Towards Understanding the Impact of AI-Generated Visual and Vocal Self-Similarity on Avatar Identification, Motivation, and Engagement in Educational Games

Tianyu Shao
School of Applied and Creative Computing
Purdue University
West Lafayette, Indiana, USA
shao122@purdue.edu

Olaoluwa Oyedokun
School of Applied and Creative Computing
Purdue University
West Lafayette, Indiana, USA
ooyedoku@purdue.edu

Christos Mousas
School of Applied and Creative Computing
Purdue University
West Lafayette, Indiana, USA
cmousas@purdue.edu

Dominic Kao
Siebel School of Computing and Data Science
University of Illinois Urbana-Champaign
Urbana-Champaign, Illinois, USA
kaod@purdue.edu

Abstract

In educational games, aligning avatar representation with the player's own characteristics may enhance engagement and learning by strengthening psychological immersion and self-identification. This project proposes a work-in-progress study to investigate how generative AI–based visual and auditory self-similarity in avatars impacts avatar identification, motivation, competence, and performance during a programming-based puzzle game. We plan to implement a 2 (Visual Similarity: high vs. low) × 2 (Auditory Similarity: high vs. low) between-subjects design, assigning participants to one of four avatar conditions in a Unity-based game teaching Java concepts, *CodeBreakers*. Participants will provide facial photos and voice samples for generating self-similar avatars using AI tools, while generic avatars will feature neutral faces and synthetic voices. Participants will complete the same game tasks. System logs, performance metrics, and self-report measures will be collected to examine how multimodal avatar similarity impacts avatar identification, intrinsic motivation, need satisfaction, and task performance. By prototyping this system, this study aims to inform the design of generative AI-powered educational technologies that support personalized learning experiences through identity alignment.

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**.

Keywords

Avatar, Education, Generative AI, Competence, Relatedness, Identification, Audio

CHI PLAY Companion '25, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-2023-9/25/10
https://doi.org/10.1145/3744736.3749363

ACM Reference Format:
Tianyu Shao, Olaoluwa Oyedokun, Christos Mousas, and Dominic Kao. 2025. Towards Understanding the Impact of AI-Generated Visual and Vocal Self-Similarity on Avatar Identification, Motivation, and Engagement in Educational Games. In *Companion Proceedings Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '25), October 13–16, 2025, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3744736.3749363

1 Introduction

In recent years, advances in generative AI have introduced new possibilities for creating personalized avatars through automated facial reconstruction and voice cloning. These tools enable the generation of high-fidelity avatars that resemble users in both appearance and voice, expanding the potential for immersive and personalized learning experiences [19, 75, 77]. In educational games, avatars have long been recognized for their ability to support player engagement and task persistence [25, 32, 45]. However, there is limited empirical understanding of how multimodal self-similarity enabled by generative AI, both in visual and auditory dimensions, influences learning outcomes and user experience in virtual environments.

To address this gap, we propose a project that investigates how avatar self-similarity in both face and voice affects player competence, avatar identification, and task performance in an educational programming game. Research has shown that resemblance between a user and their avatar, whether visual or auditory, can significantly influence self-identification, engagement, and performance [26, 30]. Visually, facial similarity has been associated with increased trust, presence, and emotional connection in both entertainment and training contexts [3, 17, 54, 73]. Moderate resemblance often offers the optimal balance between familiarity and comfort, avoiding the discomfort of exact likeness [54, 79]. More recently, auditory self-similarity has emerged as a promising but underexplored factor [37]. Studies suggest that hearing one's own voice, or a closely matched synthesized version, can improve motivation, emotional salience, and goal alignment [22, 37]. Speech synthesis technologies like Tacotron 2 [67], Glow-TTS [44], and StyleTTS2 [49] now make it possible to generate real-time personalized voices that reflect user tone and affect [19, 38, 81]. Although these tools are gaining traction in areas such as behavioral nudging and affective computing [22], they remain underutilized in educational games, particularly in tandem with facial similarity, as in this paper.

Building on these insights, our study will investigate how combining visual and auditory self-similarity in avatars influences learners' avatar identification, intrinsic motivation, need satisfaction, and task performance. Participants will complete short Java-based puzzles in an educational game *CodeBreakers* [33, 38] using avatars that vary in visual and auditory resemblance. This project is guided by the theoretical frameworks of self-perception theory [8], self-association [45], and player identification theory [73] and is informed by Gee's principles of learning [25]. We will employ triangulation in our methodology that will integrate multiple sources of data, including game performance, behavioral engagement, and self-report measures. These will include the Player Identification Scale (PIS) [71], the interest/enjoyment subscale of the Intrinsic Motivation Inventory (IMI) [51], the competence subscale of the Player Experience of Need Satisfaction (PENS) [64], an open-ended post-game experience survey analyzed using inductive thematic analysis[13], and behavioral metrics such as puzzles completed, hints used, and voluntary engagement duration. Overall, this project aims to provide new insight into how generative AI–enabled identity alignment can enhance performance and personalization in educational technology.

2 Related Work

2.1 Avatar Self-Similarity and Avatar Identification

Avatar-based systems enhance user identification, presence, and engagement in interactive learning environments [4, 21, 30]. Avatar identification [71] allows players to temporarily adopt avatar characteristics, influencing attitudes, behaviors, and motivation [45, 79], supported by self-perception theory [8]and embodiment theory [40, 57].

Visual self-similarity, where an avatar's appearance matches the user's facial features, increases trust, perceived competence, and social presence [3, 17, 54]. Users prefer and feel more connected to agents mirroring their facial features, aligning with the "mere-exposure effect," [52, 72, 76, 80]. Users also respond more quickly and attend more deeply to avatars that resemble themselves [41, 57, 58]. Studies of role model avatars further show that visual resemblance to aspirational figures perceived as competent and sharing similar demographics can promote identification and motivation in educational games [34, 38]. These effects extend beyond perception: users are more likely to mimic behavior and align socially with self-similar avatars, reinforcing a sense of ownership and identification [6, 43].

Auditory self-similarity, though less studied, has emerged as a promising factor with similarly powerful effects [37]. Research shows that players with self-similar avatar voices had significant performance increases, time spent, similarity identification, competence, relatedness, and immersion [26, 37, 38]. Vocal cues activate influence perceptions of credibility, warmth, and competence [36, 53]. Cross-modal congruence further enhances embodiment and learning, with aligned voice tone and visual features increasing presence and behavioral engagement [48, 66, 78]. Multimodal similarity enhances perceived realism and reduces cognitive dissonance [75, 81]. These findings emphasize aligning avatar design with user attributes for motivation and immersion.

2.2 Avatar Influence in Education

Avatar design is central to shaping users' cognitive and emotional engagement in educational environments [21, 27]. In game-based learning, avatars scaffold complex concepts within narrative, interaction, and embodiment [21, 25], particularly in programming games [32, 34]. Player identification with avatars increases task investment, persistence, intrinsic motivation, and self-efficacy [10, 23, 35, 39, 70]. This is supported by evidence that avatars can reflect users' ideal selves, promoting sustained engagement [9, 10, 31, 59]. Avatar identification consistently enhances learning outcomes in computational environments [32, 38]. Customization, both visual [3, 29] and auditory [37, 53], influences identification. When avatars reflect a student's voice or face, deeper engagement and increased motivation are reported [26, 38]. Generative AI tools further personalize learning agents, amplifying identification and embodiment effects [19, 23, 75, 77]. Beyond motivation, avatar embodiment affects cognitive performance and psychological states [57]. Studies show that embodying specific avatars can improve working memory, mitigate stereotype threat, and increase engagement [5, 6, 24, 34, 39, 55]. Physiological evidence, such as EEG activity and HRV patterns, suggests that avatar embodiment facilitates emotional stability and cognitive alignment conducive to learning [61].

2.3 Avatar System Design and Generative AI

Educational game design increasingly integrates avatar systems for interactivity, personalization, and conceptual learning. In programming education, games like JavaStrike [33] and Imikode [12] use avatar-mediated interfaces to teach computational concepts, providing real-time feedback within interactive narratives [46]. Avatars as instructors or peers can improve autonomy and reduce cognitive load [27, 32]. Immersive virtual environments and 3D affordances align well with logic-based and experiential learning [4, 7, 15, 25]. Generative AI has transformed educational avatar systems by enabling real-time synthesis of human-like characteristics, including facial likeness [3, 75], dynamic movements [19, 65], realistic voice [75], and emotional responsiveness [14, 81]. AI can generate 3D models from images [69], manipulate facial features [1, 3, 77], and animate static portraits [82]. Frameworks like VALID [18] and FaceGen [68] support visual similarity manipulation. These AI-driven avatars enhance realism and personalization, aligning with learner identity [23, 56, 75, 81].

Despite this progress, few studies examine how generative facial and vocal alignment together impact learning outcomes. While AI-generated voices boost identification [22, 23] and facial self-similarity increases trust [54], these modalities are largely studied in isolation. This work-in-progress study aims to bridge this gap by exploring the combined effects of AI-generated facial and vocal self-similarity on identification, competence, and task performance in an educational programming game.

2.4 Hypotheses

This project is guided by established theoretical frameworks in avatar research and educational psychology. Self-perception theory [8] suggests that individuals construct aspects of their self-concept by observing their behavior in context, including how they engage with self-resembling representations. Player identification theory [45, 74], on the other hand, posits that individuals can form meaningful psychological connections with avatars, which in turn influence need satisfaction and performance. Complementing these views, self-association theory [45] argues that users are more likely to identify with and respond favorably to avatars that resemble them, as such representations are cognitively linked to the self and evoke self-referential processing. Furthermore, Gee's principles of learning [25] emphasize the importance of identifying alignment for fostering deep engagement and persistence in game-based educational environments. Guided by these frameworks, this study focuses on quantized task performance and three theoretically grounded and outcome-relevant constructs: the Player Identification Scale (PIS) [71], which captures similarity, embodied, and wishful identification; interest/enjoyment, the primary indicator of intrinsic motivation [51]; and competence, a fundamental need within Self-Determination Theory that reflects players' perceived effectiveness—each selected for its strong conceptual relevance and empirical support in avatar-mediated learning contexts [64].

This study proposes four hypotheses concerning the effects of avatar self-similarity on user experience in an educational programming game.

Prior research has shown that visual resemblance between avatars and users increases identification by fostering a sense of recognition, embodiment, and personal relevance [3, 54]. Self-perception theory [8] and player identification theory [45, 74] suggest that individuals infer internal states and motivations from their interaction with avatars that reflect their external characteristics. Therefore, we hypothesize:

H1: Learners with avatars that exhibit high visual self-similarity will report higher levels of avatar identification, intrinsic motivation, need satisfaction, and task performance, compared to learners with generic avatars.

Auditory cues also contribute to avatar identification, particularly when the avatar speaks in the user's own voice. Studies have shown that cloned or familiar voices increase immersion and emotional connection [22, 38], while auditory customization enhances the effects of visual resemblance [37]. We propose:

H2: Learners with avatars that exhibit high auditory self-similarity will report higher avatar identification, intrinsic motivation, need satisfaction, and task performance compared to those with avatars using synthetic voices.

When visual and auditory cues are combined, their alignment may yield additive or synergistic effects on user experience. Auditory customization has been shown to enhance avatar identification, particularly when paired with visual similarity [37]. Consistent multimodal cues improve perceived realism and credibility [26, 81], and studies of multimodal learning environments show that integrating these channels can boost task performance [78]. Thus, we hypothesize:

H3: Learners experiencing both high visual and high auditory self-similarity will report higher avatar identification, intrinsic motivation, need satisfaction, and task performance than those experiencing only one or neither modality.

Finally, drawing on prior research, both visual [20, 42] and auditory self-similarity [22, 37] are shown to increase avatar identification. This enhanced avatar identification is consistently associated with positive learner outcomes, including increased need satisfaction [32, 64], and improved task performance and time spent [38]. Therefore, it is hypothesized that the combined effect of visual and auditory self-similarity will be mediated by avatar identification.

H4: The combined effect of visual and auditory self-similarity on intrinsic motivation, need satisfaction, and task performance will be mediated by avatar identification, such that higher self-similarity increases avatar identification, which in turn will lead to enhanced motivation, greater psychological need fulfillment, and improved performance outcomes.

Additionally, we aim to conduct exploratory analyses to understand how learners subjectively evaluate the perceived relevance and impact of self-similar avatars compared to generic representations. While self-similarity tends to enhance subjective avatar identification and certain positive feelings [26, 37], perceived impact on performance or overall engagement can be complex, sometimes favoring generic designs [50, 65]. Exploratory analysis is therefore necessary as the subjective evaluation of relevance and overall impact is nuanced, and baseline understandings of how various avatar types affect learners differently are still being established.

RQ1. How do learners subjectively evaluate the relevance and impact of self-similar avatars compared to generic ones during gameplay?

3 Methodology

3.1 Game Setup

The proposed study will use *CodeBreakers* [33], a Java programming game in which players solve progressively challenging code puzzles by throwing code snippets at virtual targets. The game includes 14 puzzles across six difficulty levels and was iteratively developed in collaboration with professional developers. The game will be played on desktop devices using keyboard and mouse input to maintain consistency across participants and ensure accessibility. To support this study, we plan to extend *CodeBreakers* with avatar customization capabilities by integrating ReadyPlayerMe[1] and ElevenLabs.[2] Participants will be asked to upload a facial photo, which ReadyPlayerMe will process to generate a 3D avatar resembling their appearance. A short voice sample will also be collected and cloned using ElevenLabs' speech synthesis API to produce a personalized avatar voice. These self-similar avatars will then be embedded into the *CodeBreakers* interface and used as in-game characters. This setup enables a controlled investigation into how multimodal identity alignment impacts engagement, competence, and learning performance in an educational programming context.

[1]Ready Player Me: Cross-Game Avatar Platform available at https://readyplayer.me/
[2]ElevenLabs: Prime Voice AI platform available at https://elevenlabs.io/

3.2 Intervention

This study employs a 2 (Visual Self-Similarity: high vs. low) × 2 (Auditory Self-Similarity: high vs. low) between-subjects experimental design to investigate how multimodal self-similarity influences players' psychological experiences and task engagement. Participants will be randomly assigned to one of four avatar conditions. In the high-similarity conditions, avatars will be generated using participants' facial photos and voice samples to produce personalized visual and vocal representations via ReadyPlayerMe and ElevenLabs, respectively. In contrast, the generic avatar will feature a low-poly, gender-neutral humanoid model, presented in monochrome tones and lacking individualized facial features. The voice will be a human-like, gender-neutral, monotone synthetic voice generated by ElevenLabs using a stock English model. Each participant will interact with only one avatar throughout the session to prevent condition contamination.

3.3 Qualitative and Quantitative Measures:

To assess player identification, intrinsic motivation, need satisfaction, and task performance, we will adopt a combination of validated self-report instruments and behavioral metrics. We follow best practices in measurement reporting by clearly explaining the what, how, and why for each measure [2].

3.3.1 Player Identification. To measure avatar identification, the extent to which players perceive psychological connection and self-resemblance with their avatars, we will use the Player Identification Scale (PIS) [71]. This validated self-report instrument, administered post-gameplay, evaluates three distinct dimensions of identification on a 5-point Likert scale: similarity identification (e.g., "My character is similar to me"), embodied identification (e.g., "In the game, it is as if I become one with my character"), and wishful identification (e.g., "I would like to be more like my character"). PIS is a widely adopted and validated scale specifically designed for assessing player-avatar relationships in game contexts, ensuring its content validity for our study.

3.3.2 Intrinsic Motivation. To assess intrinsic motivation, we use the interest/enjoyment subscale of the Intrinsic Motivation Inventory (IMI) [51], which is the primary component of IMI that evaluates participants' inherent satisfaction and engagement with the activity itself. This subscale, rated on a 7-point Likert scale (e.g., "I enjoyed doing this activity very much"), is widely regarded as the primary indicator of intrinsic motivation in research due to its strong psychometric properties. It is particularly appropriate for this study, as it captures how avatar design influences players' voluntary interest and enjoyment, core motivational outcomes relevant to educational games.

3.3.3 Player Experience of Need Satisfaction. We measure need satisfaction through the competence subscale of the Player Experience of Need Satisfaction (PENS) [64]. This subscale captures participants' perceived effectiveness and mastery (e.g., "I feel competent at the game"), which are central to Self-Determination Theory [16, 62] and highly predictive of motivation and persistence in learning environments [63]. Its theoretical alignment and strong psychometric foundation make it a targeted and reliable measure for our research.

3.3.4 Exploratory Analysis. To address the exploratory research question **RQ1**, we will include a post-game experience survey with open-ended questions to assess learners' subjective interpretations of avatar relevance and identity alignment. Participants will respond to prompts such as: "Did your avatar feel like an extension of yourself? Why or why not?", and "How did the avatar's appearance or voice affect your experience during the game?" These questions aim to elicit reflective responses that provide richer insight into how learners interpret multimodal avatar similarity. Responses will be analyzed using inductive thematic analysis [13] to uncover recurring patterns in participants' reflections. These thematic insights will be compared across avatar similarity conditions (e.g., high vs. low visual/auditory similarity) to examine how different forms of identity alignment shape perceived relevance and impact. This analysis aims to complement the quantitative data and highlight subjective nuances in avatar-based engagement.

3.3.5 Game Performance. To assess game performance, which refers to participants' objective learning-related task success and efficiency within the programming game, we will track two direct gameplay metrics, automatically logged by the game system during play: the number of puzzles completed (out of 14) and the number of hints used (up to 42). These metrics serve as objective and quantifiable indicators of problem-solving success and efficiency in the programming tasks, and will help determine the immediate impact of avatar design on performance outcomes.

3.3.6 Voluntary Engagement Duration. To evaluate participants' motivational persistence, we will measure voluntary engagement duration, defined as the total time participants spend in the game environment beyond the required 10-minute session. This metric captures participants' motivation to continue engaging with the task.

Table 1: Measurement Instruments.

Conceptual Group	Instruments	Time
Avatar Identification	Player Identification Scale (PIS) [71]	Post
Intrinsic Motivation	IMI – Interest/Enjoyment Subscale [51]	Post
Need Satisfaction	PENS – Competence Subscale [64]	Post
Exploratory Evaluation	Open-Ended Post-Game Prompts [13]	Post
Game Performance	Puzzles Completed, Hints Used	During
Voluntary Engagement Duration	Time Spent Beyond Required Session	During

3.4 Participants

We plan to recruit approximately 330-340 undergraduate participants from a large U.S. university. Participants will be recruited through multiple introductory computing and programming courses to ensure a sufficient sample size. This sample size is calculated using an a priori power analysis conducted using G*Power 3.1 for a 2×2 between-subjects factorial ANOVA. Assuming a small-to-medium effect size ($f = 0.20$) [60], $\alpha = 0.05$, and 95% power, the analysis indicated a minimum required sample of $N = 327$ [47]. As students enrolled in these courses are learning relevant computer science concepts, they represent an appropriate population for evaluating a Java-based educational game. In addition to core study measures, we will collect self-reported programming experience in the demographic survey at the end of the session. This will be used as a covariate in subsequent analyses to control for potential baseline skill variance. Informed consent will be obtained from all participants prior to participation. Facial photos and short voice recordings will be securely collected for the purpose of avatar generation. All procedures will be reviewed and approved by the Institutional Review Board (IRB) before data collection to ensure ethical compliance and participant privacy.

3.5 Procedure

After completing the informed consent form, participants will be randomly assigned to one of four avatar conditions: high or low visual self-similarity crossed with high or low auditory self-similarity. Each participant will upload a facial photo and a five-second short audio sample, which will be used to generate self-similar avatars via ReadyPlayerMe and ElevenLabs. Next, participants will play the game *CodeBreakers* with their assigned avatar. Participants will begin with a 2-minute guided tutorial introducing game mechanics and the avatar system. This tutorial will be narrated in a neutral synthetic voice to avoid priming effects. Participants will then play for a mandatory 10 minutes, with the option to continue voluntarily afterward. During gameplay, avatars will provide task-relevant spoken feedback, such as celebrating correct actions or offering motivational nudges when players request hints, using a voice aligned with the assigned voice condition. All participants will complete the same set of tasks regardless of avatar condition. Following gameplay, participants will complete a set of self-report questionnaires, including PIS, IMI interest/enjoyment, PENS competence, and a post-game experience survey. Demographic questions, including prior programming experience, will be presented at the end of the session to minimize potential priming effects on task-related self-perception or performance. Game performance (e.g., number of puzzles completed, number of hints used) and voluntary playing time will be automatically logged. The entire session is expected to last approximately 30–40 minutes.

3.6 Analysis

Data will be analyzed using SPSS 29 and the PROCESS macro for SPSS [28]. To evaluate H1-H3, a 2×2 between-subjects factorial ANOVA [11] will be conducted with visual self-similarity (high vs. low) and auditory self-similarity (high vs. low) as independent variables. Dependent variables will include avatar identification (PIS), IMI interest/enjoyment, PENS competence, and task performance metrics. To test H4, a mediation analysis will be conducted using PROCESS Model 4 [28] to determine whether avatar identification mediates the effect of visual and auditory self-similarity on learner outcomes. Each outcome variable (IMI interest/enjoyment, PENS competence, task performance) will be modeled separately, and indirect effects will be tested using bias-corrected bootstrapped confidence intervals [28]. Programming experience will be included as a covariate in the ANOVA and mediation models where appropriate. All significance tests will be conducted at $\alpha = 0.05$.

To support the exploratory objectives of the research question, the post-game experience survey responses will be analyzed using inductive thematic analysis [13] to identify emergent themes in learners' qualitative accounts. Thematic patterns will be compared across avatar conditions (e.g., high vs. low visual/auditory similarity) to examine how different forms of self-similarity shape user experiences. These qualitative insights will be triangulated with quantitative outcomes to contextualize individual differences in avatar identification, intrinsic motivation, need satisfaction, and task performance. Additionally, descriptive statistics and visualizations from key self-report scales (PIS, IMI interest/enjoyment, PENS competence) will be used to illustrate trends across conditions.

4 Future Work

This work-in-progress study serves as an initial step toward understanding how AI-generated visual and vocal self-similarity in avatars impacts player identification, motivation, competence, and task performance in educational games. Game development is currently underway. Future work will involve completing data collection and analysis, including examining interaction and mediation effects as outlined in our hypotheses and the exploratory research question. In preparation for full-scale deployment, we plan to conduct a pilot study to test the fidelity of avatar generation, voice synthesis, interface integration, and task flow. This will help validate the technical setup and inform minor design adjustments prior to full data collection. As an early-stage investigation, this work also opens space for future research directions. While the current design manipulates avatar resemblance through AI-enabled facial and vocal similarity, it does not incorporate user customization. Future studies could investigate how enabling avatar choice or modification affects identification and motivation, especially in educational settings where learner agency is linked to improved outcomes. Additionally, longitudinal deployments are needed to assess whether short-term motivational gains from avatar similarity persist and contribute to long-term learning and behavior change.

5 Conclusion

This work-in-progress study explores the potential of AI-generated visual and vocal self-similarity in avatars to enhance educational game experiences. While prior research has explored avatar customization in learning environments, few have systematically examined the combined effect of visual and auditory self-similarity on performance outcomes. This project is among the first to integrate generative AI technologies, specifically facial reconstruction and voice cloning, into an interactive educational game, enabling the exploration of multimodal identity alignment in a controlled setting.

Building on recent advances in generative AI, this work will bridge the gap between technical personalization and psychological resonance in educational technology. We aim to provide researchers and designers with meaningful insights into the interplay between identity, motivation, player satisfaction, and learners' performance in game-based education. Ultimately, this study seeks to spark new research into the pedagogical potential of generative AI and self-similar avatars and to inform the next generation of personalized learning systems.

References

[1] Naphtali Abudarham and Galit Yovel. 2016. Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision* 16, 3 (2016), 40–40.

[2] Lena Fanya Aeschbach, Sebastian A.C. Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. Transparency in Measurement Reporting: A Systematic Literature Review of CHI PLAY. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 233 (Oct. 2021), 21 pages. doi:10.1145/3474660

[3] C. Alves da Silva, B. Hilpert, C. Bhuvaneshwara, P. Gebhard, F. Nunnari, and D. Tzovlazti. 2023. Visual similarity for socially interactive agents that support self-awareness. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents (IVA '23)*. ACM, 3–5. doi:10.1145/3570945.3607329

[4] M. M. Asad, A. Naz, P. Churi, and M. M. Tahanzadeh. 2021. Virtual Reality as Pedagogical Tool to Enhance Experiential Learning: A Systematic Literature Review. *Education Research International* 2021 (2021), 1–17. doi:10.1155/2021/7061623

[5] D. Banakou, A. Beacco, S. Neyret, M. Blasco-Oliver, S. Seinfeld, and M. Slater. 2020. Virtual body ownership and its consequences for implicit racial bias are dependent on social context. *Royal Society Open Science* 7, 12, Article 201848 (2020). doi:10.1098/rsos.201848

[6] D. Banakou, P. D. Hanumanthu, and M. Slater. 2016. Virtual embodiment of White people in a Black virtual body leads to a sustained reduction in implicit racial bias. *Frontiers in Human Neuroscience* 10 (2016), 601. doi:10.3389/fnhum.2016.00601

[7] J. T. Bell and H. S. Fogler. 1995. The investigation and application of virtual reality as an educational tool. In *Proceedings of the American Society for Engineering Education Annual Conference, Session 2513*. Anaheim, CA.

[8] D. J. Bem. 1972. Self-perception theory. In *Advances in experimental social psychology*, Elsevier (Ed.). Vol. 6. Academic Press, 1–62. doi:10.1016/S0065-2601(08)60024-6

[9] K. Bessière, A. F. Seay, and S. Kiesler. 2007. The ideal elf: Identity exploration in World of Warcraft. *CyberPsychology & Behavior* 10, 4 (2007), 530–5. doi:10.1089/cpb.2007.9994

[10] Max V Birk, Cheralyn Atkins, Jason T Bowey, and Regan L Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. *CHI* (2016). doi:10.1145/2858036.2858062

[11] María J. Blanca, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* (2017). doi:10.7334/psicothema2016.383

[12] N. Bouali, E. Nygren, S. S. Oyelere, J. Suhonen, and V. Cavalli-Sforza. 2019. Imikode: A VR game to introduce OOP concepts. In *Koli Calling '19: Proceedings of the 19th Koli Calling International Conference on Computing Education Research*. ACM, Article 27, 2 pages. doi:10.1145/3364510.3366149

[13] V. Braun and V. Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[14] C. Butler, S. Michalowicz, L. Subramanian, and W. Burleson. 2017. More than a Feeling: The MiFace Framework for Defining Facial Communication Mappings. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 773–786. doi:10.1145/3126594.3126640

[15] B. Dalgarno and M. J. W. Lee. 2010. What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology* 41, 1 (2010), 10–32. doi:10.1111/j.1467-8535.2009.01038.x

[16] Edward Deci and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press.

[17] T. D. Do, C. I. Protko, and R. P. McMahan. 2024. Stepping into the right shoes: The effects of user-matched avatar ethnicity and gender on sense of embodiment in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2434–2443. doi:10.1109/TVCG.2024.3372067

[18] T. D. Do, S. Zelenty, M. Gonzalez-Franco, and R. P. McMahan. 2023. VALID: A perceptually validated Virtual Avatar Library for Inclusion and Diversity. *Frontiers in Virtual Reality* 4 (2023), 1248915. doi:10.3389/frvir.2023.1248915

[19] A. Drobnjak and I. Boticki. 2024. Technology Considerations in Building Virtual Educational Avatars. In *International Conference on Computers in Education*. doi:10.58459/icce.2024.4887

[20] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2006. "Alone Together?": Exploring the Social Dynamics of Massively Multiplayer Online Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Montréal Québec Canada, 407–416. doi:10.1145/1124772.1124834

[21] G. Falloon. 2010. Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology* 41, 1 (2010), 108–122. doi:10.1111/j.1467-8535.2009.00991.x

[22] C. M. Fang, P. Chua, S. W. T. Chan, J. Leong, A. Bao, and P. Maes. 2025. Leveraging AI-generated emotional self-voice to nudge people towards their ideal selves. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM. doi:10.1145/3596458.3713559

[23] M. C. Fink, S. A. Robinson, and B. Ertl. 2024. AI-based avatars are changing the way we learn and teach: Benefits and challenges. *Frontiers in Education* 9, Article 1416307 (2024). doi:10.3389/feduc.2024.1416307

[24] J. Fordham, R. Ratan, K.-T. Huang, and K. Silva. 2020. Stereotype threat in a video game context and its influence on perceptions of science, technology, engineering, and mathematics (STEM): Avatar-induced active self-concept as a possible mitigator. *American Behavioral Scientist* 64, 7 (2020), 900–926. doi:10.1177/0002764220919148

[25] J. P. Gee. 2003. *What video games have to teach us about learning and literacy*. Palgrave Macmillan.

[26] S. Guo, M. Choi, D. Kao, and C. Mousas. 2024. Collaborating with my doppelgänger: The effects of self-similar appearance and voice of a virtual character during a jigsaw puzzle co-solving task. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 7, 1, Article 4 (2024), 23 pages. doi:10.1145/3651288

[27] L. Hasenbein, P. Stark, U. Trautwein, A. C. M. Queiroz, J. Bailenson, J.-U. Hahn, and R. Göllner. 2022. Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. *Computers in Human Behavior* 133 (2022), 107282. doi:10.1016/j.chb.2022.107282

[28] Andrew F. Hayes. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

[29] D. Hepperle, C. F. Purps, J. Deuchler, and M. Wölfel. 2022. Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. *The Visual Computer* 38, 4 (2022), 1227–1244.

[30] Y.-H. Hu, H.-Y. Yu, J.-W. Tzeng, and K.-C. Zhong. 2023. Using an avatar-based digital collaboration platform to foster ethical education for university students. *Computers & Education* 196 (2023), 104728. doi:10.1016/j.compedu.2023.104728

[31] M. Javier, S. Wheeler, Q. Ni, M. Drivas, K. G. Wong, S. Fitzsimmons, R. Panda, T. Burke, S. Salazar, J. Jiang, L. C. Miller, and D. C. Jeong. 2024. The Alt-Self: Investigating the Inclusivity of Self-Avatar Representations in Social VR. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play*. 131–138. doi:10.1145/3665463.3678783

[32] D. Kao. 2018. Researching and developing the impacts of virtual identity on computational learning environments. https://api.semanticscholar.org/CorpusID:197662255

[33] D. Kao. 2019. JavaStrike: A Java programming engine embedded in virtual worlds. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. ACM, Article 58, 5 pages. doi:10.1145/3337722.3341828

[34] D. Kao and D. F. Harrell. 2015. Exploring the impact of role model avatars on game experience in educational games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, 571–576. doi:10.1145/2793107.2810291

[35] D. Kao and D. F. Harrell. 2018. The effects of badges and avatar identification on play and making in educational games. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Article 600, 19 pages. doi:10.1145/3173574.3174174

[36] D. Kao, S. T. Mubarrat, A. Joshi, S. Pandita, C. Mousas, H.-N. Liang, and R. Ratan. 2024. Exploring how gender-anonymous voice avatars influence women's performance in online computing group work. *International Journal of Human-Computer Studies* 181 (2024), 103146. doi:10.1016/j.ijhcs.2023.103146

[37] D. Kao, R. Ratan, C. Mousas, A. Joshi, and E. F. Melcer. 2022. Audio matters too: How audial avatar customization enhances visual avatar customization. In *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, Article 165, 27 pages. doi:10.1145/3491102.3501848

[38] D. Kao, R. Ratan, C. Mousas, and A. J. Magana. 2021. The effects of a self-similar avatar voice in educational games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY, Article 238 (2021), 28 pages. doi:10.1145/3474665

[39] L. K. Kaye, C. R. Pennington, and J. J. McCann. 2018. Do casual gaming environments evoke stereotype threat? Examining the effects of explicit priming and avatar gender. *Computers in Human Behavior* 78 (2018), 142–150. doi:10.1016/j.chb.2017.09.031

[40] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387. doi:10.1162/PRES_a_00124

[41] B. Kim, E. Hwang, and J. Lee. 2024. Self-Avatar Recognition in Virtual Reality: The Self Advantage Phenomenon and the Relative Importance of Motion and Visual Congruence. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 643–651. doi:10.1109/ISMAR62088.2024.00079

[42] Changsoo Kim, Sang-Gun Lee, and Minchoel Kang. 2012. I Became an Attractive Person in the Virtual World: Users' Identification with Virtual Communities and Avatars. *Computers in Human Behavior* 28, 5 (Sept 2012), 1663–1669. doi:10.1016/j.chb.2012.04.004

[43] H. Kim, J. Park, and I. K. Lee. 2023. "To be or not to be me?": Exploration of self-similar effects of avatars on social virtual reality experiences. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4794–4804. doi:10.1109/

TVCG.2023.3320240

[44] Jaehyeon Kim, Sungwon Kim, Jungil Kong, and Sungroh Yoon. 2020. Glow-TTS: A Generative Flow for Text-to-Speech via Monotonic Alignment Search. arXiv:2005.11129 [eess.AS] https://arxiv.org/abs/2005.11129

[45] C. Klimmt, D. Hefner, and P. Vorderer. 2010. The video game experience as "true" identification: A theory of enjoyable alterations of players' self-perception. *Communication Theory* 19, 4 (2010), 351–373. doi:10.1111/j.1468-2885.2009.01347.x

[46] M. Kwak, K. M. Yurov, and K. S. Floyd. 2015. A 3D learning game to foster computational thinking in K-12 education. *Issues in Information Systems* 16, III (2015), 20–29. doi:10.48009/4_iis_2015_20-29

[47] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4 (2013), 863.

[48] Luchcha Lam, Minsoo Choi, Magzhan Mukanova, Klay Hauser, Fangzheng Zhao, Richard Mayer, Christos Mousas, and Nicoletta Adamo-Villani. 2023. Effects of body type and voice pitch on perceived audio-visual correspondence and believability of virtual characters. In *ACM Symposium on Applied Perception 2023*. 1–11.

[49] Yinghao Aaron Li, Cong Han, Vinay S. Raghavan, Gavin Mischler, and Nima Mesgarani. 2023. StyleTTS 2: Towards Human-Level Text-to-Speech through Style Diffusion and Adversarial Training with Large Speech Language Models. (2023). arXiv:2306.07691 [eess.AS] https://arxiv.org/abs/2306.07691

[50] Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior* 25, 3 (2009), 695–710.

[51] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport* 60, 1 (1989), 48–58.

[52] R. M. Montoya, R. S. Horton, J. L. Vevea, M. Citkowicz, and E. A. Lauber. 2017. A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological Bulletin* 143, 5 (May 2017), 459–498. doi:10.1037/bul0000085

[53] C. Nass, Y. Moon, and N. Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology* 27, 10 (1997), 864–876. doi:10.1111/j.1559-1816.1997.tb00275.x

[54] M. Niwa, K. Masai, S. Yoshida, and M. Sugimoto. 2023. Investigating effects of facial self-similarity levels on the impression of virtual agents in serious/non-serious contexts. In *AHs '23: Proceedings of the Augmented Humans International Conference 2023*. ACM, 221–230. doi:10.1145/3582700.3582721

[55] O. Oyedokun, S. T. Mubarrat, A. Joshi, C. Mousas, and D. Kao. 2024. Exploring the Influence of Avatar Skin Tone in VR Educational Games. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play*. 227–234. doi:10.1145/3665463.3678799

[56] P. Pataranutaporn, V. Danry, J. Leong, P. Punpongsanon, D. Novy, P. Maes, and M. Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3 (2021). Issue 12. doi:10.1038/s42256-021-00417-9

[57] T. C. Peck and M. Gonzalez-Franco. 2021. Avatar Embodiment. A Standardized Questionnaire. *Frontiers in Virtual Reality* 1 (2021), 575943. doi:10.3389/frvir.2020.575943

[58] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. 2013. Putting yourself in the skin of a Black avatar reduces implicit racial bias. *Consciousness and Cognition* 22, 3 (2013), 779–787. doi:10.1016/j.concog.2013.04.016

[59] A. K. Przybylski, N. Weinstein, K. Murayama, M. F. Lynch, and R. M. Ryan. 2012. The ideal self at play: The appeal of video games that let you be all you can be. *Psychological Science* 23, 1 (2012), 69–76.

[60] Rabindra Ratan, David Beyea, Benjamin J. Li, and Luis Graciano. 2020. Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: a meta-analysis of the proteus effect. *Media Psychology* 23, 5 (2020), 651–675. doi:10.1080/15213269.2019.1623698

[61] C. V. Russoniello, K. O'Brien, and J. M. Parks. 2009. EEG, HRV and psychological correlates while playing Bejeweled II: A randomized controlled study. In *Studies in Health Technology and Informatics*, Vol. 144. 189–192.

[62] R M Ryan and E L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist* 55, 1 (2000), 68–78. doi:10.1037/0003-066X.55.1.68 arXiv:0208024 [gr-qc]

[63] Richard M Ryan, Edward L Deci, et al. 2002. Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research* 2, 3-33 (2002), 36.

[64] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (2006), 344–360. doi:10.1007/s11031-006-9051-8

[65] P. Salehi, S. Z. Hassan, S. Shafiee Sabet, G. Astrid Baugerud, M. Sinkerud Johnson, P. Halvorsen, and M. A. Riegler. 2022. Is More Realistic Better? A Comparison of Game Engine and GAN-based Avatars for Investigative Interviews of Children. In *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*. ACM, 41–49. doi:10.1145/3512731.3534209

[66] J. A. Sereno, A. Jongman, Y. Wang, P. Tupper, D. M. Behne, J. Gu, and H. Ruan. 2025. Expectation of speech style improves audio-visual perception of English vowels. *Speech Communication* 171 (2025), 103243. doi:10.1016/j.specom.2025.103243

[67] Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, Rj Skerrv-Ryan, Rif A. Saurous, Yannis Agiomvrgiannakis, and Yonghui Wu. 2018. Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Calgary, AB, Canada). IEEE Press, 4779–4783. doi:10.1109/ICASSP.2018.8461368

[68] Singular Inversions Inc. 2025. FaceGen Modeller. https://facegen.com/ Version 3.10 (or specify your version if known). Software available at https://facegen.com/.

[69] Kil-Soo Suh, Hongki Kim, and Eung Kyo Suh. 2011. What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIS Quarterly* (2011), 711–729.

[70] Selen Turkay and Charles K. Kinzer. 2017. The Relationship between Avatar-Based Customization, Player Identification, and Motivation. 48–79. doi:10.4018/978-1-5225-1817-4.ch003

[71] Jan Van Looy, Cédric Courtois, Melanie De Vocht, and Lieven De Marez. 2012. Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology* 15, 2 (2012), 197–221. doi:10.1080/15213269.2012.674917

[72] Henriette C Van Vugt, Jeremy N Bailenson, Johan F Hoorn, and Elly A Konijn. 2008. Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 2 (2008), 1–27.

[73] A. Vasalou, A. Joinson, T. Bänziger, P. Goldie, and J. Pitt. 2008. Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human-Computer Studies* 66, 11 (2008), 801–811. doi:10.1016/j.ijhcs.2008.08.002

[74] Asimina Vasalou, Adam N. Joinson, and Jeremy Pitt. 2007. Constructing my online self: avatars that increase self-focused attention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 445–448. doi:10.1145/1240624.1240696

[75] K. A. Wahba, K. A. Ahmed, M. R. Kamel, M. Fathy, P. K. H. Abdelfatah, and S. Hatem. 2024. Creating a digital human twin: Cloning voice, face, and attitude. In *Proceedings of the 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. IEEE, 199–205. doi:10.1109/MIUCC62295.2024.10783609

[76] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1643–1652.

[77] Y. Wang, L. Wang, and K. L. Siau. 2025. Human-Centered Interaction in Virtual Worlds: A New Era of Generative Artificial Intelligence and Metaverse. *International Journal of Human–Computer Interaction* 41, 2 (2025), 1459–1501. doi:10.1080/10447318.2024.2316376

[78] C. Xia, X. Fang, R. Arakawa, and Y. Sugiura. 2022. VoLearn: A cross-modal operable motion-learning system combined with virtual avatar and auditory feedback. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2, Article 81 (2022), 26 pages. doi:10.1145/3534576

[79] N. Yee, J. N. Bailenson, and N. Ducheneaut. 2009. The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research* 36, 2 (2009), 285–312. doi:10.1177/0093650208330254

[80] R. B. Zajonc. 2001. Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science* 10, 6 (2001), 224–228. doi:10.1111/1467-8721.00154 (Original work published 2001).

[81] H. Zhang, Z. Meng, M. Luo, H. Han, L. Liao, E. Cambria, and H. Fei. 2025. Towards multimodal empathetic response generation: A rich text-speech-vision avatar-based benchmark. In *WWW '25: Proceedings of the ACM on Web Conference 2025*. ACM, 2872–2881. doi:10.1145/3696410.3714739

[82] Yang Zhou, Xintong Han, Eli Shechtman, Jose Echevarria, Evangelos Kalogerakis, and Dingzeyu Li. 2020. MakeItTalk - SA'20. https://people.umass.edu/~yangzhou/MakeItTalk/