Computers & Education 198 (2023) 104764

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Promoting musical instrument learning in virtual reality environment: Effects of embodiment and visual cues

Shufan Yu [a,b,*], Qingtang Liu [a,b,**], Mina C. Johnson-Glenberg [c], Miaomiao Han [a,b], Jingjinag Ma [a,b], Shen Ba [d], Linjing Wu [a,b]

[a] School of Educational Information Technology, Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, 430079, Hubei, China
[b] Hubei Research Center for Educational Informationization, Central China Normal University, 430079, Hubei, China
[c] Department of Psychology, Arizona State University, Tempe, 85281, Arizona, USA
[d] Faculty of Education, The University of Hong Kong, 999077, Hong Kong, China

ARTICLE INFO

Keywords:
Augmented and virtual reality
Applications in subject areas
Human-computer interface
Media in education
Simulations

ABSTRACT

While virtual reality (VR) provides a great potential for musical instrument learning, little attention has been paid to the instructional design in creating a VR musical instrument. Previous research has suggested that high embodied interaction or added visual cues (e.g., distinctive colors, flashing areas) on VR-based musical instrument may aid students' learning. In this study, we investigated the feasibility and efficacy of the embodied design (low and high level of embodiment; LoEmb and HiEmb) and visual cues (low and high level of visual cues; LoViz and HiViz) on students' musical instrument learning. Four corresponding virtual Chinese dulcimers (*Yangqin*) were thereby designed. A sample of 112 university students participated in our study, and they were randomly assigned into the four conditions (LoEmb & LoViz, LoEmb & HiViz, HiEmb & LoViz, and HiEmb & HiViz). Results showed that the LoEmb design benefited students' completion rate, the HiViz improved students' playing rhythmic accuracy. Both LoEmb and HiViz decreased students' playing errors and improved their overall performance. Moreover, we found that the HiViz and HiEmb designs had a combined effect on reducing students' cognitive load and improving the instructional efficiency of learning material. These findings collectively reveal that the design of VR learning materials should carefully consider the trade-off between the level of embodiment and visual cues.

1. Introduction

Learning a musical instrument can facilitate a pupil's development of cognitive skills, academic performance, and time management skills (Hille & Schupp, 2015). However, several factors may lower the chance for students to access musical instruments, including the cost of purchasing and maintaining an instrument, a lack of guidance in learning, and a reduction of public funding on music education (Innocenti et al., 2019). Virtual reality (VR) technology offers a new alternative to musical learning and practicing by

* Corresponding author. School of Educational Information Technology, Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, 430079, China.
** Corresponding author. Hubei Research Center for Educational Informationization, Central China Normal University, 430079, Hubei, China.
E-mail addresses: yushufan1993@gmail.com (S. Yu), liuqtang@mail.ccnu.edu.cn (Q. Liu).

https://doi.org/10.1016/j.compedu.2023.104764
Received 15 July 2022; Received in revised form 17 February 2023; Accepted 19 February 2023
Available online 24 February 2023
0360-1315/

extending the music experiences beyond those offered by physical instruments (Serafin, Erkut, Kojs, Nilsson, & Nordahl, 2016; Turchet, Hamilton, & Camci, 2021). Recently, VR has been actively incorporated into the learning and experiencing of some musical instruments, including piano playing (Huang, Zhou, Yu, Wang, & Du, 2011), theremin practicing (Johnson, Damian, & Tzanetakis, 2020), drum performing (Kilteni, Bergstrom, & Slater, 2013), and music genres identification (Innocenti et al., 2019). While some studies have indicated the potential benefits of VR in musical instrument learning, the integration of instructional design or strategy on the design of a VR musical instrument (VRMI) has been rarely discussed. Literature has demonstrated that meaningful gestural and bodily interaction that aligned with the instructional material would benefit students' learning (de Koning & Tabbers, 2011; Johnson-Glenberg, Birchfield, Tolentino, & Koziupa, 2014; Johnson-Glenberg & Megowan-Romanowicz, 2017; Lindgren, Tscholl, Wang, & Johnson, 2016). The embodied design (Wilson, 2002) was subsequently considered by an increasing number of researchers, especially in the simulation-based interactive materials (Lindgren et al., 2016; Lui, McEwen, & Mullally, 2020). However, while researchers always emphasize the importance of designing high embodied interaction during practicing (de Koning & Tabbers, 2011; Johnson--Glenberg et al., 2014; Lindgren et al., 2016), the effects of embodiment on students' learning remain unclear due to a lack of studies and inconsistent results (Lindgren et al., 2016; Lui et al., 2020). Considering that operating a musical instrument involves various sensorimotor interactions, the level of embodiment was the first factor we were curious about.

Another factor that could influence learning in a VR environment is the representation of the content (Vogt, Babel, Hock, Baumann, & Seufert, 2021). Specifically, the visual cue (VizCue) on the key or strings in musical instrument systems was the major concern among researchers (Johnson et al., 2020); for instance, Johnson et al. (2020) augmented the theremin with VizCues that can provide students visual guidance and feedback for note positions. Since musical instruments entail different pitch arrangements on their strings, keys, or plane, the visual cues, as we assumed, would facilitate students' playing skills by following the notes. As depicted by Serafin et al. (2016), visual cues can help players perform better and improve their overall aesthetic experience (e.g., conversing the discrete keys into continuous musical lines and sequences). Nevertheless, whether VizCues can facilitate or affect students' musical instrument learning remains relatively unexplored. Therefore, the design of the visual cues is the second factor of interest to this paper.

In this study, we explore whether the embodied design and visual cues of VRMI could affect students' musical instrument learning. Four virtual Chinese dulcimers (i.e., yangqin) learning applications with different levels of embodiment (LE, high level of embodiment "HiEmb" & low level of embodiment "LoEmb") and levels of visual cues (LVC, high level of visual cues "HiViz" & low level of visual cues "LoViz") were implemented based on our previous work (Liu, Ba, Wu, Huang, & Li, 2018). Students' musical skill performance and cognitive load were examined to answer the question.

2. Literature review

2.1. Musical instrument learning and VR

To become proficient in a musical instrument, years of disciplined practice are necessary (Johnson et al., 2020). Enculturating in learners a fine preliminary impression and inspiring them to practice over time (Mazur & Łaguna, 2017) is particularly important. However, some factors, such as the expensive cost of the instrument, lack of feedback and personal guidance, and few chances to meet with professional teachers (Johnson et al., 2020), hinder students' willingness and ability to learn an instrument. Recently, the development of simulation technologies (e.g., augmented reality "AR" and VR) has provided potential solutions (Cui, 2022; Innocenti et al., 2019; Johnson et al., 2020; Qingtang et al., 2021), such as overlaying visual cues on physical instruments (Johnson et al., 2020), proving game-based environment (; Innocenti et al., 2019; Qingtang et al., 2021), and detecting hand posture mistakes (Johnson, Dufour, Damian, & Tzanetakis, 2016). Studies have indicated that simulated audio can elicit users' physiological reactions equivalent to those experienced in real life (Orman, 2004; Williamon, Aufegger, & Eiholzer, 2014). In addition, through digital technology, the musical performance and acoustics have seen the potential to couple closer (Çamcı & Granzow, 2019).

Unlike AR, which uses digital overlays on top of physical objects or markers (Azuma, 1997), VR can create an immersive learning environment in which users can perceive telepresence by interacting with multisensory resources in a fully digitized world (Burbules, 2006). Literature has shown the possibility of VR-Musical Instruction supporting and facilitating musical learning experiences (Innocenti et al., 2019; Johnson et al., 2020; Orman, Price, & Russell, 2017; Turchet et al., 2021). Some studies focused on the design and implementation of VRMI to enrich users' musical experience. For example, Serafin, Adjorlu, Nilsson, Thomsen, and Nordahl (2017) developed a VR drum simulator in which the VR joysticks resembled two sticks of a drummer, they contextualized the VR drum into a large stadium to overcome the "stage fear". From their observation, children around six years old can easily learn how to play it and report their satisfactory experience. Johnson et al. (2020) administered a virtual theremin system to university students. The result indicated that students were more engaged when performing immersive virtual theremin than in other non-immersive conditions. Likewise, in our previous study (Liu et al., 2018), a virtual Chinese dulcimer (yangqin) based on desktop VR was implemented by aligning the MIDI (a music technique standard) notes with dulcimer phonemes. In comparison to playing a physical yangqin, more than half of the participants considered that virtual yangqin learning had the potential to be more effective. VR technology can also foster students' musical appreciation ability. In Innocenti et al. (2019)'s study, VR technology was applied to involve students in a music genre identification activity (e.g., classical, country, jazz, and swing) with a mobile VR application. The results indicated that, with the help of VR technology, students could improve their musical experience in terms of active listening, attention, and time. Their study also confirmed the usability of VRMI for pupils with special needs. Moreover, an increasing number of studies have gradually focused on the interaction and presence of VRMI performers (e.g., Glowinski et al., 2015; Van Kerrebroeck, Caruso, & Maes, 2021). For instance, Van Kerrebroeck et al. (2021) utilized motion capture technology to facilitate two pianists' duet performance by mapping their motion to virtual avatars, the results showed that this real-time interaction could lead to performers' strong feeling of social presence.

However, while some studies confirmed the feasibility of including VR in music education, few explored the educational efficacy among students or compared it with other media. Moreover, music pedagogy and strategy have yet to see a breakthrough in the use of VR technologies (Johnson et al., 2020). To this end, integrating instructional designs on VRMI for learning and training should be considered (Serafin et al., 2016). Among the literature, embodied design (Kilteni et al., 2013) and visual cues (Johnson et al., 2020) piqued our interest.

2.2. Embodied design

The embodied theory of cognition posits that human cognition is deeply grounded in bodily interaction with the surrounding physical environment (Wilson, 2002). Specifically, the more meaningful and task-related body movements (e.g., gestures, postures) that are included, the more positive learning performance we would see (de Koning & Tabbers, 2011). To advance the study of embodiment in simulation, Johnson-Glenberg (2018) proposed an embodiment taxonomy by offering three constructs: the amount of sensorimotor engagement, gestural congruency, and sense of immersion, respectively. As such, researchers have begun to emphasize the importance of embodied design in VR research (e.g., Lindgren et al., 2016; Lui et al., 2020) and tried to examine the impact of these factors. For instance, Lindgren et al. (2016) examined two different embodied simulations of gravity and planetary motion; students were assigned to an experimental group supported by a simulation with whole-body interactions and a control group mediated by a desktop version. The results showed that the whole-body activity led to significant learning gains, higher levels of engagement, and more positive attitudes toward science. However, results are not always consistent. In Lui et al. (2020)'s work, when performing a gene regulation learning task, students in the seated VR (medium involvement of body) group performed better than those in the standing VR group (high involvement of body). Given that few studies with mixed findings related to the effect of LE, more empirical studies on the effects of embodied design with emerging technologies were needed, as suggested by Johnson-Glenberg (2018).

In musical instrument education, the virtual-real body alignment (congruency) would result in different learning attitudes and behaviors among students (Kilteni et al., 2013). For example, when designing a flute VRMI, as depicted by Serafin et al. (2016), introducing a breath-pressure sensor physical model that can be tracked could enhance the player experience by delivering haptic and visual feedback. In this sense, the high gestural congruency (interaction with a real instrument) should correlate with a high sense of embodiment (Johnson-Glenberg et al., 2014), and promote learning. Likewise, in Kilteni et al. (2013)'s study, the alignment of the virtual and physical body (full body ownership) led to significant changes in participants' body movement patterns while playing on a VR-based drum. While immersing and feeling like part of a virtual performer, the user can connect their body with their virtual performance (Atherton & Wang, 2020), thereby effortlessly increasing their musical skill. In this regard, researchers were increasingly taking into account the embodied design (e.g., bodily ownership and movement) on VRMI (Turchet et al., 2021). However, although there have been some significant advancements in VRMI, the embodied interaction design and strategy varied greatly among research and context. Therefore, based on Johnson-Glenberg (2018)'s taxonomy, we attempted to investigate what level of embodied interaction is needed when playing a virtual instrument.

2.3. Visual cues

During playing a VRMI, the presence of a virtual body representation is also important to provide the necessary visual feedback to the player (Atherton & Wang, 2020; Serafin et al., 2016; Turchet et al., 2021), which shed light on the other factor, cues that we were curious about.

In multimedia learning, VizCues (or signals) can be arrows, distinctive colors, or flashing areas that highlight the critical components of the learning material, thereby guiding students' attention (Mayer, 2002). According to Mayer (2002)'s signaling principle, learners would learn better when adding VizCues to the learning materials. He believed the VizCues could act as a cognitive guide during learning. However, research on VizCues still produced mixed findings. On the one hand, some studies reported a positive effect of VizCues on learning performance (Jamet, 2014; Yang, Su, Xu, & Hu, 2022). On the other hand, some studies indicated the limited effect of VizCues on learning (Vogt, Babel, et al., 2021). While a previous meta-analysis of 29 studies (Alpizar, Adesope, & Wong, 2020) synthesized a medium effect of VizCues on learning outcome ($d = 0.38$), the subject domain was not considered in their work, and the effect on musical learning was still unclear. Given that learners can look anywhere in a 360-degree environment, we believe that VizCues is even more important in VR to help students maintain focus on the instruments.

In musical education, as depicted Johnson et al. (2020), integrating relevant immersive visuals into a digital musical instrument can result in a new way of musical experience. In this sense, some previous research tried to apply VizCues to the design of VRMI, such as overlaying the VizCues on piano keys (Huang et al., 2011) and presenting VizCues upon the theremin (Johnson et al., 2020). Moreover, Johnson et al. (2020) tried to examine the effect of adding VizCues on the virtual theremin, but the result was not favorable for VizCue's benefits on learning performance. However, the studies on the impact of visual cues on students were still insufficient to draw any robust conclusion.

2.4. Cognitive load and instructional efficiency

The cognitive load is regarded as a multidimensional construct used to describe a learning process wherein students utilize their limited cognitive resources to process the task complexity and element interactivity of instructional materials (Paas, Tuovinen, van Merriënboer, & Aubteen Darabi, 2005; Sweller, 2010). If the cognitive resources failed to meet the learning requirement, a high cognitive load would be generated, thereby impeding students' learning success (Sweller, 2010). In VR learning environment, learners are frequently confronted with various representations, including text, images, and 3D objects in 360°; their cognitive learning process would be influenced by this media-rich stereoscopic world (Vogt, Albus, & Seufert, 2021). On the one hand, students would report more load in VR studies for the sounds, simulations, and interactions that were irrelevant to learning (Parong & Mayer, 2021), and the cumbersome manipulation tasks as well as the unfamiliarity with the technology (Makransky, Terkildsen, & Mayer, 2019; Parong & Mayer, 2021). On the other hand, they may also experience decreased cognitive load since they do not have to worry about the unforeseen issues that may arise in the physical world (Petersen, Klingenberg, Mayer, & Makransky, 2020) and can increase their attention to the digital world because the physical world is blocked out. In this regard, assessing and adjusting students' cognitive load is a necessary task for VR researchers (Makransky et al., 2019; Petersen et al., 2020; Yu, 2021). The great diversity of immersion, gesture interactions, and cognitive guidance would generate different levels of task complexity. Therefore, we were curious about what design configuration of embodiment and VizCues would have a better impact on cognitive load.

Furthermore, Paas and Van Merriënboer (1993) argued that the interpretation of cognitive load would be more meaningful when discussing learner performance. They proposed a mental effort—performance coordinate system and a formula to represent the instructional efficiency of students on cognitive tasks. Such equation may be useful in predicting which configurations would maximize performance efficiency and enhancing the research on the effectiveness of complex cognitive activities (Ba, Stein, Liu, Long, Xie, Wu, 2021; Paas & Van Merriënboer, 1993; Schroeder et al., 2017). Therefore, to reach an overall view on students' learning performance and cognitive load, instructional efficiency was also of interest to this study.

2.5. Present study

Yangqin is a traditional Chinese stringed instrument performed with a hammer (see Fig. 1). As a popular folk instrument, yangqin instruction has always drawn the interest of researchers (Tse, 2007). Yangqin shares similar physical features and tuning as other stringed instruments (Tse, 2007). Learning yangqin well can enhance students' overall musical ability and lead to improved performance on other stringed musical instruments. However, traditional yangqin teaching (i.e., face-to-face) has encountered several obstacles that could potentially impede students' learning, such as the high cost of the instrument, the limited opportunities for personalized instruction, and the lack of teachers with greater musical instrument knowledge and yangqin playing experience (Liu et al., 2018; Tse, 2007; Zhang, 2020). Moreover, yangqin has several complicated pitch areas arranged by 144 strings which makes it difficult for a beginner to hit the right note. Therefore, we are interested in providing a VR-based application to facilitate students' yangqin learning.

This study aims to investigate the effect of embodied design (i.e., HiEmb & LoEmb) and visual cues (i.e., LoViz & HiViz) on students' musical instrument learning. Specifically, students were required to learn the basic playing skills of yangqin via four versions of pre-designed VRMIs, which were upgraded based on our previous work (Liu et al., 2018). We hope to provide some general implications by looking through the lens of the musical instrument learning domain.

Given the mixed results of previous research on the effects of LE and LVC on learning outcomes, and the research gap on the cognitive load and instructional efficiency of VR studies. We generated the following research questions (RQ).

RQ 1. How does learners' yangqin playing skill vary in VR environment with different levels of embodiment and VizCues?

RQ 2. How does learners' cognitive load vary in a VR environment with different levels of embodiment and VizCues?

RQ 3. How does the instructional efficiency of virtual yangqin vary among students when designed with different levels of embodiment and VizCues?

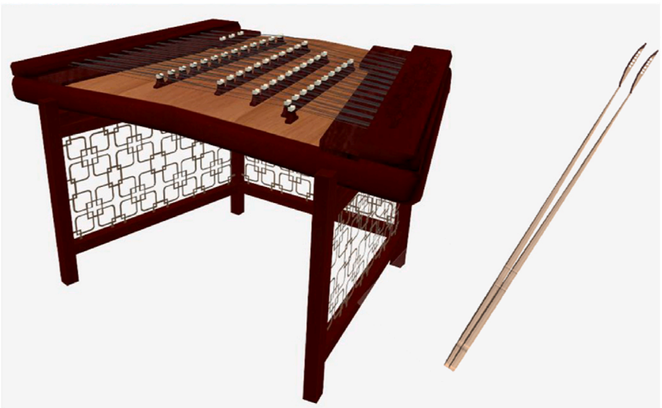

Fig. 1. The Chinese dulcimer (yangqin) and the bamboo hammers.

3. Methodology

3.1. Participants

This study recruited 120 undergraduate and graduate students from a Chinese university. Eight were removed from the final analyses because they did not finish all the learning tasks or their rhythm score could not be generally evaluated for some unexpected reasons (e.g., the screen was not recorded or was muted). There were 112 participants (age range from 18 to 25) finally included in the data analysis. All participants were in education or computer-related majors, and they did not have any prior experience with yangqin before treatment.

3.2. Research design and procedure

A 2 × 2 factorial design was implemented in this study to investigate the effects of embodied design (i.e., LE) and LVC. The first factor LE entailed two levels, in LoEmb they used a desktop-based yangqin simulation, and in HiEmb they used immersive VR version. The second factor LVC had two levels as well, in the LoViz they saw VizCues aligned with some strings, and in HiViz they saw VizCues followed the musical notes. Students were randomly assigned to these four conditions and did not know which one they were in. The overall procedure is shown in Fig. 2.

Before the intervention, students completed a demographic survey and a Likert scale question on previous VR experience (On a scale of 1–5, to what extent do you think you are familiar with VR?). Then, they were randomly assigned into the four groups. Afterward, participants received pre-training individually to acquaint themselves with the basic VR operations of different learning environments in our study.

During the experiment, each participant was allocated a virtual yangqin based on their condition. At the beginning of the learning module, participants were introduced to fundamental music theory and yangqin playing skills via the corresponding applications. Then, they were taught the fundamentals of the yangqin's string and pitch arrangements using the song "Jasmine" (a well-known Chinese folk ballad) and its musical score. After the training session, we evaluated students' musical playing skills by having them perform the song "Two tigers" (another popular Chinese song). During the learning phase, students received evaluative scores and immediate feedback on their playing performance. After completing the learning modules, all students were asked to fill out a

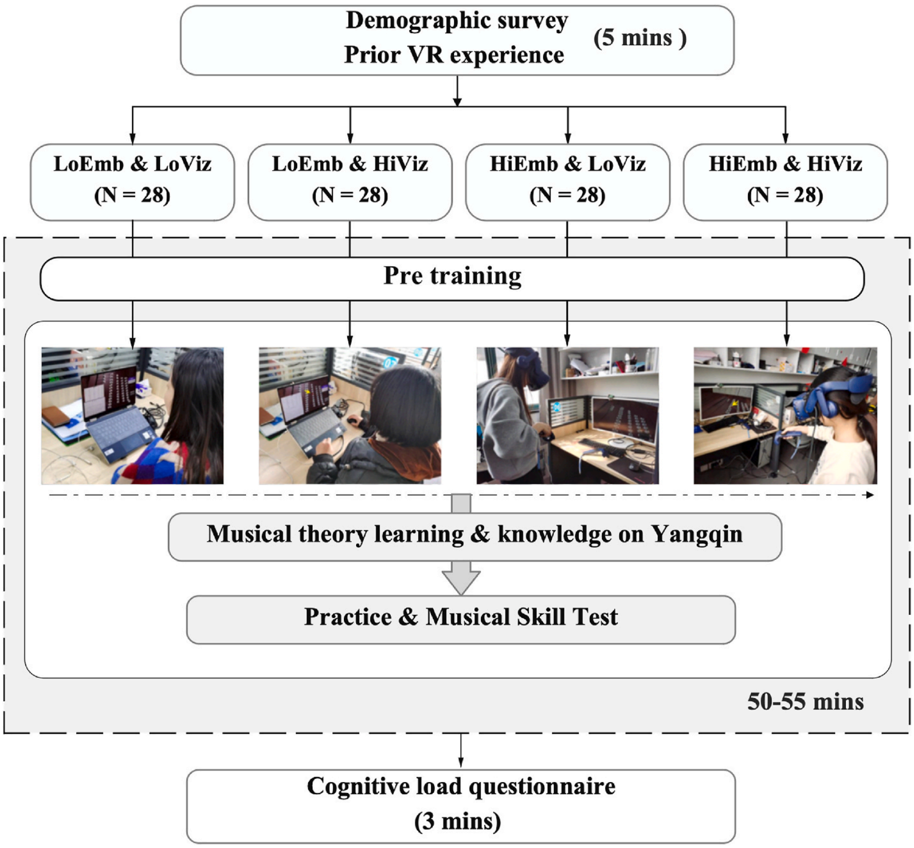

Fig. 2. Experimental procedure.

cognitive load questionnaire. Note that we purposefully chose these two songs because they are two of the most popular folk songs in China, and all participants in our study learned them during their k-12 music education. In addition, the notes of the two songs were relatively easy to understand by a beginner, as suggested by a music expert. Moreover, each student had 5 min to familiarize themselves with the song "two tigers" before the final test, and they could choose to stop at any time within the time limit. In this regard, the whole learning process took up 50–55 min, depending on the learners' practicing time.

In addition, after the entire learning process, a brief interview was randomly conducted with 3 students in each condition by asking, "How was your experience during learning the virtual yangqin?" Each of them was given 5 min to express their opinion. The interview results were only used as supplemental data, as our research questions may be almost entirely answered and explained by quantitative data. In this regard, we did not include it as one part of the main process in Fig. 2.

3.3. VR-based virtual yangqin

This study proposed four different virtual yangqins varied in LE and LVC. After discussing with a music expert, we designed the virtual yangqin learning application with three modules: learning, training, and testing (see Fig. 3: I, II, III) according to the standard learning process in musical education.

When playing a physical yangqin, hammers made of bamboo were held in each hand, and used to hit the strings to produce sound. Therefore, in terms of the design of the embodiment, we provided a desktop version (see Fig. 3(c) and (d)) in which students use a mouse to interact with strings by clicking buttons and an immersive VR (IVR) version (HTC VIVE Pro 2) (see Fig. 3(a) and (b)) wherein students can use a VR hand controller to directly control the hammer to hit the string. Specifically, the VR hand controller model will automatically switch to the virtual hammers when students interreacted with the virtual yangqin (they cannot see their hands or bodies in VR environment). In this regard, the gestural control of virtual hammers via controllers in IVR condition was aligned to the operation in physical yangqin, while the mouse-system interaction in the desktop version tallied with a low gestural congruency. Therefore, based on Johnson-Glenberg (2018)'s taxonomy, we assigned the desktop VR as LoEmb version (low gestural congruency and high immersion) and immersive VR as HiEmb version (high gestural congruency and high immersion).

As for the system per se, in the learning module, two video tutorials were provided to introduce the basic musical theory and knowledge of the yangqin, including the arrangement of pitches, the basic playing skill (e.g., how to tap the strings) of yangqin, and an example of how to play it. Moreover, a three-dimensional yangqin model as well as some pictorial and aural information were also included to impart the structure of yangqin to students. When students stepped into the training session, the song "Jasmine" was presented to train their playing skills. In this module, based on Mayer (2002)'s signaling principle and enlightened by Huang et al. (2011)'s VizCue on piano keys, the design of LVC was highlighted (see Fig. 3). Specifically, LoViz version only marked the VizCues on some key strings in different pitch areas of yangqin (See Fig. 3(1)). In contrast, HiViz version (See Fig. 3(2)) also highlights the **next** string that needed to be hit according to the notes on a musical score. Therefore, students can learn the playing skill following the time-sequenced prompts on strings.

Regarding the final test, the song "Two tigers" was provided to test their yangqin playing skills. In this context, all versions only included the configuration of LoViz to test their transfer skill. The system will collect and assess students' musical skill performance, as

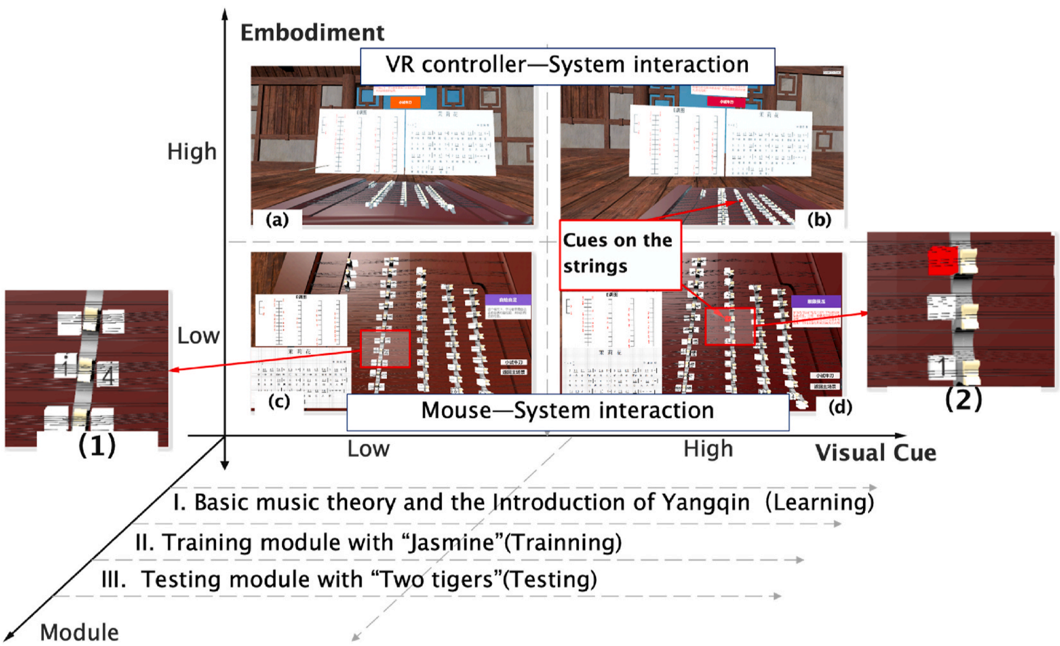

Fig. 3. The different versions of virtual yangqin and their learning module.

detailed in section 3.4.1.

Considering students were all new to yangqin, both songs in the training and testing module were in monophony so that students could easily understand when their first encounter with this instrument.

3.4. Measures

3.4.1. Musical skill performance (MSP) test

Students' musical skill on yangqin was evaluated by the system along three constructs which were 1) Completion rate, 2) Error rate, and 3) Rhythm.

1. ***Completion rate (CR)*** refers to the percentage of notes correctly hit by a participant. This was used to examine if they could finish the song (i.e., two tigers) within a specified time frame (i.e., 35 s).
2. ***Error rate (ER)*** means the ratio of wrong trials during the test. It was reflected by the percentage of wrong hits out of the total number of hits (the sum of the wrong and correct hits of a participant).
3. ***Rhythm score (RS)***. During the experiment, an interesting phenomenon was observed: although some students can correctly play the musical notes, they may make some intermittent mistakes, which may cause their songs to sound "unpleasant". In this regard, the students' RS was also considered; it was utilized to identify how well a rhythmic pattern they performed matched the actual pattern (Cui, 2022). Specifically, if the notes are either ahead of or below the beat, their RS will be reduced. RS was evaluated by a musical expert who had a 5-year professional musical experience. The total score of RS was confined to 32 points. Each student's screen was recorded so that their RS could be checked and evaluated by the expert and researchers.

For instance, in this study, the song "Two tigers" used to test students' musical skills has 32 notes. Student A finished 30 notes out of 32 notes within 35 s, and he made 10 errors during the test. In this context, the Completion rate will be 30/32 = 93.75%, the Error rate will be 10/(30 + 10) = 25%, and the rhythm score (e.g., 25) will be evaluated by a music expert according to students' rhythmic performance.

3.4.2. Cognitive load scale

The two-item cognitive load scale, which is modified based on the ratings of Cheng (2017) and the measure of Paas and Van Merriënboer (1993), was applied to assess students' mental load (ML) and mental effort (ME). ML indicates students' perceived task difficulty, while ME refers to the amount of effort they devoted to comprehending the learning content and organizing learning materials. The scale was in a seven-point Likert rating, and a higher score means a higher cognitive load. The Cronbach's α for the scale is 0.66, indicating acceptable internal consistency of the scale.

3.4.3. Data analysis

This experiment was a 2 × 2 factorial design in which the first factor under investigation was LE (LoEmb & HiEmb), and the second factor was LVC (LoViz & HiViz). The dependent variables were students' musical skill performance (MSP) test scores (i.e., CR, ER, and RS), cognitive load (CL), and the instructional efficiency (IE) of virtual yangqin.

Due to the growing emphasis on the control of prior experience and familiarity in virtual environments (e.g., Liu, Yu, Chen, Wang, & Xu, 2021; Makransky & Petersen, 2021), we considered the prior VR experience as a covariate in the analysis of MSP. Specifically, a two-way analysis of covariance (ANCOVA) was applied to investigate the effect of LE and LVC on MSP by using prior VR experience as the covariate. Moreover, A two-way analysis of variance (ANOVA) was conducted to analyze the remaining factors (CL and IE).

Regarding students' MSP, except for the separated analyses on each dimension, we proposed a calculation method to consider the combined evaluation of the three dimensions following the advice of a musical expert. This Combined Performance (CP) was calculated by equation (1).

$$\mathrm{CP} = T_{score}\left(\frac{\mathrm{Completion\_Num} - \mathrm{Error\_Num}}{\mathrm{Completion\_Num} + \mathrm{Error\_Num}}\right) + T_{score}(RS) \tag{1}$$

In this equation, the former part indicates the accuracy of pitch (Mazur & Łaguna, 2017). Specifically, we subtracted the number of errors from the completion number and divided the result by the total trials. The final result means the ratio of positive trials to the total trials.

However, some unexpected cases were also observed in the current study. For example, some students' completion number is much lower than the error number, which may lead to a negative result. Therefore, to make the result easily understood and interpreted, we transformed the score into T-score (T = 50 + 10Z), in which Z represents the normal standard scores (Z-scores). As for the latter term, to reveal the precision of rhythm (Mazur & Łaguna, 2017), we also utilized the RS's T-score. As such, the two terms were transferred to the same range of measurement, which makes them comparable. According to the musical expert, pitch accuracy shares the same importance as precision in rhythm. We therefore directly leveraged their sum to represent the finalized CP and included it in the statistical analysis.

Furthermore, the IE was also considered in this study; we applied the IE calculation method proposed by Paas and Van Merriënboer (1993) to compare the efficiency of different virtual yangqins. In this formula (see equation (2)), *P* represents the z-score of learning performance, and *R* represents the z-score of ME. Specifically, we use the CP score as students' performance, then include the standardized CP and ME in the equation as P and R to get the final IE.

$$IE = \frac{P - R}{\sqrt{2}} \tag{2}$$

Finally, for a comprehensive interpretation, the p value was used to assess the significance of each factor, and the partial η^2 was included to reflect their effect size.

4. Results

4.1. Musical skill performance (MSP)

The MSP consists of CR, ER, and RS, which will be considered in the following analysis. Fig. 4 presents the descriptive statistics of each dimension.

4.1.1. Completion rate (CR), error rate (ER), and rhythm score (RS)

After verifying the equality of error variances (CR: $p = .34$, ER: $p = .078$, RS: $p = .952$), a two-way analysis of covariance (ANCOVA) was conducted, by using LE and LVC as independent variables, prior VR experience as covariates. Moreover, The CR, ER, and RS were separately served as dependent variables. Table 1 shows the results of the two-way ANCOVA.

For the CR dimension, there was only one statistically significant main effect for LE (F (1, 107) = 4.175, $p < .05$, $\eta^2 = 0.038$). Specifically, students who played with LoEmb digital virtual yangqin (Adjusted mean = 73.1%, SE = 0.037) finished significantly more on the tasks than those in HiEmb group did (Adjusted mean = 62.2%, SE = 0.037).

Regarding the effect on ER, two main effects were found for LE (F (1, 107) = 5.363, p < .05, $\eta^2 = 0.048$) and LVC (F (1, 107) = 4.061, $p < .05$, $\eta^2 = 0.037$), while no significant interaction was presented. According to the results, learners in high embodied condition (Adjusted mean = 24.1%, SE = 0.029) made significantly more mistakes than LoEmb group (Adjusted mean = 14.3%, SE = 0.029), and the VizCue on strings did help students significantly decrease their mistakes during playing process (HiViz group: Adjusted mean = 15%, SE = 0.029; LoViz group: Adjusted mean = 23.4%, SE = 0.029).

With respect to the RS, we found the LVC significantly impact it (F (1, 107) = 5.856, $p < .05$, $\eta^2 = 0.052$). For those who played HiViz virtual yangqin (Adjusted mean = 22.72 SE = 0.999), they performed with a significantly better rhythm than those in LoViz group (Adjusted mean = 19.322, SE = 0.99).

4.1.2. Combined performance (CP)

Levene's test on the four groups' CP scores shows the equality of variances with $p = .110$. A two-way ANCOVA was subsequently conducted to compare the CP of students in different groups by using their prior VR performance as covariates. Again, we only found a

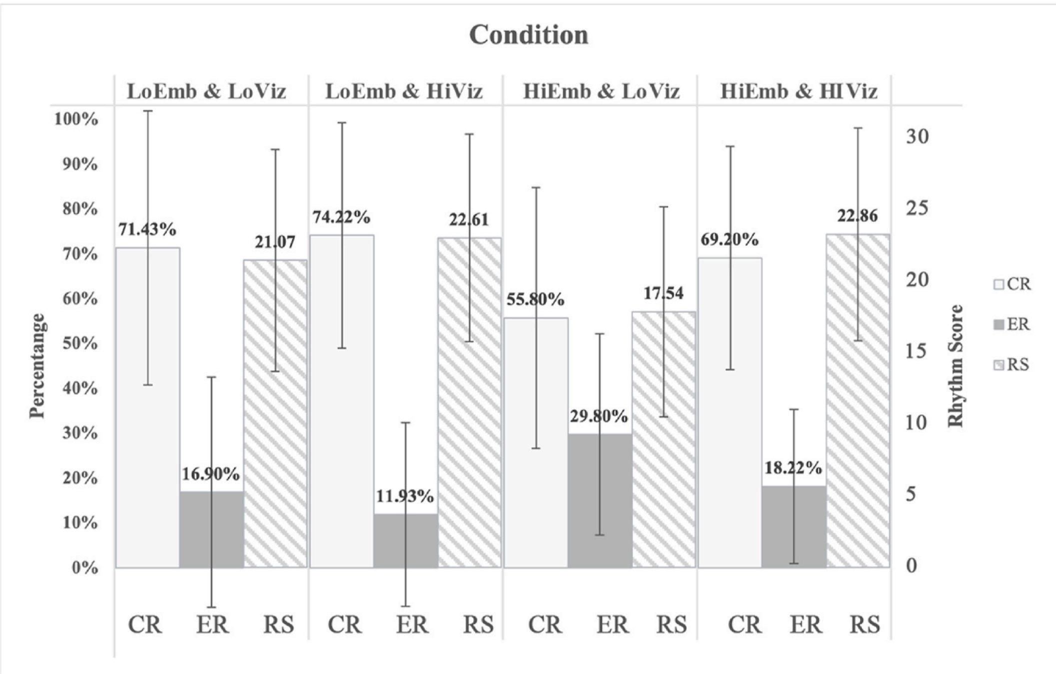

Fig. 4. Mean and SD (whiskers) of CR, ER, and RS value in each condition (CR: completion rate, ER: error rate, RS: rhythm score).

Table 1
The two-way ANCOVA result of the CR, ER, and RS.

Source	DV	SS	*df*	MS	*F*	η^2
Covariance	CR	0.021	1	0.021	0.277	0.003
	ER	0.001	1	0.001	0.030	0.000
	RS	1.105	1	1.105	0.020	0.000
LE	CR	0.319	1	0.319	4.175**	0.038
	ER	0.254	1	0.254	5.363**	0.048
	RS	68.331	1	68.331	1.265	0.012
LVC	CR	0.198	1	0.198	2.592	0.024
	ER	0.192	1	0.192	4.061**	0.037
	RS	316.444	1	316.444	5.856**	0.052
LE * LVC	CR	0.086	1	0.086	1.125	0.010
	ER	0.032	1	0.032	0.666	0.006
	RS	97.393	1	97.393	1.802	0.017
Error	CR	8.174	107	0.076		
	ER	5.068	107	0.047		
	RS	5781.824	107	54.036		

Note. $^{**}p < .05$.

significant effect for the two main factors (LE: $F\,(1, 107) = 4.156$, $p < .05$, $\eta^2 = 0.037$; LVC: $F\,(1, 107) = 6.917$, $p < .05$, $\eta^2 = 0.061$) rather than the interaction term ($F\,(1, 107) = 1.640$, $p = .203$) (See Fig. 5). Specifically, students in low embodied condition (Adjusted mean = 103.24, $SE = 2.22$) significantly outperformed their high embodied counterparts (Adjusted mean = 96.757, $SE = 2.22$) in terms of the overall performance. Meanwhile, when students played with virtual yangqin with HiViz (Adjusted mean = 104.132, $SE = 2.21$), they achieved significantly higher combined scores than those in LoViz groups (Adjusted mean = 95.87, $SE = 2.21$).

4.2. Cognitive load (CL)

To compare students' CL when they played virtual yangqin with different levels of embodiment and visual cues, a two-way ANOVA was conducted on both ME and ML respectively. Before that, we tested the ANOVA hypothesis. Levene's test verified the equality of the dependent variable's error variance with a p value of .337 on ME dimension and 0.631 on ML dimension. Therefore, it is sensible to conduct the following between-subjects tests.

As shown in Table 2, according to the results of ANOVA, we found a significant main effect of LVC on ME ($F\,(1, 108) = 5.070$, $p < .05$, $\eta^2 = 0.045$) and a marginally significant interaction effect of LE and LVC on ME ($F\,(1, 108) = 2.965$, $p < .1$, $\eta^2 = 0.027$). As such, we conducted a simple main effect analysis. See Table 3 and Fig. 6, it was found that, when treated by HiViz, students in low level of embodiment (LoEmb) condition (Mean = 3.43) invested higher ME than in HiEmb group (Mean = 2.79) with a marginally significant effect ($F\,(1, 108) = 2.842$, $p < .1$, $\eta^2 = 0.026$). Moreover, in HiEmb group, students presented significant less ME ($F\,(1, 108) = 7.895$, $p < .05$, $\eta^2 = 0.068$) when played virtual yangqin with HiViz (Mean = 2.79) than with LoViz (Mean = 3.86) (see Fig. 7).

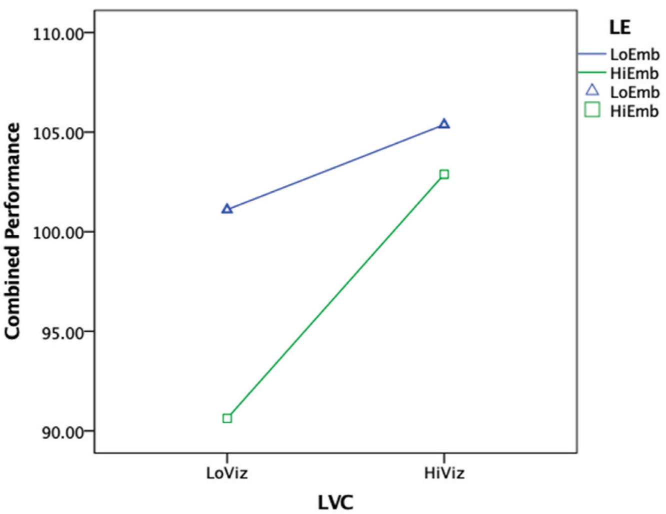

Fig. 5. The main effect of LVC and LE on CP (LE: level of embodiment, LoEmb: low level of embodiment, HiEmb: high level of embodiment; LVC: level of visual cue, LoViz: low level of visual cue, HiViz: high level of visual cue).

Table 2
The two-way ANOVA result of ME and ML.

Source	DV	SS	*df*	MS	*F*	η^2
LE	ME	0.893	1	0.893	0.439	0.004
	ML	2.893	1	2.893	1.264	0.012
LVC	ME	10.321	1	10.321	5.070**	0.045
	ML	2.286	1	2.286	0.999	0.009
LE * LVC	ME	6.036	1	6.036	2.965*	0.027
	ML	1.286	1	1.286	0.562	0.005
Error	ME	219.857	108	2.036		
	ML	247.214	108	2.289		

*Note.**$p < .1$; **$p < .05$.

Table 3
Simple main effects analysis results of LE and LVC on ME.

Sources	SS	*df*	MS	*F*	η^2	Comparison
LE						
LoViz	1.143	1	1.143	0.561	0.05	
HiViz	5.786	1	5.786	2.842*	0.026	LoEmb > HiEmb
LVC						
LoEmb	0.286	1	0.286	0.140	0.001	
HiEmb	16.071	1	16.071	7.895**	0.068	LoViz > HiViz

Note. *$p < .1$; **$p < .05$.

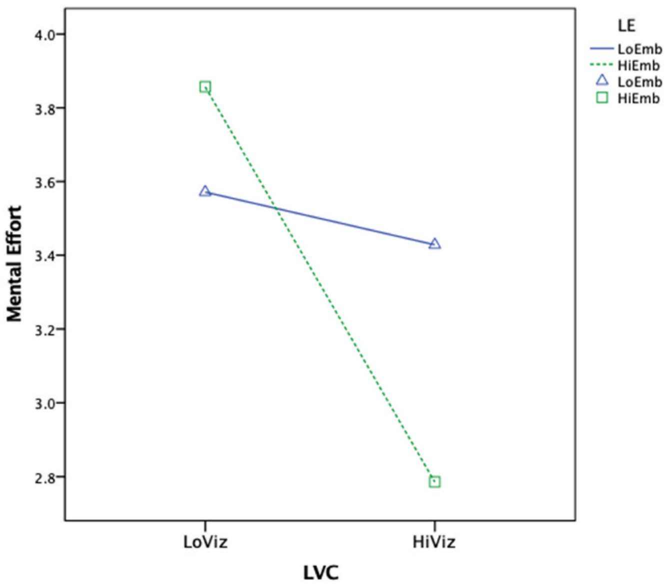

Fig. 6. The interaction effect of LE and LVC on ME (LE: level of embodiment, LoEmb: low level of embodiment, HiEmb: high level of embodiment; LVC: level of visual cue, LoViz: low level of visual cue, HiViz: high level of visual cue).

4.3. *Instructional efficiency (IE)*

After confirming the equality of error variance ($p = .764$), a two-way ANOVA was performed to compare the students' perceived IE in different conditions. The result revealed a marginally significant interaction effect of LE and LVC on IE (F (1, 108) = 3.576, $p < .1$, $\eta^2 = 0.032$) (See Fig. 7). As such, we further conducted a simple main effect test.

See Table 4, it is found that a marginally significant effect of LE on LoViz condition (F (1, 108) = 3.838, $p < .1$, $\eta^2 = 0.034$) and a significant effect of LVC on HiEmb condition (F (1, 108) = 12.363, $p < .05$, $\eta^2 = 0.103$). To be specific, when students learned virtual yangqin with LoViz, the low embodied yangqins (Mean = −0.032, SE = 0.206) were reported significantly higher IE than those with HiEmb (Mean = −0.603, SE = 0.206). Moreover, in HiEmb conditions, the virtual yangqin with HiViz (Mean = 0.422, SE = 0.206) had significantly higher IE than its LoViz counterpart (Mean = −0.603, SE = 0.206).

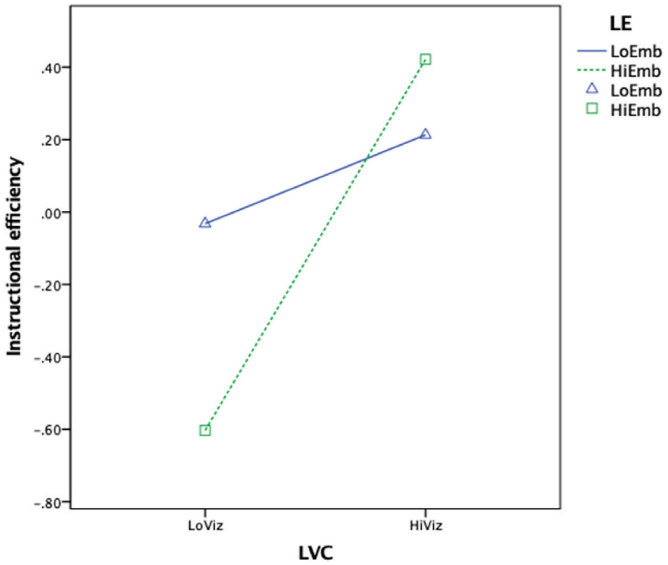

Fig. 7. The interaction effect of LE and LVC on IE (LE: level of embodiment, LoEmb: low level of embodiment, HiEmb: high level of embodiment; LVC: level of visual cue, LoViz: low level of visual cue, HiViz: high level of visual cue).

Table 4
Simple main effects analysis results of LE and LVC on IE.

Sources	SS	*df*	MS	*F*	η^2	Comparison
LE						
LoViz	4.564	1	4.564	3.838*	0.034	LoEmb > HiEmb
HiViz	0.608	1	0.608	0.511	0.005	
LVC						
LoEmb	0.843	1	0.843	0.709	0.007	
HiEmb	14.703	1	14.703	12.363**	0.103	LoViz < HiViz

Note. *$p < .1$; **$p < .05$.

5. Discussion

This study addressed the potential of embodied design and adding visual cues during learning with a VR musical instrument (VRMI). Participants were asked to learn through our proposed virtual yangqin with different levels of embodiment (LoEmb & HiEmb) and different levels of visual cues (LoViz & HiViz), and their MSP, CL, and IE were explored. The overall results are shown in Table 5. The findings are discussed below.

5.1. Impact of LE and LVC on MSP: LoEmb and HiViz matter

To address RQ1, we examined students' MSP both on the three specific categories (i.e., CR, ER, and RS) and their combined performance (CP). Generally, we found that the main effect of LE on CR (LoEmb > HiEmb) and LVC on RS (HiViz > LoViz), while the ER and CP were jointly benefited by the design of LoEmb and HiViz. Three findings can be drawn. First, the low embodied design could facilitate students' performing efficiency (higher CR) on VRMI regardless of the LVC. This finding was opposite to the main propositions of the theory of embodied cognition (Wilson, 2002), which emphasizes the importance of bodily interaction and gestural interactions that are aligned to learning tasks (de Koning & Tabbers, 2011; Johnson-Glenberg et al., 2014). However, this phenomenon can also be reflected by some qualitative evidence. According to the interview results, we found that the result would ascribe to the difference of arrangement on the musical scores; that is, LoEmb condition align the musical score with the virtual yangqin, while in HiEmb context, the notes were put in front of students. Although this would be more in line with playing a physical yangqin, it did distract beginners' attention due to the additional "string-score switch" movement. This also conforms to Mayer (2002)'s contiguity principle, which posits spatial contiguous elements (e.g., words and pictures) would benefit students' learning by reducing their extraneous processing on the difficulty of the instructional materials.

Second, the time-sequenced VizCues on strings could facilitate students playing rhythm (HiViz > LoViz). This finding not only reaffirmed that visual cues could guide a learners' attention on the learning materials per se (Alpizar et al., 2020; Jamet, 2014), it also indicated that VizCues can further increase students' sound perceptual ability (i.e., rhythmic accuracy) (Mazur & Łaguna, 2017).

Table 5
Summary of the results (MSP: musical skill performance, CR: completion rate, ER: error rate, RS: rhythm score, CL: cognitive load, ME: mental effort, ML: mental load, IE: instructional efficiency).

Groups			LoEmb		HiEmb	
			LoViz	HiViz	LoViz	HiViz
N			28	28	28	28
MSP	CR	Mean	71.4%	74.2%	55.8%	69.2%
		SD	0.31	0.25	0.29	0.25
		Comparison	LoEmb > HiEmb			
	ER	Mean	16.9%	11.9%	29.8%	18.2%
		SD	0.26	0.20	0.22	0.17
		Comparison	LoEmb < HiEmb, LoViz > HiViz			
	RS	Mean	21.07	22.61	17.54	22.86
		SD	7.63	7.12	7.21	7.31
		Comparison	HiViz > LoViz			
	CP	Mean	101.11	105.37	90.64	102.89
		SD	19.66	14.02	15.66	15.59
		Comparison	LoEmb > HiEmb HiViz > LoViz			
CL	ME	Mean	3.57	3.43	3.86	2.79
		SD	1.35	1.29	1.60	1.45
		Comparison	HiViz & LoEmb > HiViz & HiEmb HiEmb & HiViz < HiEmb & LoViz			
	ML	Mean	3.00	2.93	2.89	2.39
		SD	1.52	1.63	1.55	1.34
		Comparison	\			
IE		Mean	−0.03	0.213	−0.60	0.42
		SD	0.21	0.21	0.21	0.21
		Comparison	LoViz & LoEmb > LoViz & HiEmb HiEmb & HiViz > HiEmb & LoViz			

Meanwhile, it reaffirmed Mayer's (2002) signaling principle, which advocates adding VizCues to instructional materials. Finally, as for the ER and CP, the results revealed that they were conjunctionally positively influenced by the LoEmb and HiViz design on VR platform; specifically, under such configuration, students would make less mistakes (less ER) and perform better on the overall level skill of yangqin (higher CP). These findings again highlighted the importance of guiding attention via VizCues (Alpizar et al., 2020; Jamet, 2014) and revealed another "platform is not destiny" (Johnson-Glenberg, Bartolomea, & Kalina, 2021) in which embodied interaction should be cautiously designed in synergy with other elements (e.g., the VizCues in the current study).

5.2. Impact of LE and LVC on CL: VizCues help reduce ME by mediating the effect of LE

Participants' ML and ME were assessed to answer RQ2. We found that regardless of which condition they were in, learners all reported a low ML (<3), which indicated that the interactive elements were organized well on all the four VRMIs, thereby could reduce students perceived task difficulty (Paas & Van Merriënboer, 1994). Furthermore, we found that students who used immersive VR headset (HiEmb) without time-sequenced VizCues on its strings invested the most psychological effort (high ME) to process the task complexity. One possible reason for this is that when treated with immersive VR, students may feel it is hard to manipulate the 3D resources in a wide range of visual filed when compared with some 2D media from a monitor, thereby increasing their cognitive load (Makransky et al., 2019; Makransky & Petersen, 2021). Although the HiEmb condition showed a negative result, we found that this negative effect could be mediated by adding chronological VizCues on VRMI. Students would invest the least ME among the four conditions when HiEmb dulcimer was integrated with HiViz. This finding affirmed that adding VizCues could optimize VR design by decreasing students' extraneous cognitive load, and echoed with previous studies (Jamet, 2014; Yang et al., 2022).

5.3. Impact of LE and LVC on IE: cautiously adding VizCues on the VR platform

In terms of IE results (RQ3), we found that when designing virtual yangqin with LoViz, the desktop version (LoEmb) presented a significantly higher IE than the immersive VR version (HiEmb). Moreover, the immersive VR yangqin (HiEmb) could benefit students learning for its higher IE when designed with HiViz. Once again, the importance of design on LVC emerged, but this conclusion should be drawn carefully with some boundary conditions. First, the design of VizCues did not influence the IE of the desktop (LoEmb) version, which means the limited embodied interactions are not needed to entangle with a high level of visual cues. Second, the HiViz is more suitable for design on an IVR (HiEmb) platform since it can promote the IE of virtual yangqin with HiEmb. By reviewing students' recorded screens, we found that in HiEmb with LoViz condition, students spent more time shifting sights from strings to score, which may subsequently influence their MSP, thereby degrading the IE of virtual yangqin. However, adding time-sequenced cues on strings (HiViz) helped them focus on the learning tasks and promoted their efficiency in searching for the next notes that needed to play. These phenomena can also explain the results. Moreover, these performance-ME combined results can also imply that students in HiEmb & HiViz group could perform better by investing less psychological effort than in other conditions (Paas & Van Merriënboer, 1993).

5.4. Implications, limitations, and future work

Generally, the present study supports the promise of high-level sequential visual cues can promote students' musical skill performance, decrease their cognitive load, and optimize the design of instructional materials during learning with a VR-based yangqin. Students can benefit from the time-sequenced VizCues on strings by guiding their cognitive progress (Jamet, 2014; Yang et al., 2022). Moreover, we also found limited effects of high embodied design on students' learning and instructional design. However, to some extent, this negative effect can be mediated by adding HiViz. While our study looks through the lens of a specific domain, several general implications for designing and conducting effective VR-based instruction could be drawn.

On the one hand, we highly recommend adding necessary VizCues or prompts during design with an immersive VR (IVR) manipulative to reduce the distraction effect elicited by the high degree of freedom and wide range of horizon in the IVR environment (Makransky et al., 2019; Makransky & Petersen, 2021). However, it is noted that when the platform does not entail rich interactions (e.g., mouse-system interaction), there is no need to consider the HiViz. On the other hand, we do not need to consider the high level of embodiment when some simple simulations can solve problems. According to the quantitative results and feedback from learners, manipulating the level of embodiment at a comfortable level seems to be more important; that is, rich embodied interactions are not always needed. This is another extension of the "platform is not destiny" (Johnson-Glenberg et al., 2021), wherein the agency and embodiment designed into learning content are more important than the platform, and we should carefully select the simulation platform that fits the learning topic.

The findings could also direct future studies that we should focus more on learning material and its suitable instructional design rather than the platform. We should carefully consider the trade-off between LE and LVC when designing VR manipulatives.

Some limitations exist in the current study that put forward guidelines for future studies. First, the rhythm score was subjectively assessed by a musical expert, it may to some extent influence the result due to "human errors". To produce a more reliable result, we may invite more experts to evaluate or utilize some algorithms to automatically assess the rhythmic score via computer. Second, the VR yangqin lacks tactile feedback, which might affect the user's playing precision (Mäki-patola, Laitinen, Kanerva, & Takala, 2005). Since the VR hand controller can customize the vibration frequency, we may design different tactile feedbacks when students hit the string or make errors. Third, we directly manipulated the immersion and gesture congruency to assign the embodied level by referring to the Johnson-Glenberg (2018)'s embodiment taxonomy. Some effects may be omitted by looking through this combined perspective. A more interesting finding may emerge if we separately design the immersion and gestural congruency of manipulatives.

In addition, our study only focused on the performance of beginners during an "introductory activity", the intervention time was really limited, and the effect of virtual yangqin on students' learning should be further explored. It is suggested that an upgrade of the system and a longitudinal experiment should be conducted based on this study to improve the internal validity and further examine students' performance changes over time. What is more, as Mäki-patola et al. (2005) noted, replicating interfaces of traditional instruments in VR may not bring about useful results since it lacks a direct experience with real instruments. Therefore, we plan to examine students' transfer performance on a real instrument to further explore the effect of VR intervention. Additionally, using AR technology to overlay VizCues information on a real dulcimer to drill and practice seems to be another potential solution (Johnson et al., 2020). In future, we may also consider the inclusion of virtual hands or body to improve students' body ownership and presence during playing an instrument (Ogawa, Narumi, & Hirose, 2019, pp. 519–528).

6. Conclusion

While studies indicated that VR musical instrument could benefit students' instrument learning experience (Cui, 2022; Innocenti et al., 2019; Johnson et al., 2020; Qingtang et al., 2021), the VRMI design, as well as its educational efficacy, were somehow omitted from previous studies. In this study, we provide a first step towards the integration of VR in yangqin learning and instruction by looking at the impact of embodied design and visual cues on students' learning. From the empirical results, the great potential of visual cues and the potential effect of high embodied design were presented. Specifically, the low level of embodied design could facilitate students' playing efficiency; the high-level visual cues design could improve students playing rhythmic accuracy, and they can both decrease students' playing errors and improve their overall performance. Moreover, we also found that the design of high-level visual cues could benefit learning on the IVR platform by decreasing students' cognitive load and improving the instructional efficiency of learning material. This study outlines the importance of considering the trade-off between the level of embodiment and the level of visual cues in designing virtual learning environments.

Credit author statement

Shufan Yu: Conceptualization, methodology, data analysis, writing, reviewing, and editing. **Qingtang Liu**: Reviewing & editing. **Mina C. Johnson-Glenberg**: Reviewing & editing. **Miaomiao Han**: Software, experiment, and data collection. **Jingjinag Ma**: Reviewing & editing. **Shen Ba**: Reviewing & editing. **Linjing Wu**: Reviewing & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors are deeply grateful to members in virtual learning lab from Central China Normal University and embodied games lab from Arizona State University, especially to Man Su, Ye Li, Christine Shin-Ping Yu, Huixiao Le and Mengtong zhao who gently supported and collaborated on this research.

This work was supported by the China Scholarship Council (202106770025), Fundamental Research Funds for the Central Universities of China (2022YBZZ025), Wuhan Pre Funded Science and Technology Project (2020010601012190), Humanities and Social Sciences Research Project of the Ministry of Education of China (22YJAZH067), and National Natural Science Foundation of China (62277021).

References

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research & Development, 68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

Atherton, J., & Wang, G. (2020). Doing vs. Being: A philosophy of design for artful VR. *Journal of New Music Research, 49*(1), 35–59. https://doi.org/10.1080/09298215.2019.1705862

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Ba, S., Stein, D., Liu, Q., Long, T., Xie, K., & Wu, L. (2021). Examining the Effects of a Pedagogical Agent With Dual-Channel Emotional Cues on Learner Emotions, Cognitive Load, and Knowledge Transfer Performance. *Journal of Educational Computing Research, 59*(6), 1114–1134. https://doi.org/10.1177/07356331211992421.

Burbules, N. C. (2006). Rethinking the virtual. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *The international handbook of virtual learning environments* (pp. 37–58). Springer Netherlands. https://doi.org/10.1007/978-1-4020-3803-7_1.

Camci, A., & Granzow, J. (2019). Hyperreal instruments: Bridging VR and digital fabrication to facilitate new forms of musical expression. *Leonardo Music Journal, 29*, 14–18. https://doi.org/10.1162/lmj_a_01056

Cheng, K. H. (2017). Reading an augmented reality book: An exploration of learners' cognitive load, motivation, and attitudes. *Australasian Journal of Educational Technology, 33*(4), 53–69. https://doi.org/10.14742/ajet.2820

Cui, K. (2022). Artificial intelligence and creativity: Piano teaching with augmented reality applications. *Interactive Learning Environments*, 1–12. https://doi.org/10.1080/10494820.2022.2059520

de Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educational Psychology Review, 23*(4), 501–521. https://doi.org/10.1007/s10648-011-9173-8

Glowinski, D., Baron, N., Shirole, K., Yahia Coll, S., Chaabi, L., Ott, T., et al. (2015). Evaluating music performance and context-sensitivity with immersive virtual environments. *EAI Endorsed Transactions on Creative Technologies, 2*(2), e3. https://doi.org/10.4108/ct.2.2.e3

Hille, A., & Schupp, J. (2015). How learning a musical instrument affects the development of skills. *Economics of Education Review, 44*, 56–82. https://doi.org/10.1016/j.econedurev.2014.10.007

Huang, F., Zhou, Y., Yu, Y., Wang, Z., & Du, S. (2011). Piano AR: A markerless augmented reality based piano teaching system. *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, 2*, 47–52. https://doi.org/10.1109/IHMSC.2011.82

Innocenti, E. D., Geronazzo, M., Vescovi, D., Nordahl, R., Serafin, S., Ludovico, L. A., et al. (2019). Mobile virtual reality for musical genre learning in primary education. *Computers & Education, 139*, 102–117. https://doi.org/10.1016/j.compedu.2019.04.010

Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior, 32*, 47–53. https://doi.org/10.1016/j.chb.2013.11.013

Johnson, D., Damian, D., & Tzanetakis, G. (2020). Evaluating the effectiveness of mixed reality music instrument learning with the theremin. *Virtual Reality, 24*(2), 303–317. https://doi.org/10.1007/s10055-019-00388-8

Johnson, D., Dufour, I., Damian, D., & Tzanetakis, G. (2016). Detecting pianist hand posture mistakes for virtual piano tutoring. *ICMC 2016 - 42nd International Computer Music Conference, Proceedings*, 167–170.

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.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86–104. https://doi.org/10.1037/a0034008.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications, 2*(1), 24. https://doi.org/10.1186/s41235-017-0060-9.

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. *Frontiers in Robotics and AI, 5* (JUN), 1–19. https://doi.org/10.3389/frobt.2018.00081

Kilteni, K., Bergstrom, I., & Slater, M. (2013). Drumming in immersive virtual reality: The body shapes the way we play. *IEEE Transactions on Visualization and Computer Graphics, 19*(4), 597–605. https://doi.org/10.1109/TVCG.2013.29

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education, 95*(January), 174–187. https://doi.org/10.1016/j.compedu.2016.01.001

Lui, Q., Ba, S., Wu, L., Huang, J., & Li, H. (2018). Virtual Dulcimer Auxiliary Teaching System Based on Musical Instrument Digital Interface. In *Proceedings - 2018 International Symposium on Educational Technology, ISET 2018* (pp. 82–86). https://doi.org/10.1109/ISET.2018.00027

Lui, M., McEwen, R., & Mullally, M. (2020). Immersive virtual reality for supporting complex scientific knowledge: Augmenting our understanding with physiological monitoring. *British Journal of Educational Technology, 51*(6), 2181–2199. https://doi.org/10.1111/bjet.13022

Liu, Q., Yu, S., Chen, W., Wang, Q., & Xu, S. (2021). The effects of an augmented reality based magnetic experimental tool on students' knowledge improvement and cognitive load. *Journal of Computer Assisted Learning, 37*(3), 645–656. https://doi.org/10.1111/jcal.12513.

Mäki-patola, T., Laitinen, J., Kanerva, A., & Takala, T. (2005). Experiments with virtual reality instruments. *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression. March 2014*, 11–16 http://dl.acm.org/citation.cfm?id=1085946.

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*. 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
Mayer, R. E. (2002). Multimedia learning. Psychology of learning and motivation - advances in research and theory. https://doi.org/10.5926/arepj1962.41.0_27.
Mazur, Z., & Łaguna, M. (2017). Assessment of instrumental music performance: Definitions, criteria, measurement. *Edukacja, 115–128*. https://doi.org/10.24131/3724.170508
Ogawa, N., Narumi, T., & Hirose, M. (2019). *Virtual hand realism affects object size perception in body-based scaling*. IEEE Conference on Virtual Reality and 3D User Interfaces (VR). https://doi.org/10.1109/VR.2019.8798040, 2019.
Orman, E. K. (2004). Effect of virtual reality graded exposure on anxiety levels of performing musicians: A case study. *Journal of Music Therapy, 41*(1), 70–78. https://doi.org/10.1093/jmt/41.1.70
Orman, E. K., Price, H. E., & Russell, C. R. (2017). Feasibility of using an augmented immersive virtual reality learning environment to enhance music conducting skills. *Journal of Music Teacher Education, 27*(1), 24–35. https://doi.org/10.1177/1057083717697962
Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research & Development, 53*(3), 25–34. https://doi.org/10.1007/BF02504795
Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(4), 737–743. https://doi.org/10.1177/001872089303500412
Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371. https://doi.org/10.1007/BF02213420
Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241. https://doi.org/10.1111/jcal.12482
Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology, 51*(6), 2098–2114. https://doi.org/10.1111/bjet.12991
Qingtang, L., Yuwei, J., Jindian, L., Miaomiao, H., Jingjing, M., & Shufan, Y. (2021). Design and Implementation of Virtual Sanxian Teaching System. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)* (pp. 938–943). https://doi.org/10.1109/TALE52509.2021.9678770.
Schroeder, N. L. (2017). The influence of a pedagogical agent on learners' cognitive load. *Educational Technology & Society, 20*(4), 138–147. https://www.jstor.org/stable/26229212.
Serafin, S., Adjorlu, A., Nilsson, N., Thomsen, L., & Nordahl, R. (2017). Considerations on the use of virtual and augmented reality technologies in music education. In *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)* (pp. 1–4). https://doi.org/10.1109/KELVAR.2017.7961562
Serafin, S., Erkut, C., Kojs, J., Nilsson, N. C., & Nordahl, R. (2016). Virtual reality musical instruments: State of the art, design principles, and future directions. *Computer Music Journal, 40*(3), 22–40. https://doi.org/10.1162/COMJ_a_00372
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5
Tse, P.-S. P. (2007). Innovation and reform of the hammered dulcimer yangqin in contemporary China [University of Hawai'i] http://hdl.handle.net/10125/20752.
Turchet, L., Hamilton, R., & Camci, A. (2021). Music in extended realities. *IEEE Access, 9*(January), 15810–15832. https://doi.org/10.1109/ACCESS.2021.3052931
Van Kerrebroeck, B., Caruso, G., & Maes, P. J. (2021). A methodological framework for assessing social presence in music interactions in virtual reality. *Frontiers in Psychology, 12*(June). https://doi.org/10.3389/fpsyg.2021.663725
Vogt, A., Albus, P., & Seufert, T. (2021). Learning in virtual reality: Bridging the motivation gap by adding annotations. *Frontiers in Psychology, 12*(March). https://doi.org/10.3389/fpsyg.2021.645032
Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model. *Computers & Education, 171*(April), Article 104235. https://doi.org/10.1016/j.compedu.2021.104235
Williamon, A., Aufegger, L., & Eiholzer, H. (2014). Simulating and stimulating performance: Introducing distributed simulation to enhance musical learning and performance. *Frontiers in Psychology, 5*(FEB), 1–9. https://doi.org/10.3389/fpsyg.2014.00025
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322
Yang, F., Su, C., Xu, W., & Hu, Y. (2022). Effects of developing prompt scaffolding to support collaborative scientific argumentation in simulation-based physics learning. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2022.2041673
Yu, Z. (2021). A meta-analysis of the effect of virtual reality technology use in education. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2021.1989466, 0(0.
Zhang, X. (2020). The influence of educational concept on dulcimer teaching mode in the new era and the corresponding reform. In *2020 2nd asia-pacific Conference on Advance in education, Learning and teaching (ACAELT 2020)*. https://doi.org/10.25236/acaelt.2020.189. *Acaelt*, 899–902.

Shufan Yu is a Ph.D. candidate at the School of Educational Information Technology at Central China Normal University (CCNU). His research interests include Augmented/Virtual /Augmented Reality (AR/VR/MR) in education and human-computer interactive technology.

Qingtang Liu is a professor at the School of Educational Information Technology at CCNU. His recent research interests include technology-enhanced learning, AR/VR in education, e-learning and teacher professional development.

Mina C. Johnson-Glenberg is a research professor in the Department of Psychology at Arizona State University. Her research interests include STEM education, educational games, embodied cognition, and AR/VR/MR (XR) in education.

Miaomiao Han is Master student at the School of Educational Information Technology at CCNU. Her research interests include Augmented/Virtual Reality in education, and music education.

Jingjinag Ma is a Ph.D. student at the School of Educational Information Technology at CCNU. Her research interests include Augmented/Virtual /Augmented Reality (AR/VR/MR) in education, and museum learning.

Shen Ba is a Post-Doctoral Fellow at the Faculty of Education at The University of Hong Kong. His research interests include technology-enhanced learning, pedagogical agent, and music education.

Linjing Wu is an associate professor at the School of Educational Information Technology at CCNU. Her recent research interests include technology-enhanced learning, e-learning and natural language processing.