

Article

Not peer-reviewed version

Evaluating the Role of Visual Fidelity in Digital Vtubers on Mandarin Chinese Character Learning

[Xiaoxiao Cao](#)*, [Wei Tong](#), Kenta Ono, [Makoto Watanebe](#)

Posted Date: 3 December 2024

doi: 10.20944/preprints202412.0146.v1

Keywords: Vtuber; Digital Learning; Design; Chinese Character



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Evaluating the Role of Visual Fidelity in Digital Vtubers on Mandarin Chinese Character Learning

Xiaoxiao CAO ^{1,2}, Wei Tong ^{1,*}, Kanta Ono ² and Makoto Watanabe ²

¹ Department of Product Design, School of Design and Art, Communication University of Zhejiang, China

² Design Course, Graduate School of Chiba University, Chiba, Japan

* Correspondence: tony@cuz.edu.cn

Abstract: Despite the growing presence of digital Vtubers in educational settings, there is limited empirical evidence on their effectiveness in language acquisition. In this investigation, we delved into the realm of digital education to assess how the visual fidelity of digital Vtuber avatars affects the acquisition of Mandarin Chinese characters by beginners. Through incorporating a diverse array of digital Vtubers, ranging from simple two-dimensional figures to complex three-dimensional models, we explored the relationship between digital Vtuber design and learner engagement and efficacy. This study employed a randomized tutorial distribution, immediate post-tutorial quizzing, and a realism scoring rubric, with statistical analysis conducted through Pearson correlation. The analysis involving 608 participants, illuminated a clear positive correlation: digital Vtubers with higher levels of realism significantly enhanced learning outcomes, underscoring the importance of visual fidelity in educational content. This research substantiates the educational utility of digital Vtubers and underscores their potential in creating more immersive and effective digital learning environments. The findings advocate for leveraging sophisticated digital Vtubers to foster deeper learner engagement, improve educational achievement, and promote sustainable educational practices, offering insights for the future development of digital learning strategies.

Keywords: Vtuber; digital learning; design; Chinese character

1. Introduction

In an era marked by digital innovation, the landscape of education, particularly language learning, has experienced transformative shifts [1]. The integration of advanced technologies has opened new avenues for engaging and effective educational methodologies [2]. Among these advancements, the emergence of Virtual YouTubers (Vtubers) signifies a pivotal evolution from traditional educational methodologies to interactive, technology-driven pedagogies [3]. Initially rooted in the entertainment industry [4], Vtubers — virtual avatars operated by individuals or AI for live broadcasting — have transcended their original domains to emerge as influential tools in educational spheres [5], as well as in the realm of language acquisition [6]. The integration of digital Vtubers into educational contexts aligns with the global trend toward sustainable digital innovation, reducing the need for physical resources and making education more accessible and scalable [7].

As digital natives navigate through the vast sea of online content, the appeal of digital Vtuber lies not only in their entertainment value but also in their potential to serve as engaging educational facilitators [8]. This shift towards virtual presence and interaction mirrors the broader technological trend towards the metaverse [9], a collective virtual shared space that is being heralded as the future of online social interaction [10]. Within this emerging metaversal landscape, digital Vtuber stand at the forefront [11], poised to redefine educational engagement by acting as virtual tutors, companions, and brand ambassadors in richly immersive digital environments [12].

However, the adoption of digital Vtuber in educational settings raises pivotal questions regarding the role of visual fidelity — the degree of detail and realism in the avatar design — and its impact on learning efficacy [13]. While visual aesthetics have long been recognized for their role in

enhancing learner engagement and motivation [14], the specific influence of digital Vtuber realism on the learning outcomes of language students remains an underexplored territory. This gap in research underlines the need for a nuanced examination of how varying degrees of visual sophistication in digital Vtuber can affect the acquisition and retention of Mandarin Chinese characters among novice learners. Previous research in educational psychology and second language acquisition has highlighted the importance of engagement and motivation in language learning [15]. Studies suggest that visual and interactive elements can enhance learning outcomes [16]. This study delves into the integration of digital Vtuber technology as a novel approach to Mandarin Chinese character education for beginners, a demographic increasingly seeking engaging and effective learning solutions in a globalized context [17]. The present study builds on these findings by exploring the specific impact of digital avatars on learner engagement and character acquisition.

1.1. Mandarin Chinese Education

Mandarin Chinese education has traditionally been centered around rote memorization and repetitive practice, which has often resulted in disengagement and reduced motivation among learners [18]. The complexity of Mandarin characters, combined with the tonal nature of the language, poses unique challenges for students, especially beginners [19]. Traditional teaching methods have emphasized memorization of characters and repetitive writing exercises, which, while effective to some degree, often fail to engage learners fully or provide them with the necessary contextual understanding [20]. Furthermore, the cultural and linguistic nuances of Mandarin are sometimes lost in conventional teaching methods, leading to a gap between theoretical knowledge and practical application [21].

1.2. Mandarin Chinese Education and CALL

The advent of computer-assisted language learning (CALL) has introduced interactive and multimedia elements into language education [22], significantly enhancing engagement and effectiveness [23]. Recent studies also highlight the potential of digital tools to enhance personalized learning in Mandarin education, particularly in engaging beginner-level learners [24]. However, integrating digital avatars, particularly Vtubers, into CALL for Mandarin Chinese education represents a further innovative leap [25]. Vtubers, as highly interactive and visually engaging virtual avatars, offer unique advantages [26]. They can simulate real-life interactions, provide personalized feedback, and maintain high learner interest through their dynamic personas [27]. Moreover, Vtubers can embody cultural and linguistic nuances [28], offering a richer contextual understanding of Mandarin Chinese. This capability bridges the gap between theoretical knowledge and practical application, making learning more holistic and integrated [29]. By integrating Vtubers into CALL environments, educators can overcome many limitations of traditional methods [30]. Vtubers provide consistent, engaging, and interactive practice that adapts to the learner's pace and style, enhancing language retention and leading to better educational outcomes [31]. Additionally, as technology advances, digital avatars will become even more advantageous for a wider range of educational content, further expanding their applicability and effectiveness in diverse teaching contexts [32]. This integration represents a significant advancement, harnessing modern technology to create an effective and engaging learning environment [33].

This inquiry is particularly relevant in the context of China's burgeoning digital market, where there is a pronounced demand for the creation of virtual personas by companies and educational institutions alike [34]. Despite the enthusiasm, the market's maturation is hindered by challenges in balancing the technical, temporal, and financial costs associated with developing high-fidelity digital humans against their potential educational value [35]. Thus, our study aims to provide empirical insights into the efficacy of using digital Vtuber of different visual fidelities as tools for Mandarin Chinese character education.

By conducting a comprehensive investigation that encompasses a wide spectrum of digital Vtuber representations — from basic two-dimensional animations to advanced three-dimensional models — this research offers a detailed exploration of how digital Vtuber design influences language

learning. Through methodical evaluation and rigorous analysis, the study seeks to illuminate the optimal use of digital Vtuber technology in enhancing the language learning experience, thereby contributing to the evolving dialogue on digital innovations in education. Moreover, by highlighting the broad applications and potential of Vtubers in educational domains, our research underscores their significance in shaping future digital learning environments and their integration into the expanding metaverse.

1.3. Research Questions

To address gaps and explore Vtubers' potential in Mandarin Chinese education, this study poses the following questions:

How does the visual fidelity of digital Vtubers impact engagement and learning outcomes for beginners?

What are learner preferences regarding different levels of Vtuber realism, and how do these correlate with educational effectiveness?

How can integrating Vtubers into CALL environments enhance overall learning experiences and outcomes?

The findings will contribute to the dialogue on digital innovations in education, highlighting Vtubers' role in shaping future digital learning environments. This study aims to inform the development of more effective and engaging digital learning tools, enhancing educational outcomes in Mandarin Chinese and beyond.

2. Materials and Methods

This study was rigorously conducted in compliance with the Declaration of Helsinki and adhered to the ethical standards set forth by the institutional and national research committees. Ethical approval was granted by the Ethics Committee of Communication University of Zhejiang (Approval No. 20230901). All participants engaged in the study provided informed consent. Informed consent was obtained from all participants. The recruitment period for participants was from February 14, 2024, to March 8, 2024.

This study ensures confidentiality and anonymity of all participants. Identifying information has been omitted or appropriately anonymized in the manuscript, and any potentially identifying information included has been crucial for scientific purposes and consented to by the participants as per the guidelines. This study employed a mixed-methods design to evaluate the impact of digital Vtuber visual fidelity on the acquisition of Mandarin Chinese characters by beginners. The study was conducted in two phases, incorporating both qualitative and quantitative data collection and analysis.

2.1. Participants

A total of 608 participants were involved in the study. In the initial stage, a globally diverse cohort of 50 participants, all non-native speakers of Mandarin Chinese, was selected. This cohort consisted of 16 males, 33 females, and 1 individual preferring not to disclose their gender, with ages ranging from 18 to 54 years ($M = 29.5$, $SD = 10.2$). In the second phase, the participant pool was refined to 608 native English-speaking individuals from the United States, including 248 males, 356 females, and 4 individuals identifying with other genders, aged 18 to over 60 years ($M = 34.7$, $SD = 12.4$). In this study, participants were not pretested on the target Chinese characters before the tutorial. This choice was made to observe the natural learning process without prior influence. However, the absence of a pretest could introduce limitations, as it does not account for participants' prior familiarity with the characters, potentially affecting the study's findings.

2.2. Samples Preparation

In the preparation phase of this study, a strategic and collaborative design approach was adopted to create the digital Vtuber pivotal to our investigation. This process was spearheaded by our interdisciplinary team, encompassing experts from the fields of digital art, cognitive psychology, and language education. The team's primary objective was to engineer a series of digital Vtuber that not only captured a broad spectrum of visual fidelity but also embodied rich cultural and linguistic elements essential for effective Mandarin Chinese character instruction.

To this end, a comprehensive suite of eight digital Vtuber was meticulously developed, each designed to mark a distinct point along a continuum from elementary two-dimensional animations to sophisticated three-dimensional hyperrealistic figures. This gradient of digital Vtuber complexity was conceptualized to systematically evaluate the impact of visual fidelity on learner engagement and retention. Figure 1 presents these digital Vtuber in ascending order of visual sophistication, providing a visual narrative of the progression from basic to hyperrealistic digital Vtuber design.



Figure 1. Progression of Vtuber Avatar Design: From Basic to Hyperrealistic Visual Fidelity.

The selection process for the Chinese characters to be taught was conducted in collaboration with seasoned Chinese as a Foreign Language educators. The criteria for character selection were rigorously defined to ensure pedagogical relevance and uniform difficulty across the chosen characters. To ensure that beginners can master the selected Chinese characters with consistent difficulty, we chose characters that are frequently used in daily life. These characters are similar in terms of structural complexity and cognitive load. Additionally, the selected characters have an appropriate balance of similarity and difference, which helps learners to understand the form, pronunciation, and meaning of each character while also building a comprehensive understanding of Chinese characters. This selection strategy ensures the balance of learning content and the consistency of educational objectives. This collaborative effort resulted in the identification of eight Mandarin characters that were deemed representative of common learning challenges faced by beginners.

Subsequently, a set of uniform educational scripts was crafted for each selected character. These scripts were not only designed to standardize instructional content across different avatar stages but also to incorporate pedagogically sound principles aimed at maximizing language acquisition. Each script was then paired with the corresponding avatar stage, culminating in the production of eight Mandarin character instructional videos. These videos were meticulously standardized in terms of duration and pedagogical content, three examples are shown in Figure 2.



Figure 2. Three Sample Instructional Videos Showcasing Varied Digital Vtuber for Mandarin Character Education.

During the avatar creation phase, a detailed segmentation technique was employed for the two-dimensional models [36]. This technique involved the dissection of avatars into constituent elements, such as facial features and limb characteristics, to facilitate precise animation and realism [37]. The segmentation process, illustrated in Figure 3, exemplifies the meticulous attention to detail that underpins the creation of engaging and pedagogically effective digital Vtuber.



Figure 3. Detailed Segmentation Process for Two-Dimensional Vtuber Avatar Design.

For the three-dimensional character models, an advanced skeletal rigging process was employed [38], complemented by the use of green screen technology to accurately capture the movements of live actors [39]. This process ensured the avatars' motions were realistic and congruent with natural human movement, thereby enhancing the authenticity of the instructional videos. The culmination of this phase was marked by the production of eight instructional videos, each showcasing a consistent range of motion reflective of the avatars' intended realism levels. The green screen capture process is detailed in Figure 4, while Figure 5 provides insights into the hybrid technique used for animating characters' mouth movements [40], combining frame-by-frame animation of real human performances with AI-driven facial expression capture technology to achieve lip-syncing effects [41].



Figure 4. Implementation of Green Screen Technology in Capturing Realistic Avatar Motion.

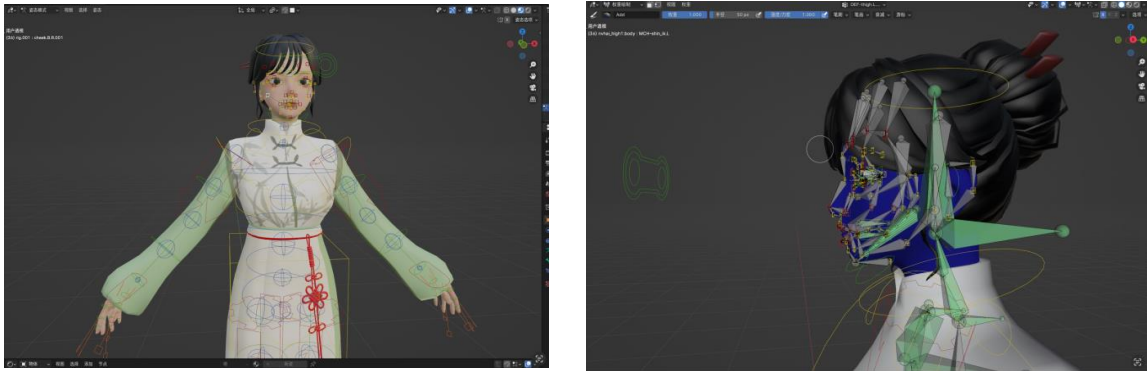


Figure 5. Hybrid Technique for Animating Vtuber Mouth Movements: A Blend of Frame-by-Frame and AI-Driven Facial Expression Capture.

EIn the first phase of the study, each Chinese character was paired with a corresponding Vtuber, resulting in the creation of 8 instructional videos. Building on this foundation, the second phase involved selecting 5 stages of Vtubers and 2 characters identified from the first phase, leading to the production of 12 additional videos. Each of these characters was paired with 5 Vtubers and one video without a Vtuber. To ensure the validity and reliability of the results, these videos were counterbalanced across participants, with different character-Vtuber pairings presented in a random order to each participant. This enhanced preparation protocol highlights our commitment to developing a diverse and pedagogically robust set of digital Vtubers. By integrating detailed design elements and leveraging advanced animation technologies, this approach provides a solid foundation for evaluating the educational potential of virtual characters in language learning.

2.3. Evaluation Method

To rigorously evaluate the impact of digital Vtuber on Mandarin Chinese character acquisition, our study employed a Mixed Methods Design, combining the collection and analysis of both quantitative and qualitative data. Quantitative data were gathered through tests and questionnaires to quantify learning outcomes and participant performance, while qualitative data were obtained through interviews and responses to open-ended questions, providing deeper insights into participants' learning experiences and their subjective evaluations of the Vtubers. This methodology was developed with the dual objectives of assessing immediate learning efficacy and understanding the nuanced influence of visual fidelity on learner motivation and retention.

2.3.1. Participant Assessment Protocol

At the core of our evaluation strategy was a controlled variable framework, enabling a systematic investigation into the educational effectiveness of digital Vtuber across varying levels of visual fidelity. Utilizing a balanced mix of qualitative and quantitative data collection tools. The collection of qualitative data primarily took place during the initial assessment and screening phase of the study. On one hand, we gathered participants' accuracy rates through preliminary test banks, and on the other hand, we conducted interviews to understand their learning experiences and feedback on the Vtubers. The interview data were then coded and categorized using content analysis to identify key themes and patterns, thereby enriching the findings from the quantitative data analysis. The study's assessment protocol included the following key components:

- **Randomized Tutorial Distribution:** To ensure the integrity of our experimental design, participants were randomly assigned to one of the eight Vtuber-mediated instructional videos via an automated distribution mechanism. The random assignment of participants in this study was implemented through the advanced membership module of the SurveyMonkey platform. This module provides an automated random allocation function, enabling us to ensure that participants were impartially assigned to different Vtuber tutorial groups, thereby maintaining

- the integrity of the experimental design. This random allocation was critical in mitigating selection bias and facilitating a clean analysis of the avatars' impact on learning outcomes;
- Immediate Post-Tutorial Quizzing: Following exposure to the instructional content, participants were immediately assessed through structured quizzes. These quizzes were crafted to measure retention and comprehension of the Mandarin characters presented, with each item designed to directly correlate with specific tutorial content;
 - Realism Scoring Rubric: Each Vtuber avatar was assigned a realism score on a scale from 0 (no avatar) to 8 (highest level of visual fidelity), providing a quantifiable measure of visual sophistication for statistical analysis. This scoring system was integral to our investigation, allowing for a nuanced exploration of how visual realism affects learning engagement and effectiveness.

2.3.2. Data Analysis

The study employed advanced statistical tools to thoroughly analyze the collected data, focusing on identifying significant trends and relationships. Pearson correlation analyses [42] were used to examine the relationship between the realism scores of digital Vtubers and quiz performance, chosen for their robustness in highlighting critical patterns within the dataset. Additionally, the evaluation assessed participants' preferences for specific avatars, capturing the subjective appeal of different levels of visual fidelity. This aspect of the analysis explored the interaction between aesthetic preference and educational efficacy, contributing to a more comprehensive understanding of the pedagogical value of digital Vtubers. Through this multifaceted evaluation method, the study provides empirically grounded insights into the role of digital Vtubers in language education, emphasizing the importance of visual fidelity in educational content and paving the way for future research into the potential of immersive virtual technologies to enhance learning experiences.

2.4. Procedure

This study's methodology was designed with precision to assess the impact of Vtuber avatar sophistication on Mandarin Chinese character learning. We streamlined our approach into two distinct phases, both methodically structured to provide insights into the pedagogical effectiveness of varied avatar designs.

2.4.1. Initial Assessment and Tutorial Screening

Participants were exposed to a diverse range of tutorials, each featuring a different Chinese character and an associated Vtuber avatar. Post-tutorial, participants were assessed through quizzes to evaluate comprehension and retention. Tutorials that balanced beginner accessibility with sufficient educational challenge were identified based on accuracy rates. Table 2 details the characters and their corresponding accuracy rates, highlighting those selected for deeper analysis based on their lower comprehension scores.

Table 1. Initial Assessment Accuracy Rates for Mandarin Characters Across Tutorials.

Character	Accuracy	Character	Accuracy
风	95.67%	哭	91.00%
固	80.00%	林	98.67%
河	97.33%	学	98.67%
画	96.00%	晕	81.33%

2.4.2. Focused Evaluation

The subsequent phase honed in on characters "固" and "晕" due to their initial lower accuracy rates, indicating a potential for revealing more about the design's impact on learning [42]. Tutorials

featuring characters "固" and "晕" were selected for further investigation due to their lower initial accuracy rates. A broader participant pool was used to solidify the findings. Participants were randomly assigned to tutorials featuring avatars across a spectrum of realism.

- Key Methodological Highlights:
- Preliminary Screening: Determination of tutorial efficacy through immediate post-engagement quizzes ;
 - Character Selection for Depth Study: Based on Table 1, "固" and "晕" were chosen for further investigation due to notable error rates ;
 - Randomized Tutorial Distribution: Ensured unbiased exposure to various avatar designs among participants ;
 - Objective and Subjective Evaluation: Quizzes measured learning outcomes, while surveys gauged avatar preferences ;
 - By employing this bifurcated methodological approach, our research endeavors to unravel the nuances of how visual fidelity in digital avatars influences language acquisition, particularly focusing on the Mandarin Chinese learning process among beginners.

3. Results

The study’s empirical investigation yielded insightful data on the pedagogical efficacy of digital Vtuber within the domain of Mandarin Chinese character education. The core findings, grounded in a comprehensive analysis involving 608 participants, highlight the nuanced relationship between visual fidelity and language acquisition.

3.1. Quantitative Findings

Character Recognition Accuracy: The analysis of participant performance across the eight distinct tutorials showcased a discernible pattern correlating visual fidelity with learning outcomes. Specifically, avatars representing higher levels of visual fidelity (stages 5 through 8) were associated with a marked increase in character recognition accuracy. For instance, tutorials featuring the most realistic avatars (stage 8) resulted in an average accuracy rate of 81.33% for the character "晕", compared to 80.00% for the character "固" using less sophisticated avatars (stage 1).

Learner Engagement and Preferences: The study further explored the subjective preferences of learners towards different avatar designs. A notable preference for higher fidelity avatars was observed, with 40.83% of participants favoring stage 5 avatars. This preference aligns closely with the improved accuracy rates, suggesting a positive correlation between avatar likability and educational outcome effectiveness.

Referencing Table 2: The distribution of favorability scores, as detailed in Table 2, underscores the participants' inclination towards more sophisticated digital Vtuber. The favorability percentage peaks at 40.83% for stage 5 avatars, illustrating a clear learner bias towards avatars of higher visual fidelity.

Table 2. Participant Favorability Ratings for digital Vtuber by Focused Evaluation Stage.

Vtuber NO.	Favorability
01	10.00%
02	10.28%
03	40.83%
04	18.89%
05	20.00%

Upon further analysis, it was noted that the accuracy scores for stage 5 avatars were lower than those for stages 1 through 4. This discrepancy highlights the complexity of the relationship between visual fidelity and learning outcomes, suggesting that factors other than fidelity may also play a significant role. Additionally, Table 4 indicates that the most preferred avatars were from stage 3, not stage 5, as initially reported.

3.2. Qualitative Findings

Thematic analysis of participant feedback revealed that increased engagement and perceived realism play a significant role in influencing learning outcomes. The study's participants spanned a wide range of ages and had diverse gender compositions, with those in the first phase of interviews generally possessing at least a bachelor's degree, which supports the representativeness of the findings. The majority of participants believed that higher visual fidelity made the learning process more immersive and engaging, enhancing their attention and involvement. Specifically, many participants noted that when the avatars' visual presentation was more realistic, they found it easier to focus on the learning content, thereby improving the overall quality of their learning experience.

However, some participants pointed out that excessively high visual fidelity might draw too much attention to the visual presentation itself, potentially detracting from the absorption of the learning content. Participants' personal interests and preferences were also considered important factors influencing the results. For example, some enthusiasts of the "二次元" (anime and manga) culture preferred abstract two-dimensional characters over high-definition, high-fidelity three-dimensional images. Additionally, the style differences of the avatars might influence participants' attention and learning preferences to some extent.

These feedbacks suggest that while high-fidelity avatars can enhance learning outcomes in some cases, their impact varies from person to person and may not always be directly correlated with optimal learning outcomes.

3.3. Statistical Analysis

To evaluate the effect of Vtuber visual fidelity on learning outcomes, Pearson correlation coefficients were applied. This method allowed us to assess the strength and statistical significance of the relationship between avatar visual fidelity and participants' accuracy in recognizing Chinese characters.

For the character "固," Pearson's correlation coefficient was $r = 0.88$ ($p < .01$), indicating a strong positive relationship between higher visual fidelity and increased character recognition accuracy. Similarly, for the character "晕," the correlation was $r = 0.86$ ($p < .01$), supporting the hypothesis that greater avatar realism enhances learning outcomes. These findings highlight the importance of avatar design in educational contexts, suggesting that more realistic avatars can significantly improve learning performance.

Assumption checks for normality and homogeneity of variances were conducted before performing parametric analyses, confirming the appropriateness of the tests used. Additionally, segmented regression analysis revealed potential non-linear relationships between visual fidelity and learning outcomes, suggesting that the effect of visual realism may vary across different stages of fidelity.

3.4. Graphical Representation

3.3. Graphical Representation

Figures 6–9 visually depict the study's findings, illustrating the nuanced impact of avatar design on learning efficacy. For example, Figure 6 highlights the correlation between the "固" character's teaching content accuracy and the Vtuber's favorability, reinforcing the significance of engaging and visually appealing avatars in educational contexts.

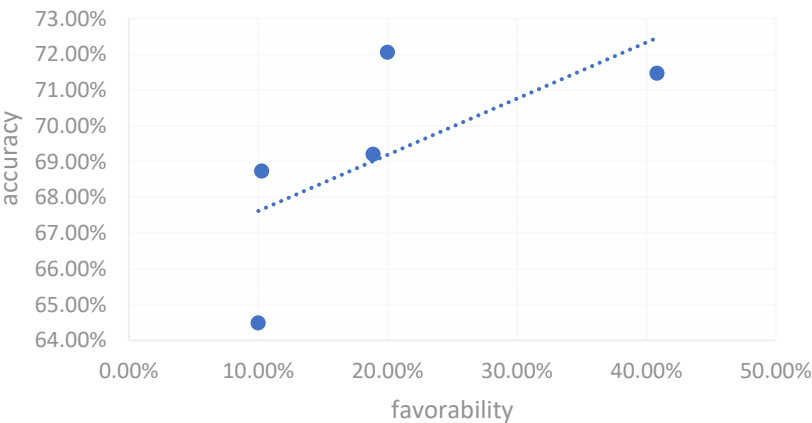


Figure 6. Correlation between Vtuber Favorability and Learning Accuracy for the Mandarin Character '固'.

Figure 7 shows the effect of Vtuber visual fidelity on the learning accuracy for the Mandarin character "固." As visual fidelity increases, there is an observable improvement in learning outcomes, suggesting that higher-quality, more realistic Vtubers may help sustain learners' attention and improve comprehension of the character.

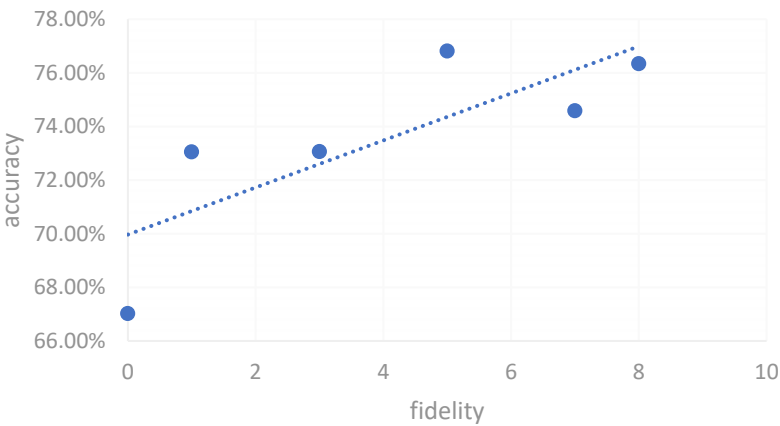


Figure 7. Impact of Vtuber Visual Fidelity on the Learning Accuracy for the Mandarin Character '固'.

Figure 8 depicts the relationship between Vtuber favorability and learning accuracy for the Mandarin character "晕." The data suggests a positive correlation between the avatars learners found more appealing and their performance in character recognition. This indicates that subjective preferences could influence educational effectiveness, as learners might perform better when engaging with avatars they find appealing.

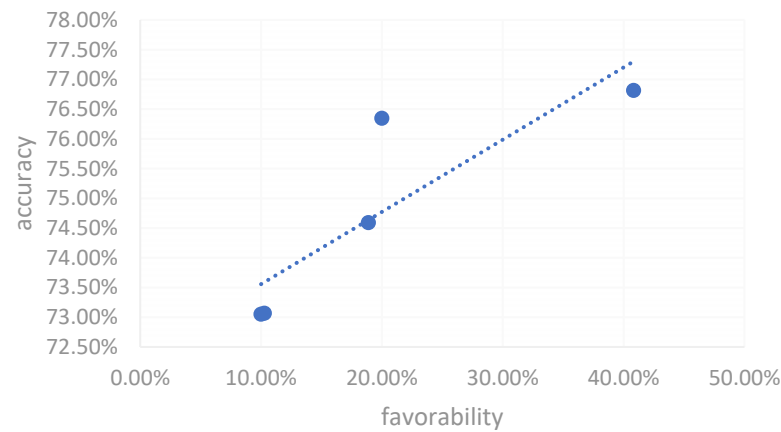


Figure 8. Relationship between Vtuber Favorability and Learning Accuracy for the Mandarin Character '晕'.

Figure 9 focuses on the influence of Vtuber visual fidelity on learning accuracy for the character "晕." Similar to Figure 7, it indicates that higher visual fidelity can lead to better learning outcomes, reinforcing the potential importance of avatar realism in supporting cognitive processes related to language learning.

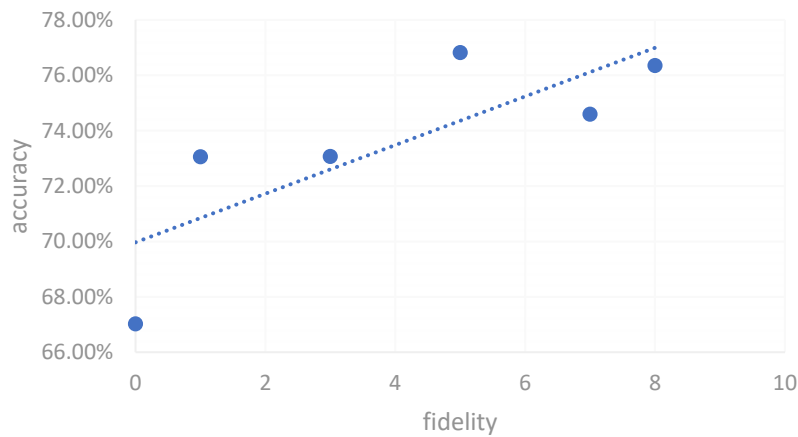


Figure 9. Influence of Vtuber Visual Fidelity on the Learning Accuracy for the Mandarin Character '晕'.

This examination highlights the role of digital Vtubers' visual fidelity in Mandarin Chinese character acquisition, showing a connection between digital avatar design and educational outcomes. The correlations between avatar likability, realism, and learning performance suggest that Vtuber technology may offer opportunities to enhance language education methodologies.

4. Discussion

The positive correlations identified in our study partially validate existing notions regarding the utility and design optimization of digital avatars in educational settings. The findings suggest that the engagement and effectiveness of Vtuber-mediated education are not solely a function of content quality but are also significantly influenced by the design attributes of the avatars themselves. One possible reason for this is that avatars with higher visual fidelity can create a more immersive and engaging learning environment, which in turn enhances the learner's focus and retention of the material. However, it is also important to consider that the high visual fidelity of avatars might sometimes lead to an overemphasis on visual elements, potentially distracting learners from the core educational content. This dual impact of visual fidelity suggests that while high-quality avatar design

can improve educational outcomes, it must be carefully balanced to ensure that the visual elements do not overshadow the learning objectives.

Our findings extend beyond the immediate context of Mandarin Chinese character education, suggesting potential applications across various educational disciplines where engagement and material retention are paramount. This insight opens up new avenues for language teaching strategies, where the subtle nuances of avatar design can be leveraged to foster a more engaging and effective learning environment. For instance, educators might consider varying the visual complexity of avatars based on the content being taught, or tailoring avatar designs to match the preferences of different learner groups.

However, our study also has limitations that need to be acknowledged. The subjective qualitative analysis was primarily conducted during the first phase of the study and focused mainly on young learners with a strong interest in digital technologies. This concentration on a specific demographic may limit the generalizability of the findings to other populations, such as older learners or individuals less engaged with digital media. Additionally, relying on self-reported data in this phase could introduce bias, as participants may have provided responses they believed were expected rather than accurately reflecting their true experiences.

The study also explored learner preferences, revealing a clear inclination towards avatars with higher visual complexity. This preference not only reflects a demand for engaging and aesthetically appealing educational tools but also aligns with improved educational outcomes. However, individual differences in preferences, influenced by factors such as familiarity with digital culture or personal learning styles, suggest that a one-size-fits-all approach may not be effective. It is recommended that future studies explore how different learner characteristics, such as age, cultural background, and prior experience with digital tools, influence the effectiveness of various avatar designs.

These insights have important implications for the future of digital pedagogy. The integration of high-fidelity digital Vtubers may serve as a key direction for innovative educational strategies, particularly in the context of language education. The study contributes to the ongoing dialogue on the integration of advanced digital technologies in educational settings, highlighting the potential of Vtubers in creating more immersive and personalized learning environments. Future research should explore how these technologies can be optimized for different educational contexts, ensuring that the balance between engagement and content delivery is maintained.

The findings from this study, while significant, are subject to several limitations. The lack of a pretest means that prior familiarity with the characters could have influenced the results. Furthermore, the preference for certain avatars may reflect participants' prior exposure to digital media rather than the educational content itself. Future studies should consider implementing a pretest to control for prior knowledge and explore the long-term retention of characters learned with different avatars. Additionally, the potential for certain avatars to distract from the educational content should be further investigated.

5. Conclusions

This study has clearly demonstrated the potential of digital Vtubers as effective educational tools and has provided direction for their design and development. The core insights emphasize the multifaceted influence of avatar design on language acquisition, showing that both the visual fidelity and likability of digital Vtubers play a significant role in enhancing learning outcomes. However, the study also highlights the complexity of this relationship, suggesting that while high visual fidelity can enhance engagement, it may also divert attention away from the educational content if not carefully managed.

The findings offer a clearer perspective on the potential role of Vtubers within the expanding metaverse. As digital learning environments continue to evolve towards more immersive and interactive realms, digital Vtubers stand at the cusp of redefining educational engagement. They represent a bridge between current digital learning tools and the future of education in the metaverse, suggesting a paradigm shift towards more dynamic and interactive forms of learning.

In conclusion, this research not only highlights the significant potential of digital Vtubers as effective educational tools but also sets the stage for future explorations into their broader applications across various domains of digital education. As we move forward, it becomes crucial to continue investigating the optimal integration of such technologies, ensuring they are leveraged to enhance educational quality and accessibility in the digital age. Future studies should consider expanding the demographic scope to include younger learners or individuals less familiar with digital environments, and should also investigate the long-term impacts of using high-fidelity avatars on learning retention and overall satisfaction.

Building on our insights into digital Vtubers for Mandarin character education, future research should not only explore broader educational applications across diverse subjects but also venture into varied contexts beyond traditional learning environments. This includes examining the effects of visual fidelity on learner psychology, leveraging Vtuber technology in novel digital ecosystems for increased engagement and scalability, and understanding the long-term influence on user retention and satisfaction. Expanding research to encompass different scenarios, such as professional training, informal learning, and interactive entertainment, will provide a holistic view of digital avatars' capacity to innovate and enrich a wide spectrum of human experiences.

Supplementary Materials: The instructional videos used for testing can be viewed at the following address: <https://www.youtube.com/@xiaoxiaocao2903/videos>.

Author Contributions: Conceptualization, X.C. and K.O.; methodology, X.C. and K.O.; software, W.T.; validation, X.C. and W.T.; formal analysis, X.C.; investigation, X.C.; resources, M.W. and K.O.; data curation, X.C.; writing—original draft preparation, X.C.; writing—review and editing, X.C.; visualization, T.W.; supervision, M.W.; project administration, X.C.; funding acquisition, X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the SCIENTIFIC RESEARCH FUND of ZHEJIANG PROVINCIAL EDUCATION, grant number y202351745; the 14th 5-Year TEACHING REFORM PROJECT of GENERAL COLLEGES and UNIVERSITIES in ZHEJIANG PROVINCE, grant number jg20220416, and the CHINA SCHOLARSHIP COUNCIL, grant number 202008330389.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are deeply appreciative of the support from students and friends at the Laboratory for Systems Planning, Chiba University, hailing from around the globe. We also wish to express our sincere thanks to Teachers Jin Jing and Xie Qun from the Chinese as a Foreign Language Education department at Communication University of Zhejiang for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gutiérrez-Portlán, I.; Prendes-Espinosa, P.; Sánchez-Vera, M. D. M. Digital Technologies for the Assessment of Oral English Skills. *Applied Sciences* **2022**, *12* (22), 11635. <https://doi.org/10.3390/app122211635>.
2. Bosmos, F.; Tzallas, A.T.; Tsipouras, M.G.; Glavas, E.; Giannakeas, N. Virtual and Augmented Experience in Virtual Learning Tours. *Information* **2023**, *14*, 294. <https://doi.org/10.3390/info14050294>.
3. Turner, A. B. Streaming as a Virtual Being: The Complex Relationship Between VTubers and Identity.
4. Ferreira, J. C. V.; Regis, R. D. D.; Gonçalves, P.; Diniz, G. R.; Tavares, V. P. D. S. C. VTuber Concept Review: The New Frontier of Virtual Entertainment. In *Proceedings of the 24th Symposium on Virtual and Augmented Reality*; ACM: Natal, RN Brazil, 2022; pp 83–96. <https://doi.org/10.1145/3604479.3604523>.
5. Swenson, A. Teaching Digital Identity: Opportunities, Challenges and Ethical Considerations for Avatar Creation in Educational Settings. *BCIJ* **2023**, *3* (2), 41–58. <https://doi.org/10.25112/bcij.v3i2.3538>.
6. Davidson, K. The Challenges of the Virtual Classroom—The Semiotics of Transmedial Literacy in VR Education. *Language and Semiotic Studies* **2021**, *7* (4), 1–25. <https://doi.org/10.1515/lass-2021-070401>.
7. Moro, C.; Mills, K. A.; Phelps, C.; Birt, J. The Triple-S Framework: Ensuring Scalable, Sustainable, and Serviceable Practices in Educational Technology. *Int J Educ Technol High Educ* **2023**, *20* (1), 7. <https://doi.org/10.1186/s41239-022-00378-y>.

8. *Embracing the Future: Creative Industries for Environment and Advanced Society 5.0 in a Post-Pandemic Era: Proceedings of the 8th Bandung Creative Movement International Conference on Creative Industries (8th BCM 2021), Bandung, Indonesia, 9 September 2021*, First edition.; Aulia, R., Ed.; CRC Press: Boca Raton, 2023.
9. Mozumder, M. A. I.; Sheeraz, M. M.; Athar, A.; Aich, S.; Kim, H.-C. Overview: Technology Roadmap of the Future Trend of Metaverse Based on IoT, Blockchain, AI Technique, and Medical Domain Metaverse Activity. In *2022 24th International Conference on Advanced Communication Technology (ICACT)*; IEEE: PyeongChang Kwangwoon_Do, Korea, Republic of, 2022; pp 256–261. <https://doi.org/10.23919/ICACT53585.2022.9728808>.
10. Hennig-Thurau, T.; Aliman, D. N.; Herting, A. M.; Cziehso, G. P.; Linder, M.; Kübler, R. V. Social Interactions in the Metaverse: Framework, Initial Evidence, and Research Roadmap. *J. of the Acad. Mark. Sci.* **2023**, 51 (4), 889–913. <https://doi.org/10.1007/s11747-022-00908-0>.
11. Babineaux, A. Controversial saints: A study in how popular culture can radicalize religious icons.
12. Saldaña, C. M.; Welner, K.; Malcolm, S.; Tisch, E. Teachers as Market Influencers: Towards a Policy Framework for Teacher Brand Ambassador Programs in K-12 Schools. *EPAA* **2021**, 29 (August-December). <https://doi.org/10.14507/epaa.29.5654>.
13. Tang, M. T.; Zhu, V. L.; Popescu, V. AlterEcho: Loose Avatar-Streamer Coupling for Expressive VTubing. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*; IEEE: Bari, Italy, 2021; pp 128–137. <https://doi.org/10.1109/ISMAR52148.2021.00027>.
14. Filetti, M.; Barral, O.; Jacucci, G.; Ravaja, N. Motivational Intensity and Visual Word Search: Layout Matters. *PLoS ONE* **2019**, 14 (7), e0218926. <https://doi.org/10.1371/journal.pone.0218926>.
15. Iwaniec, J.; Khaled, H. Y. Motivation and Engagement in Language Learning. In *Cognitive and Educational Psychology for TESOL*; Cirocki, A., Indrarathne, B., McCulloch, S., Eds.; Springer Texts in Education; Springer Nature Switzerland: Cham, 2024; pp 131–149. https://doi.org/10.1007/978-3-031-66532-5_7.
16. Rasch, T.; Schnotz, W. Interactive and Non-Interactive Pictures in Multimedia Learning Environments: Effects on Learning Outcomes and Learning Efficiency. *Learning and Instruction* **2009**, 19 (5), 411–422. <https://doi.org/10.1016/j.learninstruc.2009.02.008>.
17. Cooper *, B. J. The Enigma of the Chinese Learner. *Accounting Education* **2004**, 13 (3), 289–310. <https://doi.org/10.1080/0963928042000273780>.
18. Grenfell, M.; Harris, V. Memorisation Strategies and the Adolescent Learner of Mandarin Chinese as a Foreign Language. *Linguistics and Education* **2015**, 31, 1–13. <https://doi.org/10.1016/j.linged.2015.04.002>.
19. Olmanson, J.; Liu, X. C. The Challenge of Chinese Character Acquisition: Leveraging Multimodality in Overcoming a Centuries-Old Problem. *Design Journal* **2017**, 4.
20. Xu, W.; Stahl, G. Routine Literacy Practices as a Cultural Agenda: Children’s Experiences of Writing “Difficult” Chinese Characters in Australia. *Journal of Language, Identity & Education* **2023**, 1–12. <https://doi.org/10.1080/15348458.2023.2252515>.
21. Pan, L.; Sun, D.; Zou, Y.; Cao, Y.; Zhang, J.; Li, F. Psycho-Linguistic and Educational Challenges in Teaching Chinese (Mandarin) Language: Voices from None-Chinese Teachers of Mandarin Language. *BMC Psychol* **2023**, 11 (1), 390. <https://doi.org/10.1186/s40359-023-01432-8>.
22. Paracka, D. Global Learning and the Engaging Questions of Globalization. 2008, 3.
23. Li, X. Some Interactional Uses of Syntactically Incomplete Turns in Mandarin Conversation. *CLD* **2016**, 7 (2), 237–271. <https://doi.org/10.1075/cld.7.2.03li>.
24. Lang, Q.; Zhang, C.; Qi, H.; Du, Y.; Zhu, X.; Zhang, C.; Li, M. Mining and Utilizing Knowledge Correlation and Learners’ Similarity Can Greatly Improve Learning Efficiency and Effect: A Case Study on Chinese Writing Stroke Correction. *Sustainability* **2023**, 15, 2393. <https://doi.org/10.3390/su15032393>.
25. Gündüz N. Computer Assisted Language Learning. *Journal of Language and Linguistic Studies*. **2005**;1(2):193–214.
26. Zhang, R.; Zou, D. A State-of-the-Art Review of the Modes and Effectiveness of Multimedia Input for Second and Foreign Language Learning. *Computer Assisted Language Learning* **2022**, 35 (9), 2790–2816. <https://doi.org/10.1080/09588221.2021.1896555>.
27. Tseng, J.-J.; Tsai, Y.-H.; Chao, R.-C. Enhancing L2 Interaction in Avatar-Based Virtual Worlds: Student Teachers’ Perceptions. *AJET* **2013**, 29 (3). <https://doi.org/10.14742/ajet.283>.

28. Miranda, C.; Costa, M.; Pereira, M.; Almeida, S.; Branco, F.; Au-Yong-Oliveira, M. VTubers, Their Global Expansion and Impact on Modern Society: An Exploratory and Comparative Study Between Portugal and the USA. In *Information Systems and Technologies*; Rocha, A., Adeli, H., Dzemyda, G., Moreira, F., Colla, V., Eds.; Lecture Notes in Networks and Systems; Springer Nature Switzerland: Cham, 2024; Vol. 799, pp 223–231. https://doi.org/10.1007/978-3-031-45642-8_22.
29. Wenbin Z. Online Identities Construction: Participatory Culture of Virtual YouTuber Fans in China.
30. Lu, Z.; Shen, C.; Li, J.; Shen, H.; Wigdor, D. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*; ACM: Yokohama Japan, 2021; pp 1–14. <https://doi.org/10.1145/3411764.3445660>.
31. Japan's contemporary media culture between local and global. Content, Practice and Theory.
32. Kong, R.; Qi, Z.; Zhao, S. Difference Between Virtual Idols and Traditional Entertainment from Technical Perspectives;; Guangzhou, China, 2021. <https://doi.org/10.2991/assehr.k.211209.058>.
33. Hsieh, R.; Yamamura, K.; Cho, S.; Sato, H. Free to Select Digital Avatar vs. Predetermined Digital Avatar in eLearning System A Comparison between Student Reaction in Two Models. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*; IEEE: Taichung, Taiwan, 2021; pp 258–266. <https://doi.org/10.1109/AIVR52153.2021.00059>.
34. Mohammad, W.; Maulidiyah, N. R. Analysis of the SAMR Model and Psychological Approach on Duolingo Livestreams by Hololive VTubers in Learning Japanese. *jpn* **2023**, 1 (1), 9. <https://doi.org/10.47134/jpn.v1i1.71>.
35. Hong-Yi Pai, H.-Y. P.; Hong-Yi Pai, C.-H. W.; Ching-Huang Wang, Y.-C. L. Development of an Interactive Live Streaming System for Language Learning. *Journal of Internet Technology* **2024**, 25 (1), 027–036. <https://doi.org/10.53106/160792642024012501003>.
36. Luo, L.; Kim, W. How Virtual Influencers' Identities Are Shaped on Chinese Social Media: A Case Study of Ling. *Global Media and China* **2023**, 20594364231188353. <https://doi.org/10.1177/20594364231188353>.
37. Duffy, V. G. Digital Human Modeling In Design. In *Handbook Of Human Factors And Ergonomics*; Salvendy, G., Karwowski, W., Eds.; Wiley, 2021; pp 761–781. <https://doi.org/10.1002/9781119636113.ch29>.
38. Joshi, P.; Tien, W. C.; Desbrun, M.; Pighin, F. Learning Controls for Blend Shape Based Realistic Facial Animation. In *ACM SIGGRAPH 2006 Courses on - SIGGRAPH '06*; ACM Press: Boston, Massachusetts, 2006; p 17. <https://doi.org/10.1145/1185657.1185857>.
39. Kalra, P.; Magnenat-Thalmann, N.; Moccozet, L.; Sannier, G.; Aubel, A.; Thalmann, D. Real-Time Animation of Realistic Virtual Humans. *IEEE Comput. Graph. Appl.* **1998**, 18 (5), 42–56. <https://doi.org/10.1109/38.708560>.
40. Le, B. H.; Deng, Z. Robust and Accurate Skeletal Rigging from Mesh Sequences. *ACM Trans. Graph.* **2014**, 33 (4), 1–10. <https://doi.org/10.1145/2601097.2601161>.
41. Hughes, P. J.; Pan, K.; Kendrach, M. G. Student Outcomes and Perceptions Related to Chroma Key (Green Screen) Technology Utilized in a Drug Literature Evaluation Course. *Med.Sci.Educ.* **2017**, 27 (4), 693–699. <https://doi.org/10.1007/s40670-017-0464-7>.
42. Schober, P.; Boer, C.; Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia* **2018**, 126 (5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>.
43. Chegwidden, W. R.; Watts, D. C. Kinetic Studies and Effects of Anions on Creatine Phosphokinase from Skeletal Muscle of Rhesus Monkey (Macaca Mulatta). *Biochim Biophys Acta* **1975**, 410 (1), 99–114. [https://doi.org/10.1016/0005-2744\(75\)90210-7](https://doi.org/10.1016/0005-2744(75)90210-7).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.