 2024; Lin et al., 2024; Zhao & McClure, 2024). Desktop VR, a non-immersive and cost-effective format, has been shown to enhance learning without specialized equipment (Chen et al. 2022a, b; Cromley et al. 2023; De Lorenzis et al. 2024; Johnson-Glenberg et al. 2021; Makransky and Lilleholt 2018; Makransky et al. 2019; Zhou et al. 2018). Yet most VR applications emphasize procedural demonstration or conceptual visualization (Makransky et al., 2019; Yang et al., 2024), with limited attention to generative learning through content creation. Embedding generative practice in desktop VR may therefore extend instructional videos from primarily presenting information toward supporting learners in actively constructing and applying knowledge.

Generative practice via VR may transform videos into powerful tools for deeper learning, but two issues require study: its effects on retention, transfer, attention, and load, and how the spatial affordances of interactive elements shape learning through embodiment (Jiang & Chen, 2022; Kincaid & Corsun, 2003). Multitasking between video viewing and practice may divide learners' attention and momentarily shift focus between competing tasks (Ralph et al., 2015), making it vital to understand how such dynamics affect cognitive efficiency and performance. Despite its promise, clear design guidance remains scarce.

The present study addresses these gaps by investigating the effects of integrating VR-based generative practice into instructional videos and examining the spatial design of the VR practice area. The study first compares VR-based generative practice with video-only learning in terms of retention, transfer, attention, and cognitive load. It then examines how the position (left vs. right) and size (small vs. large) of the practice area influence learning outcomes, attention, cognitive load, perceived enjoyment, and layout usability, as well as whether position and size interact in shaping these outcomes. Accordingly, the following research questions are proposed:

RQ1: How does integrating VR-based generative practice into instructional videos, compared to video-only learning, affect learners' performance, attention, and cognitive load?

RQ2: How does the position of the VR-based generative practice area influence retention, transfer, attention, cognitive load, perceived enjoyment, and layout usability?

RQ3: How does the size of the VR-based generative practice area influence retention, transfer, attention, cognitive load, perceived enjoyment, and layout usability?

RQ4: Do the position and size interact in their effects on learning and experience?

The corresponding hypotheses are developed in Sect. 2 based on generative learning theory, embodied cognition theory, the split-attention effect, and the cognitive-affective theory of learning with media.

Literature review, theoretical background, and hypotheses

Generative learning theory and strategies

Generative learning theory suggests that learning is enhanced when learners actively select, organize, and integrate new information with prior knowledge rather than merely receive it (Fiorella & Mayer, 2016, 2021). Generative activities such as drawing, explaining, visualizing, and enacting can support deeper processing (Fiorella, 2023; Fiorella & Mayer, 2016).

Recent empirical studies provide converging evidence for the benefits of generative strategies in technology-mediated learning (Zhang et al. 2025a, b). In video-based learning, drawing during viewing enhanced the benefits of segmented instruction, and drawing accuracy mediated learning gains (Zhang et al. 2025a, b). Scaffolded self-explanation prompts also improved learning performance and reduced cognitive load (Wang & Xu, 2024). Beyond video contexts, explaining to oneself or others improved retention, and explaining to others further enhanced transfer and activated brain networks related to attention, working memory, metacognitive processing, and social processing (Zhu et al., 2024). In online learning, interactive graphic organizers improved retention and comprehension, directed attention toward generative representations, and increased satisfaction while reducing perceived difficulty (Wang et al., 2021). Collectively, this evidence supports expecting VR-based generative practice to improve retention (H1a), transfer (H1b), and attention (H1c) through deeper knowledge construction (Chi & Wylie, 2014; Fiorella, 2023; Fiorella & Mayer, 2016).

However, their effectiveness depends on activity design, instructional support, and cognitive demands (Skulmowski, 2024). Although some scaffolded activities reduce cognitive load (Wang & Xu, 2024), active production may compete with content processing for working-memory resources. In video learning, such tasks may divide attention between processing and production (Beege and Ploetzner 2025; Zhang et al. 2025a, b). Thus, coordinating video processing with VR-based generative practice is expected to impose greater cognitive load than video viewing alone (H1d).

Taken together, these theoretical and empirical findings lead to the following hypothesis: H1: Compared with video-only learning, VR-based generative practice is expected to improve retention (H1a) and transfer (H1b), increase attention allocation (H1c), and result in higher cognitive load (H1d).

Embodied cognition theory and interface layout in multiscreen and interactive learning

Embodied cognition theory suggests that cognition is grounded in interactions among sensorimotor systems, bodily states, and the environment (Foglia & Wilson, 2013; Stalheim & Somby, 2024; Wilson, 2002). From this perspective, interface layout may therefore affect learning by altering perception-action demands and coordination effort.

Evidence on spatial positioning is mixed. Holmes et al. (2012) found more gaze on the main than the secondary screen. However, Vatavu and Mancas (2015) reported no left-right attention difference. Murano and Lomas (2015) observed fewer errors and clicks with top and left menus, though McCarthy et al. (2004) and Kalbach and Bosenick (2003) found no significant performance differences across menu positions. Thus, existing interface research does not provide direct evidence that right-side placement is universally superior. The prediction in the present study is derived from embodied alignment: perception-action coupling is influenced by factors such as handedness, reachability, and habitual motor routines (Bridgeman & Tseng, 2011; Ranzini et al., 2022; Stalheim & Somby, 2024; Thorvald et al., 2020; Wilson, 2002). Related evidence indicates performance asymmetries associated with hand preference and experience in mouse control (Peters & Ivanoff, 1999), and faster responses when stimulus location is spatially compatible with the responding hand (Wilf et al., 2013). Although this relationship has not been directly tested in VR-based generative learning, for predominantly right-handed learners, positioning the practice area on the right may better align with habitual motor routines, potentially reducing interaction

effort and leaving more cognitive resources available for learning. This reasoning supports the expectation that right-side positioning may facilitate retention (H2a) and transfer (H2b).

Size has a more direct motor-control basis. Larger targets reduce precision demands and facilitate selection (MacKenzie, 1992; Parhi et al., 2006; Wang et al., 2024), and Wang et al. (2024) showed that larger motor and visual spaces improved selection efficiency. Interface layout can also affect learning efficiency and cognitive load (Zhang et al., 2022), while appropriately managed sensorimotor interaction can support learning and transfer (Lyu & Deng, 2024; Pouw et al., 2014; Zou et al., 2025). Larger practice areas are therefore expected to improve retention (H3a) and transfer (H3b).

A key gap is that prior studies have focused mainly on general interfaces, multiscreen viewing, target selection, mobile learning, or embodied learning broadly. Position, size, and especially their combined effects remain underexplored in VR-based generative practice. Because larger areas may reduce precision demands while right-side placement may align with habitual motor routines, the large-right configuration is expected to provide the best conditions for retention (H4a) and transfer (H4b).

Split-attention effect

The split-attention effect proposes that coordinating spatially separated information sources imposes cognitive costs because learners must divide attention and mentally integrate them (Chandler & Sweller, 1992; Mayer & Moreno, 1998; Sweller et al., 2011). Cierniak et al. (2009) found that split-source formats led to poorer learning outcomes than integrated formats, and that both extraneous and germane load contributed to this effect. Pouw et al. (2019) revealed that greater spatial separation slowed information integration and reduced integrative eye movements. In a systematic review of the spatial split-attention principle, Schroeder and Cenkci (2020) proposed that the benefits of integrated designs may improve allocation of germane resources and integrative cognitive processing rather than reduce extraneous cognitive load.

In the present study, learners must coordinate instructional video content with a separate VR-based practice area. Unlike conventional split-attention research focused mainly on information-source separation or integration, little is known about how the position and size of an interactive practice area affect attention and cognitive load during such alternation.

For position, performance asymmetry (Peters & Ivanoff, 1999) and spatial-compatibility evidence (Wilf et al., 2013) suggest that layouts aligned with dominant-hand routines may facilitate interaction. Thus, for predominantly right-handed learners, right-side placement may make switching between video and practice more efficient, leading to greater attention allocation (H2c) and lower cognitive load (H2d). For size, larger targets reduce precision demands and facilitate selection (MacKenzie, 1992; Parhi et al., 2006; Wang et al., 2024), while interface layout can affect learning-related cognitive load (Zhang et al., 2022). Larger practice areas are therefore expected to facilitate attention allocation (H3c) and reduce cognitive load (H3d).

These mechanisms may also combine. Efficient coordination of spatially separated activities can support processing (Pouw et al., 2019; Schroeder & Cenkci, 2020), while larger targets and spatially compatible positioning may reduce motor and switching demands (MacKenzie, 1992; Parhi et al., 2006; Peters & Ivanoff, 1999; Wang et al., 2024; Wilf et al., 2013). Direct evidence for the interaction of position and size in VR-based generative learning remains limited. However, the combined reduction in target-selection and spatial-coordination demands suggests that the large-right configuration may produce the greatest attention (H4c) and lowest cognitive load (H4d).

Cognitive-affective theory of learning with media

The Cognitive-Affective Theory of Learning with Media (CATLM) emphasizes that interactive designs influence not only cognitive processing but also learning experience (Moreno, 2006). Park et al. (2014) suggested that affective factors can facilitate both cognitive and affective outcomes, providing a basis for considering learners' subjective experience. In immersive learning environments, the Cognitive Affective Model of Immersive

Learning proposes that factors such as interest, motivation, embodiment, and cognitive load jointly influence outcomes in virtual reality (Makransky & Petersen, 2021). More broadly, recent theoretical work suggests that learning in digital environments may also involve social processes in addition to cognitive and affective ones (Schneider et al., 2022).

However, little is known about how spatial properties of an interactive practice area, particularly its position, size, and their combination, influence enjoyment and usability in VR-based generative learning. When interface arrangements reduce motor-control effort and facilitate interaction, learners may experience greater ease of use and perceived control. Accordingly, right-side positioning, when better aligned with the interaction routines of predominantly right-handed learners, is expected to modestly enhance perceived enjoyment (H2e) and layout usability (H2f). Larger interaction areas, by reducing precision demands, are expected to enhance enjoyment (H3e) and usability (H3f). Their combination may further reduce interaction difficulty and increase perceived control; therefore, the large-right configuration is expected to yield the highest enjoyment (H4e) and usability (H4f).

Taken together, the hypotheses concerning position, size, and their interaction are as follows:

H2: Compared with left-side positioning, right-side positioning is expected to improve retention (H2a) and transfer (H2b), increase attention allocation (H2c), reduce cognitive load (H2d), and yield modest improvements in perceived enjoyment (H2e) and layout usability (H2f).

H3: Compared with smaller interaction areas, larger interaction areas are expected to improve retention (H3a) and transfer (H3b), increase attention allocation (H3c), reduce cognitive load (H3d), and increase perceived enjoyment (H3e) and layout usability (H3f).

H4: Position and size are expected to interact, with the large-right configuration producing the most favorable outcomes across retention (H4a), transfer (H4b), attention (H4c), cognitive load (H4d), enjoyment (H4e), and usability (H4f), and the small-left configuration producing the least favorable outcomes.

Methods

Participants and research design

A total of 100 university students (50 females, 50 males, ages 18–31, $M=20.5$, $SD=2.69$), all of whom were right-handed and learning English as a foreign language. The study used a $2\times2+1$ factorial design with two between-subjects factors: position (left vs. right) and area size (large vs. small) of the interaction area, plus a control group. The effect size was based on the data (>0.44) reported in Kyaw et al. (2019). Using G*Power 3.1 (Version 3.1.9.7; Faul et al., 2007) calculations (effect size $f=0.45$, power $=0.80$, alpha $=0.05$), the target sample size was 90 participants in total, which was met.

Application and generative interaction activities

The application, built with Unity 3D and Blender, teaches colour vocabulary to foreign language learners through interactive and generative activities. It features two areas: an interactive area and a video area. The interactive area functions as the practice space, where learners can select target words, adjust brush settings, rotate objects, erase, and capture snapshots. The video area allows play, pause, fast-forward, and navigation.

As part of the generative learning tasks, learners engaged with nine English colour words, that is, aqua, azure, sienna, mauve, fuchsia, vitelline, tangerine, chartreuse, and ultramarine, by actively applying them in the design of artistic donuts, thereby fostering deeper semantic processing and integration. The design tasks were inspired by *The Tale of the Peach Blossom Spring*; for example, one assignment required students to create a donut using chartreuse and sienna, reflecting the village's cultural association with wine and pastries.

The 5-minute, 52-second video explains pronunciation, meaning, origin, symbolism, and usage of the colours. The interface offers two sizes: large (1152 × 800) and small (768 × 800). When the interactive area is positioned on the left with a large size, the video area is on the right with a small size. As shown in Fig. 1, the four experimental conditions vary by interactive area position and size: LVRL (The VR interactive area is on the left with a larger size, i.e., Left VR and Large), RVRL (Right VR and Large), LVRS (Left VR and Small), and RVRS (Right VR and Small). The control group views only the instructional video (IV). Tasks are consistent across all conditions.

Measures

Prior knowledge

The prior knowledge test had nine fill-in-the-blank questions on the nine words. Participants wrote meanings of each word if they knew, earning one point per correct answer (max nine points). None had prior knowledge of these words.

Retention and transfer

We measured retention and transfer using 27 multiple-choice questions (nine words × three types). The retention test had 18 questions, and the transfer test had nine. Participants earned 1 point per correct answer, with a maximum of 27 points. To discourage guessing, one option was "I do not know."

The first retention type showed a coloured bar, asking participants to select the corresponding word. The second, also the retention type, showed an image with multiple colours, and participants selected the matching word for one colour. The transfer test gave a word, asking participants to choose the matching scene or object. The McDonald's ω values (Hayes & Coutts, 2020; Shaw, 2021) indicated high internal consistency (retention: 0.83, transfer: 0.72, overall: 0.86).

Subjective scales

Attention, layout usability, and perceived enjoyment were assessed using a seven-point Likert scale (1 = strongly disagree, 7 = strongly agree; as shown in Table 1). Higher scores indicated a higher level of agreement. Attention

Fig. 1 The interface of the application and four experimental conditions (translated)

Table 1 Subjective ratings

Variable	Numbers of questions	Type	Reliability (McDonald's ω)	Reference
Attention	3	7-point Likert	0.875	Revised from (Salomon, 1984)
Perceived enjoyment	3		0.913	(Tokel & İsler, 2015)
Layout usability	6		0.922	Self-designed

was revised for the context, perceived enjoyment was referenced, and layout usability was self-designed with six items:

1. The layout of this interactive learning environment makes me feel comfortable.
2. The layout conforms to my usage habits.
3. The layout is simple and easy for me to use.
4. I feel that the size and position of the interactive area in the layout are appropriate.
5. I feel that the size and position of the video area in the layout are appropriate.
6. I feel that the design of the layout aligns with what I expected.

Cognitive load and attention allocation

This study compares learners' cognitive load across five conditions using average fixation duration (0.1–0.5), an eye-tracking indicator that correlates positively with cognitive load (Alemdag & Cagiltay, 2018; Krejtz et al., 2018; van Gog & Jarodzka, 2013). We examined attention to interactive and video areas via total fixation duration in Areas of Interest (AOIs) and explored attention to the instructor using fixation duration percentage (Rosch & Vogel-Walcutt, 2013).

Procedure and data analysis

Participants were first instructed about the experiment, signed the consent form, and completed demographic and prior knowledge tests. Then, they watched a video, interacted with the app on a laptop for tasks, and completed posttests, subjective ratings, and a brief interview for feedback, taking approximately 40 min overall. Kolmogorov-Smirnov tests and visual inspections confirmed normal distribution of most data across the five conditions, allowing for ANOVA analysis. Effect sizes in ANOVAs were measured using ηp^2 (Richardson, 2011), with thresholds for small (0.0099), medium (0.0588), and large (0.1379) effects. We used Kruskal-Wallis tests for non-normally distributed data.

Results

Table 2 shows the descriptive statistics for learning performance, subjective scales, and average duration of fixation across the five groups.

Retention

One-way ANOVA showed no significant differences in retention across the five groups ($F(4,95) = 1.830$, $p = 0.129$, $\eta p^2 = 0.072$), with the RVRL group performing highest ($M = 14.95$). These results did not support H1a.

Among the four VR conditions, two-way ANOVA revealed no main effect of position on retention ($F(1,76) = 0.868$, $p = 0.355$, $\eta p^2 = 0.011$), providing no evidence for H2a. However, a significant main effect of

Table 2 Descriptive statistics

Measures	LVRL ($n=20$)	RVRL ($n=20$)	LVRS ($n=20$)	RVRS ($n=20$)	IV ($n=20$)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Retention	13.65 (3.42)	14.95 (3.28)	12.30 (3.48)	12.45 (3.72)	12.85 (4.06)
Transfer	5.75 (1.94)	5.85 (2.11)	4.40 (1.76)	5.00 (1.56)	4.20 (1.79)
Attention	4.33 (1.74)	5.38 (1.26)	4.15 (1.62)	5.00 (1.27)	5.50 (0.83)
Perceived enjoyment	5.63 (0.85)	5.68 (1.05)	5.28 (1.19)	5.72 (0.77)	N/A
Layout usability	5.47 (1.11)	5.73 (0.99)	4.29 (1.59)	4.70 (1.05)	N/A
Cognitive load (Average duration of fixation)	0.32 (0.05)	0.30 (0.04)	0.36 (0.04)	0.33 (0.05)	0.25 (0.05)

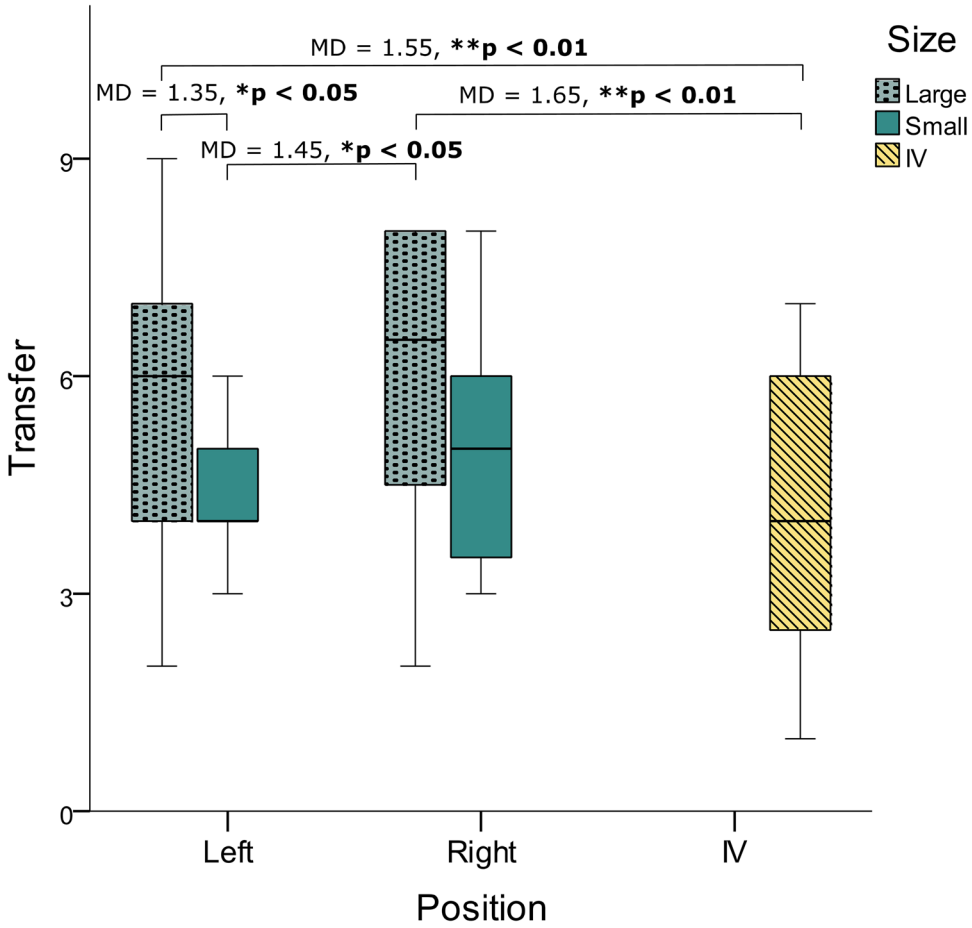

Fig. 2 Differences in the transfer

area size was found ($F(1,76)=6.118$, $p=0.016$, $\eta p^2 = 0.075$), indicating that larger interaction areas enhanced retention. This result supported H3a.

No interaction effect on retention was observed between position and area size ($F(1,76)=0.546$, $p=0.462$, $\eta p^2 = 0.007$). Thus, no evidence was found to support H4a.

Transfer

One-way ANOVA indicated significant differences in learning transfer ($F(4,95)=3.356$, $p=0.013$, $\eta p^2 = 0.124$; see Fig. 2). Tukey's HSD tests revealed that the LVRL group scored higher than the LVRS group (Mean Difference = 1.35, $p<0.05$) and IV group (MD = 1.55, $p<0.01$). Similarly, RVRL outperformed LVRS (MD = 1.45, $p<0.05$) and IV (MD = 1.65, $p<0.01$). No other group differences were significant. These results partially

supported H1b, as only the two large-area VR conditions (LVRL and RVRL) significantly outperformed the video-only condition.

Among the four VR conditions, two-way ANOVA showed no significant main effect of position on transfer ($F(1,76)=0.713$, $p=0.401$, $\eta p^2 = 0.009$), providing no evidence for H2b. In contrast, a significant main effect of area size was observed ($F(1,76)=7.044$, $p=0.010$, $\eta p^2 = 0.085$), with larger areas yielding higher transfer scores than smaller ones. This result supported H3b.

No interaction effect on transfer was found between position and area size ($F(1,76)=0.364$, $p=0.548$, $\eta p^2 = 0.005$). Thus, no evidence was found to support H4b.

Attention

The Kruskal-Wallis test showed a slight difference in attention levels across the five conditions ($H(4)=9.55$, $p=0.049$), but post hoc tests with Bonferroni correction revealed no significant pairwise differences. The IV group ($M=5.50$, $SD=0.83$) and the RVRL group ($M=5.38$, $SD=1.26$) had the highest values. These results provided limited evidence and only partially supported H1c.

Among the four VR conditions, two-way ANOVA revealed a significant main effect of interactive area position on attention ($F(1,75)=8.067$, $p=0.006$, $\eta p^2 = 0.097$), with right-side interactive areas attracting more attention than left-side areas. This result supported H2c. No significant main effect of area size was found ($F(1,75)=0.713$, $p=0.401$, $\eta p^2 = 0.009$), providing no evidence for H3c.

No significant interaction effect between position and size was observed ($F(1,75)=0.087$, $p=0.768$, $\eta p^2 = 0.001$). Thus, no evidence was found to support H4c.

Attention allocation

The application included three predefined areas of interest (AOIs): AOI1 (interactive area), AOI2 (video area), and AOI3 (teacher area). Table 3 presents the Wilcoxon signed-rank test results for total fixation duration in AOI1 and AOI2 across the four experimental groups. Across all VR conditions, learners spent significantly more time fixating on the interactive area than on the video area, indicating a prioritization of interactive practice over passive video viewing.

Significant differences were also observed in fixation duration toward AOI3 (teacher area) across the five experimental conditions ($H(4)=18.81$, $p=0.001$). As shown in Fig. 3, multiple comparisons revealed that the IV group devoted a significantly higher percentage of fixation time to the teacher area than the LVRL group ($p=0.002$, $r=0.38$), the RVRL group ($p=0.017$, $r=0.32$), the LVRS group ($p=0.012$, $r=0.33$), and the RVRS group ($p=0.008$, $r=0.34$). No significant differences were found among the four VR groups. These findings provide process-level evidence that VR-based generative practice consistently redirected learners' attention from the video and teacher areas toward the interactive tasks, thereby partially supporting H1c.

Table 3 Total duration of fixation in AOI1 and AOI2

	AOIs	Total duration of fixation in AOI M (SD)	Z	P
LVRL	AOI1	606.84 (231.60)	-3.72[b]	0.00
	AOI2	236.18 (120.17)		
RVRL	AOI1	554.95 (180.22)	-3.81[b]	0.00
	AOI2	236.08 (116.19)		
LVRS	AOI1	480.31 (129.77)	-3.51[b]	0.00
	AOI2	256.09 (115.12)		
RVRS	AOI1	609.39 (143.96)	-3.78[b]	0.00
	AOI2	254.47 (96.92)		

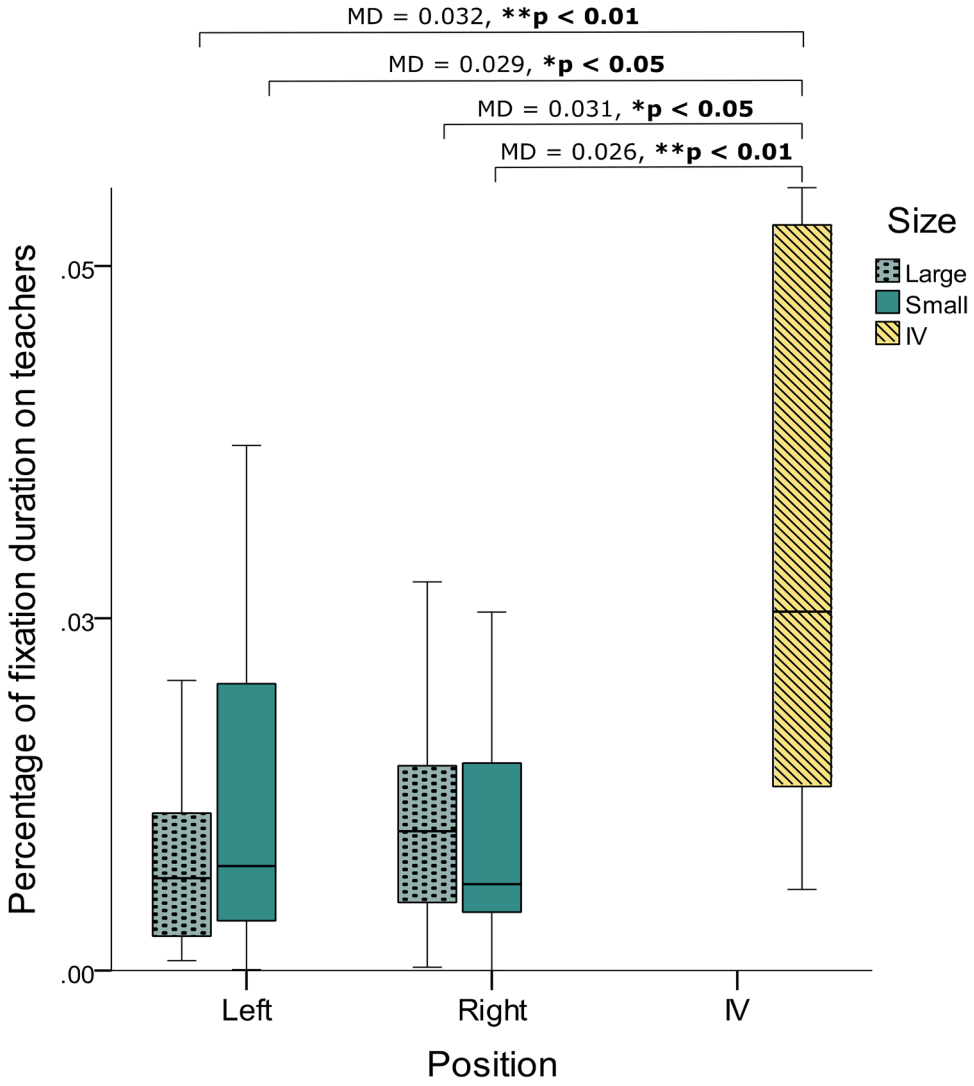

Fig. 3 Pairwise comparisons of fixation-duration percentages

Cognitive load

Table 2 presents the descriptive statistics of cognitive load, indexed by average fixation duration (0.1–0.5), which has been shown to correlate positively with cognitive load. The Kruskal-Wallis test revealed significant differences across the five conditions ($H(4)=35.76$, $p<0.001$; see Fig. 4). Multiple comparisons indicated that the LVRS group had significantly longer fixation durations than the RVRL group ($p=0.008$, $r=0.34$). In contrast, the IV group showed significantly shorter fixation durations than the LVRL ($p=0.004$, $r=0.36$), LVRS ($p<0.001$, $r=0.59$), and RVRS groups ($p=0.002$, $r=0.38$). These results partially supported H1d because three of the four VR conditions showed significantly higher cognitive load than the video-only condition, whereas the difference between RVRL and IV was not significant.

Two-way ANOVAs revealed significant main effects of both position ($F(1,72)=5.967$, $p=0.017$, $\eta p^2 = 0.077$) and size ($F(1,72)=9.154$, $p=0.003$, $\eta p^2 = 0.113$). Learners using right-side interactive areas showed lower cognitive load than those using left-side areas, supporting H2d. Learners with larger interaction areas reported lower cognitive load than those with smaller ones, supporting H3d. No significant interaction effect between position and size was observed ($F(1,72)=1.013$, $p=0.318$, $\eta p^2 = 0.014$). Thus, no evidence was found to support H4d.

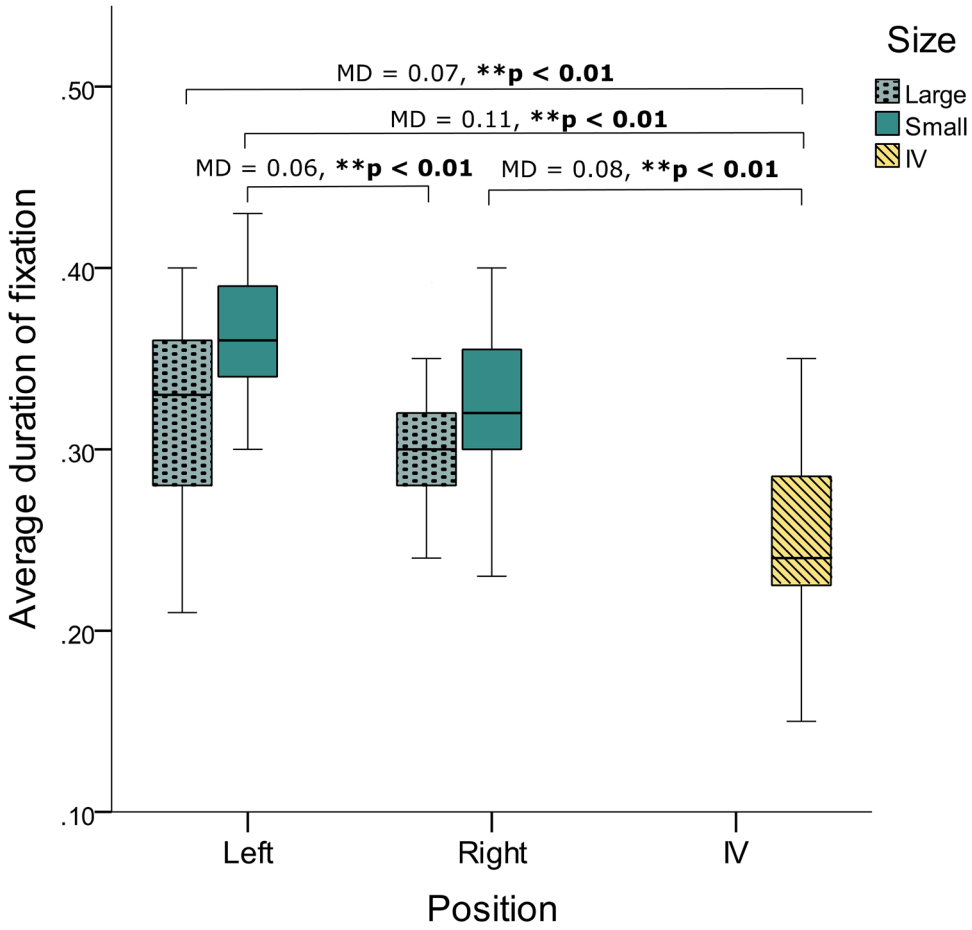

Fig. 4 Group differences in average fixation duration

Perceived enjoyment

Perceived enjoyment scores were high across all four experimental groups, with mean values exceeding 5, indicating generally high levels of enjoyment among learners in the VR conditions. The Kruskal-Wallis test showed no significant differences in perceived enjoyment ratings across conditions ($H(3)=1.56$, $p=0.670$). The RVRS group ($M=5.72$, $SD=0.77$) and RVRL group ($M=5.68$, $SD=1.05$) reported descriptively higher enjoyment; however, these differences were not statistically significant.

However, two-way ANOVAs indicated no significant main effects of position ($F(1,74)=1.149$, $p=0.287$, $\eta p^2 = 0.015$) or size ($F(1,74)=0.473$, $p=0.494$, $\eta p^2 = 0.006$) on perceived enjoyment, nor was there a significant interaction effect ($F(1,74)=0.720$, $p=0.399$, $\eta p^2 = 0.010$). These results provided no evidence for H2e, H3e, and H4e.

Layout usability

The Kruskal-Wallis test revealed a significant difference in layout usability across the four VR conditions ($H(3)=13.75$, $p=0.003$; see Fig. 5). Bonferroni-corrected pairwise comparisons showed that the RVRL group rated the layout significantly higher than the LVRS group ($p=0.015$, $r=0.34$) and the RVRS group ($p=0.038$, $r=0.31$). No significant differences were found among the other groups.

Two-way ANOVAs indicated no significant main effect of interactive area position on layout usability ($F(1,75)=1.517$, $p=0.222$, $\eta p^2 = 0.020$), providing no evidence for H2f. However, a significant main effect of interactive area size was found ($F(1,75)=16.676$, $p<0.001$, $\eta p^2 = 0.182$), indicating that groups using large interactive areas reported significantly higher layout usability scores than those using small areas. This supported

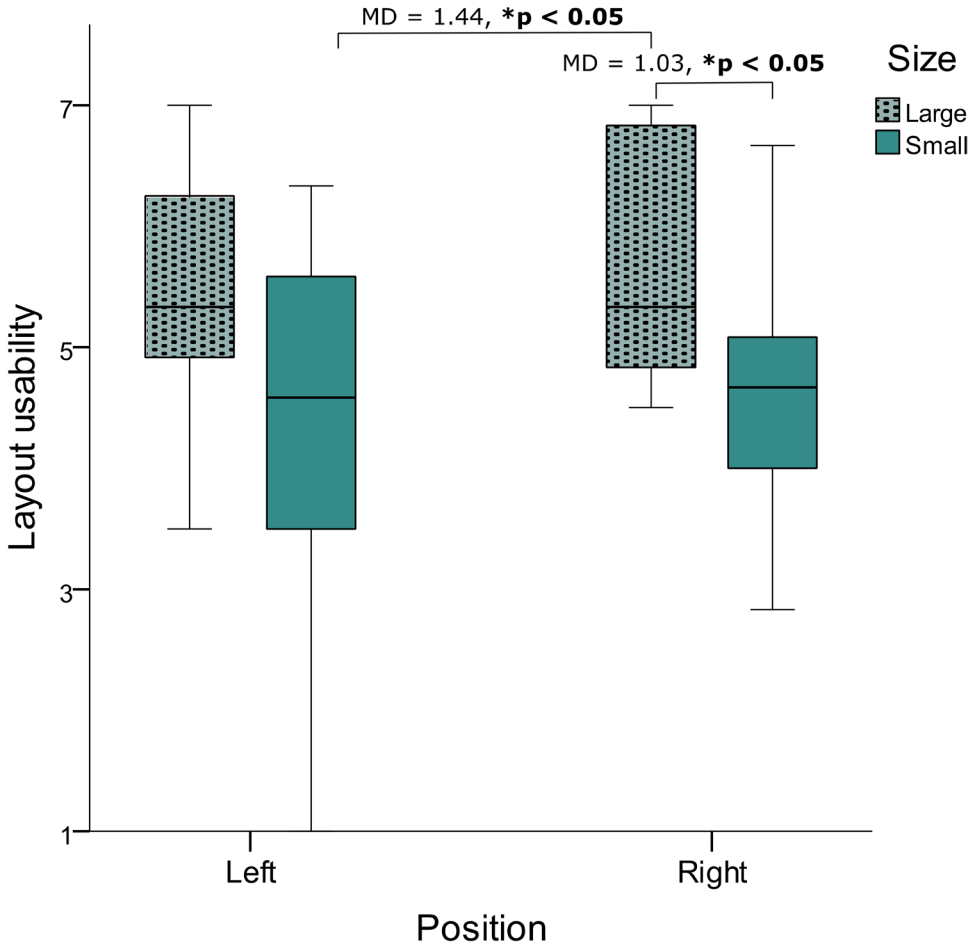

Fig. 5 Differences in the layout usability

H3f. The interaction effect between position and size was not significant ($F(1,75) = 0.081$, $p = 0.777$, $\eta p^2 = 0.001$), indicating no evidence to support H4f.

Table 4 summarizes the results of hypothesis testing.

General discussion

This study investigated the impact of integrating VR-based generative practice into instructional videos, with a focus on the spatial design of the interaction (i.e., position and size). Three key contributions are as follows.

The effect of VR-based generative practice compared to video-only learning

The results showed that the transfer benefits of VR-based generative practice depended on practice-area size. The two large-area VR groups (LVRL and RVRL) significantly outperformed the video-only group, whereas the two small-area VR groups did not. This pattern is consistent with prior work showing that drawing during video learning enhances learning gains (Zhang et al. 2025a, b), scaffolded self-explanation improves performance and can reduce cognitive load (Wang & Xu, 2024), and explaining to others promotes transfer and deeper cognitive engagement (Zhu et al., 2024). Eye-tracking results further showed that learners in all four VR conditions allocated more fixation time to the interactive area than to the video area, while video-only learners devoted more attention to the instructor than learners in each VR condition. These findings provide process-level evidence of

Table 4 The results of hypothesis testing

Hypothesis	Description	Result
H1a	VR practice > Video-only on retention	×
H1b	VR practice > Video-only on transfer	O
H1c	VR practice > Video-only on attention	O
H1d	VR practice > Video-only on cognitive load	O
H2a	Right > Left on retention	×
H2b	Right > Left on transfer	×
H2c	Right > Left on attention	√
H2d	Right < Left on cognitive load	√
H2e	Right > Left on enjoyment	×
H2f	Right > Left on usability	×
H3a	Larger > Smaller on retention	√
H3b	Larger > Smaller on transfer	√
H3c	Larger > Smaller on attention	×
H3d	Larger < Smaller on cognitive load	√
H3e	Larger > Smaller on enjoyment	×
H3f	Larger > Smaller on usability	√
H4a	Right + Large → best retention	×
H4b	Right + Large → best transfer	×
H4c	Right + Large → most attention	×
H4d	Right + Large → lowest cognitive load	×
H4e	Right + Large → highest enjoyment	×
H4f	Right + Large → best usability	×

Note: √ (supported, statistically significant results in the predicted direction); × (not supported, no significant effects or effects in the opposite direction); O (partially supported, overall differences observed but no significant pairwise effects or only limited evidence)

attention allocation toward active interaction. Consistent with prior evidence (Tai et al., 2022), this study reveals the potential of appropriately designed interactive practice to support vocabulary transfer.

The results therefore qualify rather than uniformly confirm the expected benefits of VR-based generative practice. Retention did not differ significantly between VR and video-only learning, and transfer advantages emerged only in the large-area conditions. Moreover, three VR conditions showed higher cognitive load than video-only learning, whereas RVRL did not differ significantly from the video-only condition. Because perceived enjoyment was measured only in the four VR groups, no conclusion can be drawn regarding whether VR-based generative practice increased enjoyment relative to video-only learning. From the perspective of generative learning theory, the observed transfer advantages in the large-area conditions may reflect learners' active selection, organization, integration, and representation of information through interaction (Fiorella & Mayer, 2021). More recent work further suggests that explaining, visualizing, and enacting are complementary forms of generative sense-making (Fiorella, 2023). In the present study, VR-based practice combined visualizing and enacting, providing opportunities for learners to engage with learning content through active interaction rather than passive viewing.

Practically, instructional videos should incorporate structured generative tasks when the goal is transfer and application rather than content delivery alone. VR-based practice can shift learners from watching to constructing, but it should be closely aligned with instructional content and designed to support meaningful generation rather than distraction.

The effect of spatial positioning

Second, spatial positioning influenced attention and cognitive load but did not significantly affect retention, transfer, perceived enjoyment, or layout usability. Right-side placement was associated with greater attention allocation and lower cognitive load than left-side placement, suggesting a process-level rather than a direct learning-performance advantage. This pattern is broadly consistent with prior interface research showing that layout can influence gaze patterns, usability, and task efficiency, although findings in general interface settings remain mixed (Holmes et al., 2012; Kalbach & Bosenick, 2003; McCarthy et al., 2004; Murano & Lomas, 2015; Vatavu

& Mancas, 2015). It also aligns with learning-related evidence that layout can affect learning efficiency, usability, and cognitive load (Ragan et al., 2012; Zhang et al., 2022).

Theoretically, these findings can be explained by embodied cognition and split-attention perspectives. Embodied cognition suggests that cognitive processing is shaped by perception-action couplings such as handedness, reachability, and habitual scan paths (Bridgeman & Tseng, 2011; Ranzini et al., 2022; Thorvald et al., 2020; Wilson, 2002). For right-handed learners, right-side placement may better align with embodied motor-control routines, reducing interaction effort and attentional switching costs, a pattern also consistent with handedness-related asymmetries in mouse control (Peters & Ivanoff, 1999). From a split-attention perspective, positioning may also matter because learners must shift and integrate attention between instructional video and a separate practice area (Chandler & Sweller, 1992; Mayer & Moreno, 1998; Pouw et al., 2019; Schroeder & Cenkci, 2020). In this sense, the present results do not simply show that "right is better than left"; rather, they suggest that positioning effects emerge when a layout aligns better with learners' embodied routines and reduces coordination demands in dual-space learning. Based on this pattern, we propose the Attentional Alignment Principle: interactive layouts support more efficient active learning when they better align with learners' embodied routines across task spaces.

Practically, spatial positioning should be treated as an instructional design factor rather than only a visual choice. For right-handed learners, right-side placement may be a useful default because it aligns with embodied routines, may reduce cognitive load, and supports more efficient attention allocation.

The effect of practice-area size

Third, we found that practice-area size significantly influenced retention, transfer, cognitive load, and layout usability, but not attention allocation or perceived enjoyment. Larger interaction areas produced higher retention and transfer, lower cognitive load, and better usability than smaller areas, regardless of whether the area was positioned on the left or right. These findings suggest that the benefits of a larger workspace are not due to attracting more attention per se. One possible interpretation is that larger practice areas may reduce interaction demands and facilitate generative practice, potentially involving lower motor effort and extraneous cognitive load. However, the roles of these factors remain to be further examined. This interpretation aligns with prior research showing that larger displays increase arousal and improve memory outcomes (Reeves et al., 1999; Segijn et al., 2017; Vatavu & Mancas, 2015). Interface layout can shape learning efficiency, usability, and cognitive load in learning environments (Zhang et al., 2022). Vatavu and Mancas (2015) found that larger screens attract more visual attention. The study by Reeves et al. (1999) indicated that larger screens elicit higher arousal levels and positively impact memory retention compared to smaller screens. Thus, the present study adds evidence that, in VR-supported generative learning, size matters not only for interaction comfort but also for meaningful learning outcomes.

We term this the "Workspace Expansion Principle," which proposes that expanding the spatial affordances of generative practice areas (e.g., larger interaction interfaces) may reduce interaction demands, thereby freeing cognitive resources for generative processing and supporting learning performance. However, the roles of motor effort and extraneous cognitive load in this proposed explanation remain to be further examined. Future studies should directly measure these variables and investigate the mechanisms underlying the observed effects.

Practically, these findings suggest that the spatial size of interactive practice areas should be treated as a core instructional design feature. In designing VR-supported instructional videos, providing learners with adequate practice space may therefore be one of the most effective ways to improve both transfer-oriented learning and usability.

Limitations and future directions

This study has several limitations. First, the advantage of right-side practice areas was observed only among right-handed participants, requiring validation with left-handed or mixed populations. Second, the experiment was a one-time intervention; longitudinal or repeated designs are needed to examine sustained effects. Third, the generative task was limited to virtual object design, which restricts task diversity. Future research should test whether benefits extend to other generative formats such as text production, visual representation, oral expression, or cross-modal tasks. Additionally, the focus on vocabulary learning limits generalizability. Further work should investigate whether the proposed two principles apply to higher-order skills and abstract knowledge learning. Finally, average fixation duration was used only as an eye-tracking indicator associated with cognitive load, rather than as direct evidence of extraneous load or motor effort. The roles of motor effort and extraneous cognitive load in explaining the observed effects of spatial design therefore remain to be further examined. Future studies should directly measure these variables and combine eye-tracking with subjective and behavioral measures to investigate the mechanisms underlying these effects.

Abbreviations

VR	Virtual reality
M	Mean
SD	Standard deviation
MD	Mean difference
LVRL	The VR interactive area is on the left with a larger size, i.e., Left VR and Large
RVRL	Right VR and large
LVRS	Left VR and small
RVRS	Right VR and small
IV	The control group views only the instructional video
AOI	Areas of interest
ANOVA	Analysis of variance

Acknowledgements We would like to thank Ms Jialin Yu and Mr Dong Yang for their technical support in the AI teacher generation program.

Author contributions XL contributed to the conceptualization, methodology, formal analysis, visualization, and original draft preparation. YZ led the conceptualization, methodology, supervision, investigation, formal analysis, validation, writing of the original draft, reviewing and editing, and funding acquisition. BD assisted with visualization. HZ carried out the formal analysis. LC and BL contributed to resources. TX contributed to conceptualization, methodology, software development, resources, and funding acquisition. All authors read and approved the final manuscript.

Funding This work was supported by the Research Projects of the National Natural Science Foundation of China under Grant [62377039]; the National Key R&D Program of China under Grant [2022YFC3303600]; CAAI-CANN Open Fund, developed on OpenI Community; Higher Education Research Fund of Northwestern Polytechnical University in 2025, No. GJJJM202504.

Data availability The data are available from the corresponding author upon request.

Declarations

Ethical approval The study followed the ethical guidelines of the institutional and national research committee, as well as WMA Declaration of Helsinki. All participants were informed of the voluntary nature of their participation and asked to provide written informed consent. They were assured of anonymity and informed about the purpose of the research, how their data would be used, and that no risks were associated with their involvement.

Consent to participate All participants signed informed consent agreement.

Artificial intelligence generated content None.

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Mamun, A. M., & Lawrie, G. (2023). Student-content interactions: Exploring behavioural engagement with self-regulated inquiry-based online learning modules. *Smart Learning Environments, 10*(1). https://doi.org/10.1186/s40561-022-00221-x

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Beege, M., & Ploetzner, R. (2025). Learning from interactive video: The influence of self-explanations, navigation, and cognitive load. *Instructional Science*, *53*(1), 99–119. https://doi.org/10.1007/s11251-024-09693-5

Boateng, R., Boateng, S. L., Awuah, R. B., Ansong, E., & Anderson, A. B. (2016). Videos in learning in higher education: Assessing perceptions and attitudes of students at the University of Ghana. *Smart Learning Environments, 3*(1). https://doi.org/10.1186/s40561-016-0031-5

Bridgeman, B., & Tseng, P. (2011). Embodied cognition and the perception–action link. *Physics of Life Reviews*, *8*(1), 73–85. https://doi.org/10.1016/j.plrev.2011.01.002

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233–246. https://doi.org/10.1111/j.2044-8279.1992.tb01017.x

Chen, B., Wang, Y., & Wang, L. (2022a). The effects of virtual reality-assisted language learning: A meta-analysis. *Sustainability, 14*(6). https://doi.org/10.3390/su14063147

Chen, C. M., Li, M. C., & Lin, M. F. (2022b). The effects of video-annotated learning and reviewing system with vocabulary learning mechanism on English listening comprehension and technology acceptance. *Computer Assisted Language Learning*, *35*(7), 1557–1593. https://doi.org/10.1080/09588221.2020.1825093

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Cromley, J. G., Chen, R., & Lawrence, L. (2023). Meta-analysis of STEM learning using virtual reality: Benefits across the board. *Journal of Science Education and Technology*, *32*(3), 355–364. https://doi.org/10.1007/s10956-023-10032-5

De Lorenzis, F., Visconti, A., Restivo, S., Mazzini, F., Esposito, S., Garofalo, S. F., Marmo, L., Fino, D., & Lamberti, F. (2024). Combining virtual reality with asymmetric collaborative learning: A case study in chemistry education. *Smart Learning Environments*, *11*(1), 43. https://doi.org/10.1186/s40561-024-00331-8

Deng, R., & Gao, Y. (2024). Effects of embedded questions in pre-class videos on learner perceptions, video engagement, and learning performance in flipped classrooms. *Active Learning in Higher Education*, *25*(3), 473–487. https://doi.org/10.1177/14697874231167098

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiorella, L. (2023). Making sense of generative learning. *Educational Psychology Review, 35*(2). https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., & Mayer, R. E. (2021). The generative activity principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 339–350). Cambridge University Press. https://doi.org/10.1017/9781108894333.036

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *WIREs Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Geri, N., Winer, A., & Zaks, B. (2017). Challenging the six-minute myth of online video lectures: Can interactivity expand the attention span of learners? *Online Journal of Applied Knowledge Management (OJAKM)*, *5*(1), 101–111. https://doi.org/10.36965/OJAKM.2017.5(1)101-111

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *But… Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Holmes, M. E., Josephson, S., & Carney, R. E. (2012). Visual attention to television programs with a second-screen application. In *Proceedings of the symposium on eye tracking research and applications* (pp. 397–400). Association for Computing Machinery. https://doi.org/10.1145/2168556.2168646

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Hung, I. C., & Chen, N. S. (2018). Embodied interactive video lectures for improving learning comprehension and retention. *Computers & Education*, *117*, 116–131. https://doi.org/10.1016/j.compedu.2017.10.005

Jiang, L., & Chen, Y. H. (2022). Menu design on small display user interfaces: Measuring the influence of menu type, number of preview items, and menu breadth on navigation efficiency. *International Journal of Human–Computer Interaction*, *38*(7), 631–645. https://doi.org/10.1080/10447318.2021.1954781

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*, *37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567

Kalbach, J., & Bosenick, T. (2003). Web page layout: A comparison between left- and right-justified site navigation menus. *Journal of Digital Information, 4*(1).

Karnchanapayap, G. (2023). Activities-based virtual reality experience for better audience engagement. *Computers in Human Behavior, 146,* Article 107796. https://doi.org/10.1016/j.chb.2023.107796

Kim, J., Glassman, E. L., Monroy-Hernández, A., & Morris, M. R. (2015). RIMES: Embedding interactive multimedia exercises in lecture videos. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1535–1544). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702186

Kincaid, C. S., & Corsun, D. L. (2003). Are consultants blowing smoke? An empirical test of the impact of menu layout on item sales. *International Journal of Contemporary Hospitality Management*, *15*(4), 226–231. https://doi.org/10.1108/09596110310475685

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE, 13*(9), e0204709. https://doi.org/10.1371/journal.pone.0203629. Article e0203629.

Kyaw, B. M., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., Divakar, U., Masiello, I., Kononowicz, A. A., Zary, N., & Car, L. T. (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration. *Journal of Medical Internet Research, 21*(1), Article e12959. https://doi.org/10.2196/12959

Lee, S. M., Yang, Z., & Wu, J. G. (2024). Live, play, and learn: Language learner engagement in the immersive VR environment. *Education and Information Technologies*, *29*(9), 10529–10550. https://doi.org/10.1007/s10639-023-12215-4

Lin, X. P., Li, B. B., Yao, Z. N., Yang, Z., & Zhang, M. (2024). The impact of virtual reality on student engagement in the classroom–a critical review of the literature. *Frontiers in Psychology*, *15*, 1360574. https://doi.org/10.3389/fpsyg.2024.1360574

Lyu, C., & Deng, S. (2024). Effectiveness of embodied learning on learning performance: A meta-analysis based on the cognitive load theory perspective. *Learning and Individual Differences, 116,* Article 102564. https://doi.org/10.1016/j.lindif.2024.102564

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human–Computer Interaction*, *7*(1), 91–139. https://doi.org/10.1207/s15327051hci0701_3

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Marchionini, G. (2003). Video and learning redux: New capabilities for practical use. *Educational Technology*, *43*(2), 36–41. https://www.jstor.org/stable/44428819

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

McCarthy, J. D., Sasse, M. A., & Riegelsberger, J. (2004). Could I have the menu please? An eye tracking study of design conventions. In *People and computers XVII—Designing for society: Proceedings of HCI 2003* (pp. 401–414). Springer. https://doi.org/10.1007/978-1-4471-3754-2_25

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*(3), 149–158. https://doi.org/10.1111/j.1365-2729.2006.00170.x

Murano, P., & Lomas, T. J. (2015). Menu positioning on web pages: Does it matter? *International Journal of Advanced Computer Science and Applications, 6*(4). https://doi.org/10.14569/IJACSA.2015.060419

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on human-computer interaction with mobile devices and services* (pp. 203–210). Association for Computing Machinery. https://doi.org/10.1145/1152215.1152260

Park, B., Plass, J. L., & Brünken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction*, *29*, 125–127. https://doi.org/10.1016/j.learninstruc.2013.05.005

Peters, M., & Ivanoff, J. (1999). Performance asymmetries in computer mouse control of right-handers, and left-handers with left- and right-handed mouse experience. *Journal of Motor Behavior*, *31*(1), 86–94. https://doi.org/10.1080/00222899909601894

Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51–72. https://doi.org/10.1007/s10648-014-9255-5

Pouw, W., Rop, G., de Koning, B., & Paas, F. (2019). The cognitive basis for the split-attention effect. *Journal of Experimental Psychology: General*, *148*(11), 2058–2075. https://doi.org/10.1037/xge0000578

Ragan, E. D., Endert, A., Bowman, D. A., & Quek, F. (2012). How spatial layout, interactivity, and persistent visibility affect learning with large displays. In *Proceedings of the international working conference on advanced visual interfaces* (pp. 91–98). Association for Computing Machinery. https://doi.org/10.1145/2254556.2254576

Rahimi, A. R., & Sevilla-Pavón, A. (2025). The role of interactive, constructive, active, and passive learning activities (ICAPCALL) in shaping students' online engagement, and learning approaches to virtual exchange (SAVE): A bisymmetric approach. *Smart Learning Environments, 12*(1). https://doi.org/10.1186/s40561-024-00363-0

Ralph, B. C. W., Thomson, D. R., Seli, P., Carriere, J. S. A., & Smilek, D. (2015). Media multitasking and behavioral measures of sustained attention. *Attention Perception & Psychophysics*, *77*(2), 390–401. https://doi.org/10.3758/s13414-014-0771-7

Ranzini, M., Semenza, C., Zorzi, M., & Cutini, S. (2022). Influences of hand action on the processing of symbolic numbers: A special role of pointing? *PLOS ONE, 17*(6), Article e0269697. https://doi.org/10.1371/journal.pone.0269557

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology*, *1*(1), 49–67. https://doi.org/10.1207/s1532785xmep0101_4

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

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

Sablić, M., Mirosavljević, A., & Škugor, A. (2021). Video-based learning (VBL)—past, present and future: An overview of the research published from 2008 to 2019. *Technology Knowledge and Learning*, *26*(4), 1061–1077. https://doi.org/10.1007/s10758-020-09455-5

Salomon, G. (1984). Television is easy and print is tough: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schroeder, N. L., & Cenkci, A. T. (2020). Do measures of cognitive load explain the spatial split-attention principle in multimedia learning environments? A systematic review. *Journal of Educational Psychology*, *112*(2), 254–270. https://doi.org/10.1037/edu0000372

Segijn, C. M., Voorveld, H. A. M., Vandeberg, L., & Smit, E. G. (2017). The battle of the screens: Unraveling attention allocation and memory effects when multiscreening. *Human Communication Research*, *43*(2), 295–314. https://doi.org/10.1111/hcre.12106

Shaw, B. P. (2021). Meeting assumptions in the estimation of reliability. *The Stata Journal: Promoting Communications on Statistics and Stata*, *21*(4), 1021–1027. https://doi.org/10.1177/1536867X211063407

Shen, W. W., Lin, J. M., Cheng, W. K., & Hong, Z. W. (2023). Developing and evaluating an online video-assisted collocation learning system for EFL students. *Interactive Learning Environments*, *31*(8), 5292–5306. https://doi.org/10.1080/10494820.2021.2003408

Shih, R. C. (2010). Blended learning using video-based blogs: Public speaking for English as a second language students. *Australasian Journal of Educational Technology*, *26*(6), 883–897. https://doi.org/10.14742/ajet.1048

Skulmowski, A. (2024). Learning by doing or doing without learning? The potentials and challenges of activity-based learning. *Educational Psychology Review, 36*(1). https://doi.org/10.1007/s10648-024-09869-y

Stalheim, O. R., & Somby, H. M. (2024). An embodied perspective on an augmented reality game in school: Pupil's bodily experience toward learning. *Smart Learning Environments, 11*(1). https://doi.org/10.1186/s40561-024-00308-7

Sweller, J., Ayres, P., & Kalyuga, S. (2011). The split-attention effect. In J. Sweller, P. Ayres, & S. Kalyuga (Eds.), *Cognitive load theory* (pp. 111–128). Springer. https://doi.org/10.1007/978-1-4419-8126-4_9

Tai, T. Y., Chen, H. H. J., & Todd, G. (2022). The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Computer Assisted Language Learning*, *35*(4), 892–917. https://doi.org/10.1080/09588221.2020.1752735

Thorvald, P., Lindblom, J., & Kolbeinsson, A. (2020). Embodied interactions in cognitive manufacturing. In H. Ayaz (Ed.), *Advances in neuroergonomics and cognitive engineering* (pp. 419–426). Springer International Publishing. https://doi.org/10.1007/978-3-030-20473-0_41

Tokel, S. T., & İsler, V. (2015). Acceptance of virtual worlds as learning space. *Innovations in Education and Teaching International*, *52*(3), 254–264. https://doi.org/10.1080/14703297.2013.820139

Van Gemert, T., Hornbæk, K., Knibbe, J., & Bergström, J. (2023). Towards a better future: A study of using virtual reality while lying down. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3544548.3580963

van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). Springer. https://doi.org/10.1007/978-1-4419-5546-3_10

Vatavu, R. D., & Mancas, M. (2015). Evaluating visual attention for multi-screen television: Measures, toolkit, and experimental findings. *Personal and Ubiquitous Computing*, *19*(5–6), 781–801. https://doi.org/10.1007/s00779-015-0862-z

Wang, L., & Xu, G. (2024). Self-explanation prompts in video learning: An optimization study. *Education and Information Technologies*, *29*(17), 23441–23462. https://doi.org/10.1007/s10639-024-12806-9

Wang, X., Mayer, R. E., Zhou, P., & Lin, L. (2021). Benefits of interactive graphic organizers in online learning: Evidence for generative learning theory. *Journal of Educational Psychology*, *113*(5), 1024–1037. https://doi.org/10.1037/edu0000606

Wang, Q. J., Ma, X. X., Lu, Y. N., Wang, D. M., & Sun, Y. H. (2024). The impact of visual and motor space size on gaze-based target selection. *Australian Journal of Psychology, 76*(1). https://doi.org/10.1080/00049530.2024.2309384

Wilf, M., Holmes, N. P., Schwartz, I., & Makin, T. R. (2013). Dissociating between object affordances and spatial compatibility effects using early response components. *Frontiers in Psychology*, *4*, 591. https://doi.org/10.3389/fpsyg.2013.00591

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Yang, C., Zhang, J., Hu, Y., Yang, X., Chen, M., Shan, M., & Li, L. (2024). The impact of virtual reality on practical skills for students in science and engineering education: A meta-analysis. *International Journal of STEM Education, 11*(1). https://doi.org/10.1186/s40594-024-00487-2

Yu, Q., Gou, J., Li, Y., Pi, Z., & Yang, J. (2024). Introducing support for learner control: Temporal and organizational cues in instructional videos. *British Journal of Educational Technology*, *55*(3), 933–956. https://doi.org/10.1111/bjet.13408

Yürüm, O. R., Yıldırım, S., & Taşkaya-Temizel, T. (2023). An intervention framework for developing interactive video lectures based on video clickstream behavior: A quasi-experimental evaluation. *Interactive Learning Environments*, *31*(10), 6611–6626. https://doi.org/10.1080/10494820.2022.2042312

Zhang, M., Hou, G., & Chen, Y. C. (2022). Effects of interface layout design on mobile learning efficiency: A comparison of interface layouts for mobile learning platform. *Library Hi Tech*, *41*(5), 1420–1435. https://doi.org/10.1108/LHT-12-2021-0431

Zhang, Y., Li, R., Pi, Z., & Yang, J. (2025a). Active learning strategies in video learning: A meta-analysis. *Educational Research Review, 48,* Article 100708. https://doi.org/10.1016/j.edurev.2025.100708

Zhang, Y., Xu, K., Pan, Y., Pi, Z., & Yang, J. (2025b). The effects of segmentation design and drawing on video learning: A mediation analysis. *Active Learning in Higher Education*, *26*(1), 3–23. https://doi.org/10.1177/14697874231180601

Zhang, Y., He, J., Pi, Z., & Yang, J. (2026a). Question generation type and organizational cues in video learning: A study on their effects on online learning behaviors. *Active Learning in Higher Education*, *27*(1), 83–103. https://doi.org/10.1177/14697874241286706

Zhang, Y., Pi, Z., Xu, K., Pan, Y., & Yang, J. (2026b). The interactive effects of drawing task, prompt format, and visual aids on video-based learning: Attention and drawing accuracy as mediators. *Active Learning in Higher Education*, *27*(1), 43–63. https://doi.org/10.1177/14697874241270536

Zhao, X., & McClure, C. D. (2024). Gather.Town: A gamification tool to promote engagement and establish online learning communities for language learners. *RELC Journal*, *55*(1), 240–245. https://doi.org/10.1177/00336882221097216

Zheng, H., Branch, R. M., Ding, L., Kim, D., Jung, E., Lu, Z., Li, T., Pan, Z., & Yoon, M. (2024). The combination of segmentation and self-explanation to enhance video-based learning. *Active Learning in Higher Education*, *25*(2), 285–302. https://doi.org/10.1177/14697874221126920

Zhou, Y., Ji, S., Xu, T., & Wang, Z. (2018). Promoting knowledge construction: A model for using virtual reality interaction to enhance learning. *Procedia Computer Science*, *130*, 239–246. https://doi.org/10.1016/j.procs.2018.04.035

Zhu, W., Wang, F., Mayer, R. E., & Liu, T. (2024). Effects of explaining a science lesson to others or to oneself: A cognitive neuroscience approach. *Learning and Instruction*, *91*, 101897. https://doi.org/10.1016/j.learninstruc.2024.101897

Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouwehand, K., & Paas, F. (2025). The synergy of embodied cognition and cognitive load theory for optimized learning. *Nature Human Behaviour*, *9*(5), 877–885. https://doi.org/10.1038/s41562-025-02152-2

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xiao Liu[1] · Yun Zhou[1] · Bingyu Dong[1] · Hui Zhang[1] · Lingzhi Cai[1] · Bin Li[2] · Tao Xu[3]

✉ Yun Zhou
zhouyun@snnu.edu.cn

✉ Tao Xu
xutao@nwpu.edu.cn

Xiao Liu
200462@snnu.edu.cn

Bingyu Dong
dongbingyu@snnu.edu.cn

Hui Zhang
z0915h@snnu.edu.cn

Lingzhi Cai
cailingzhi@snnu.edu.cn

Bin Li
binli2@cityu.edu.hk

[1] Faculty of Education, Shaanxi Normal University, South Chang'an Road 199, Yanta District, Xi'an, Shaanxi Province 710062, P.R. China

[2] Department of Linguistics and Translation, City University of Hong Kong, Tat Chee Ave., Kowloon Tong, Hong Kong, Hong Kong S.A.R.

[3] School of Software, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an, Shaanxi Province 710072, P.R. China