CREATIUM

Research

Evaluating Hyper-Realistic AI Avatars in Instruction: Effects on Learning and Engagement

Dr. Kelly Puzio

Director of Research, Creatium

Abstract

This experiment investigated the extent to which adding a hyper-realistic AI avatar to a narrated instructional video yielded learning benefits over audio narration alone. U.S.-based teachers (N = 245) were randomly assigned to one of three conditions: PowerPoint-only (descriptive baseline), Creatium video with audio narration, or Creatium video with avatar narration. The primary analyses contrasted the audio group (n = 84) with the avatar group (n = 80), using pretest knowledge as a covariate. The AI avatar condition demonstrated superior learning efficiency: participants achieved 5–6 percentage points higher posttest scores while spending 9% less time learning. In statistical models, the avatar advantage—characterized by better performance despite shorter viewing time—yielded a modest effect size ($g = +0.26$) that approached statistical significance ($p = .059$). Engagement and NPS were statistically equivalent between formats, suggesting that the presence of the avatar did not diminish the overall user experience. Descriptive comparisons confirmed the expected modality progression (PowerPoint-only < audio narration < avatar narration). These findings provide evidence that Creatium's current-generation AI avatars can enhance both learning effectiveness and efficiency without compromising the overall learning experience, highlighting their potential for scalable, time-conscious professional learning initiatives.

Background

As organizations invest significant resources in digital learning infrastructure, a critical question emerges: can AI-generated learning experiences match or exceed the effectiveness of human teaching? This study examines one aspect of this frontier—hyper-realistic AI avatars—to understand their impact on learning outcomes, with particular attention to how prior knowledge influences their effectiveness.

Throughout this paper, we use terminology to describe different types of virtual instructors. Pedagogical agents encompass all virtual characters designed for instruction, from simple animations to complex digital humans. Avatars specifically refer to visual representations of instructors or learners in educational media. Hyper-realistic AI avatars—the focus of this study—use artificial intelligence to generate lifelike digital instructors with synthesized and synchronized speech, natural movements, and human-like appearance.

The Cognitive Theory of Multimedia Learning (CTML) conjectures that pairing spoken narration with relevant visuals can reduce cognitive load and foster deeper processing through dual channels (Mayer, 2009). Social Agency Theory extends this framework, proposing that human-like cues (e.g., conversational voice, facial expressions) activate social schemas that boost motivation and learning (Mayer, Sobko, & Mautone, 2003). Supporting evidence includes laboratory studies (Wang & Antonenko, 2017) and MOOC analytics from 6.9 million sessions showed that student engagement (viewing time) and problem-solving increased when students see an instructor in videos (Guo, Kim, & Rubin, 2014).

However, pedagogical agent effectiveness has remained inconsistent. An early review found mixed results, noting that poorly designed agents became "costly distractors" (Heidig & Clarebout, 2011). A recent meta-analysis (N = 2,104) reported modest gains ($g = +0.20$) for multimedia agents, with simple 2-D figures outperforming elaborate 3-D avatars ($g = +0.38$ vs. $+0.11$)—supporting cognitive load warnings against unnecessary visual detail (Castro-Alonso, Wong, Adesope, & Paas, 2021). In one study manipulating an agent's embodiment, learners who interacted with a fully embodied animated agent outperformed those in a voice-only condition on both near- and far-transfer tests, suggesting that coordinated verbal and nonverbal cues can enhance problem solving (Lusk & Atkinson, 2007).

Yet newer research suggests that design quality matters more than realism alone. According to Kerac et al. (2025), a literature review of avatar-design in e-learning found that students prefer pedagogical agents with natural, human-like movements, and that knowledge acquisition was enhanced when those nonverbal cues were well-designed. Collectively, these findings indicate that current AI avatar technology may have reached a level of sophistication where implementation quality—rather than merely pursuing photorealism—determines effectiveness.

The effectiveness of AI avatars depends significantly on learners' existing knowledge, with benefits decreasing or even reversing as expertise increases. Cognitive Load Theory's expertise-reversal principle indicates that supports that help novices can frustrate or burden experts (Kalyuga, Ayres, Chandler, & Sweller, 2003). Thus, visible agents might aid novice learning and—at the same time—impede experts who process content efficiently.

Beyond visual design, the quality of avatar-learner interaction and the degree of human realism present distinct challenges for educational effectiveness**.** A systematic review of 31 studies concluded that agent-student dialogue could strengthen learner self-regulation and motivation, though current agents lacked human teachers' adaptive capabilities (Sikström, Valentini, Sivunen, & Kärkkäinen, 2022). Additionally, avatars approaching but not achieving human realism may trigger "uncanny valley" discomfort that offsets social-presence benefits (Mori, MacDorman, & Kageki, 2012).

Current evidence suggests three core predictions about AI avatars in instruction: First, visible avatars may enhance learning over audio-only narration by engaging social and cognitive processes. Second, avatar impact varies by prior knowledge, with larger benefits for novices and potential drawbacks for experts. Third, effective avatars must avoid unnecessary visual complexity that distracts from content. Despite growing interest, few studies have systematically examined these variables using today's hyper-realistic AI avatars. This experiment addresses this gap by comparing audio-narrated and avatar-narrated videos—both using identical AI-generated content—to determine when and for whom avatars improve professional learning outcomes.

Research Methods

This investigation examined the impact of AI avatars in video-based instruction with a focus on learning outcomes, engagement, and recommendation intention. The following questions guided this investigation:

1. To what extent does avatar-based narration affect learning outcomes compared to audio-only narration?
2. Does prior knowledge level moderate the effectiveness difference between avatar-based and audio-only narration on learning outcomes?
3. Does avatar-based narration result in higher learner engagement and satisfaction compared to audio-only narration, as measured by participant engagement and Net Promoter ratings?

Study Design and Participants

This study used a between-subjects design in which 245 teachers recruited through Prolific were randomly assigned to one of three delivery formats. The PowerPoint-only group (PowerPoint) received a static slide deck, served as a descriptive baseline for the well-documented modality effect, and was reported for contextual reference only. The two focal groups both viewed Creatium videos featuring proprietary AI technology: one with AI-generated voice narration presenting the slides (Audio) and an identical video augmented with Creatium's hyper-realistic, AI avatar (Avatar).

All participants were U.S.-based English-speaking teachers, with one self-reporting birth in Jamaica, one in Canada, and two in Vietnam. The data for 10 users were excluded from the analytic sample due to incomplete assessment data, failed attention checks, or user reports of not completing the learning module. Table 1 presents the demographic characteristics of participants by condition:

Table 1
Participant Characteristics, by Condition

Characteristic	**PowerPoint** (*n* = 71)	**Audio** (*n* = 84)	**Avatar** (n = 80)
Age (Mean, SD)	42.17 (14.22)	41.67 (13.03)	39.39 (10.33)
Sex (% Female)	73%	75%	78%
Ethnicity (% White)	77%	74%	71%
Education (% Graduate Degrees)	69%	71%	68%
Employment (% Full-Time)	85%	83%	87%

As shown in Table 1, the two experimental groups were comparable across demographic variables. Participants in both conditions were predominantly female (73-78%), primarily White (71-77%), and employed full-time (83-87%), with a mean age of approximately 41 years. In addition, most participants had graduate degrees (68-71%).

Task and Conditions

Participants in all conditions received instructional content on interacting with different personality types and strategies for effective communication across personality differences. The experiment consisted of three conditions. In the PowerPoint condition, participants received a PDF of a PowerPoint presenting information about personality types and communication strategies. This slide deck contained text and images on each slide. In the audio condition, participants viewed a Creatium video featuring the same content, where the video was narrated by Creatium's proprietary male AI-generated voice to provide more details and context about the information presented on the slides. In the Avatar condition, participants viewed a Creatium video that was identical to the Audio condition with the addition of Creatium's hyper-realistic male AI avatar. The only difference between the Audio and Avatar conditions was the inclusion of the AI avatar that was present on every slide and whose mouth was engineered to lip-sync the narrated words using Creatium's technology.

Data Collection

All data were collected through a survey instrument hosted by the SurveyMonkey platform. Prior to instruction, all participants completed a multiple-choice pretest measuring their baseline knowledge of personality types and communication strategies. Following the instructional intervention, participants completed: (1) a different multiple-choice posttest to measure their knowledge of personality types and communication strategies, (2) an engagement measure (How engaging did you find this learning experience?), and (3) a Net Promoter Score (NPS) measure on a 0-10 scale. Prior to this experiment, both knowledge assessments were pilot tested with different participants and aligned with the content presented in the videos and interactive learning resources. Test scores were calculated as the proportion of correct responses (0-1).

Analysis

Data from the audio-only (n = 84) and avatar (n = 80) groups were analyzed with three sequential models. First, an ANCOVA predicted posttest percentage from instructional condition while controlling for pretest knowledge. Second, the model was expanded to include a Condition × Pretest interaction to test for expertise reversal. Third, time was entered as an additional covariate to control for differential viewing duration. Engagement and NPS were examined simultaneously with a MANCOVA (covariate = pretest).

Results

Descriptive Statistics

Table 2 presents the means and standard deviations for all measures across all experimental conditions.

Table 2
Descriptive Statistics, by Condition

Measure	**PowerPoint** (*n* = 71)	**Audio** (*n* = 84)	**Avatar** (*n* = 80)
Pretest	45.95 (17.90)	45.09 (21.43)	45.00 (20.15)
Posttest	71.13 (21.15)	75.00 (21.82)	79.88 (20.34)
Engagement	3.56 (1.42)	3.96 (1.40)	3.45 (1.28)
Net Promoter	5.54 (2.92)	6.35 (2.86)	5.32 (2.61)
Time (min)	14.00 (6.08)	18.27 (7.08)	16.55 (5.90)

As shown in Table 2, pretest scores were comparable across all three conditions (PowerPoint: 45.95%, Audio: 45.09%, Avatar: 45.00%), indicating successful randomization. Posttest scores showed an increasing trend from PowerPoint (71.13%) to Audio (75.00%) to Avatar (79.88%), suggesting potential differences in the instructional effectiveness of these different conditions.

Learning Outcomes and Efficiency

Analyses investigated whether adding an AI avatar to narrated video improved learning relative to audio-only narration. The sample comprised 84 participants in the audio condition (coded 0) and 80 in the avatar condition (coded 1). Pretest means were virtually identical—45.09 (SD = 21.43) for audio and 45.00 (SD = 20.15)—$t(162) = 0.027$, $p = .978$, indicating successful random assignment. Posttest means favored the avatar condition, 79.88 (SD = 20.34) versus 75.0 (SD = 21.8), and the two groups viewed the lesson for 16.55 (SD = 5.90) and 18.3 (SD = 7.08) minutes, respectively.

Table 3 (Model 1) presents the primary ANCOVA in which posttest score was regressed on condition with pretest performance as a covariate. Prior knowledge strongly predicted learning, $\beta = 0.56$, $SE = 0.09$, $t(161) = 6.24$, $p < 0.001$. After adjustment, the avatar produced a 4.91-point advantage, $\beta = 4.91$, $SE = 2.96$, $t(161) = 1.66$, $p = .10$, corresponding to a small effect size ($g = +0.26$, $R^2 = 0.196$).

Table 3
Hierarchical Regression Predicting Posttest Performance

Predictor	**Model 1**	**Model 2**	**Model 3**
Intercept	51.91 (3.98)*	51.74 (4.06)*	46.68 (4.71)*
Pretest (%)	0.56 (0.09)**	0.55 (0.09)**	0.50 (0.09)**
Condition	4.91 (2.96)†	4.83 (2.98)†	6.08 (3.20)†
Pre×Condition	—	0.02 (0.13)	0.02 (0.13)
Time (min)	—	—	0.38 (0.17)*
R^2	0.196	0.196	0.228

Note. Model 1 controls for pretest score; Model 2 adds the Condition × Pretest interaction; Model 3 adds time. Unstandardized coefficients are shown with standard errors.
*n = 164. †p < .10; *p < .05; **p < .001.*

To examine expertise reversal, a Condition × Pretest interaction was added (Model 2). The interaction was negligible, $\beta = 0.02$, $SE = 0.13$, $t(160) = 0.14$, $p = .89$, showing that the avatar advantage did not depend on prior knowledge.

Because learning time varied by condition, viewing minutes were included as an additional covariate (Model 3). Time positively predicted achievement, $\beta = 0.38$, $SE = 0.17$, $t(159) = 2.23$, $p = .027$, and slightly increased the avatar coefficient to 6.08 points, $\beta = 6.08$, $SE = 3.20$, $t(159) = 1.90$, $p = .059$. Thus, the treatment effect persisted and approached statistical significance (p = .059) even after controlling for time differences.

A multivariate ANCOVA evaluated engagement and NPS outcomes. With pretest as covariate, the combined test for engagement and NPS was non-significant, Wilks $\Lambda = 0.986$, $F(2, 160) = 1.14$, $p = .32$; univariate contrasts were likewise trivial (engagement $p = .47$; NPS $p = .63$), and adding time did not alter this pattern.

Overall, the AI avatar condition yielded encouraging results, with learners scoring 5–6 percentage points higher on the posttest than the audio-only group despite spending 1.7 minutes less time on the learning module (16.6 vs. 18.3 minutes). The avatar advantage—characterized by better performance despite shorter viewing time—yielded a modest effect size ($g = +0.26$) and approached statistical significance ($p = .059$) in models adjusting for prior knowledge and time-on-task. Engagement and NPS were statistically equivalent between formats, indicating that the learning gains did not come at the cost of user experience.

Limitations

Several limitations constrain the interpretation of these findings. The study examined only a single content domain delivered in a brief, one-time instructional session, precluding examination of effectiveness across different disciplines or potential novelty effects that may diminish with repeated exposure. The AI avatar was limited to a single male presenter with specific visual and vocal characteristics, preventing examination of how avatar demographics, appearance styles, or voice qualities might moderate learning outcomes. Additionally, participants were not informed whether they were viewing an AI-generated or human instructor, leaving unanswered how transparency about AI use might affect learning outcomes and trust. The reliance on self-reported engagement measures and single-administration knowledge assessments limits understanding of deeper learning processes and long-term retention. Finally, the sample consisted exclusively of teachers recruited through Prolific, whose pedagogical knowledge and instructional design experience may have influenced how they perceived and processed AI avatar instruction differently than typical adult learners, potentially limiting generalizability to broader contexts.

Discussion

The present study investigated the extent to which layering an AI avatar onto an otherwise identical, AI-narrated video influenced learning in terms of knowledge and efficiency. Controlling for prior knowledge, teachers who learned with the avatar scored about 4.9 percentage points higher on the posttest than peers who heard narration alone. Remarkably, this learning advantage occurred despite avatar viewers spending 1.7 minutes less time with the material (16.6 vs. 18.3 minutes), a 9% reduction in viewing time. When time differences were controlled for, the avatar advantage increased to 6.1 percentage points and approached statistical significance (p = .059). The direction and magnitude align with Social Agency Theory's claim that even subtle social cues can foster deeper cognitive processing. The simultaneous improvement in both effectiveness (higher scores) and efficiency (less time) challenges conventional wisdom that more time-on-task necessarily yields better outcomes. Importantly, the avatar did no harm: engagement and NPS ratings were statistically indistinguishable from the audio condition. These findings suggest the technology has room to mature but is already a viable enhancement that can deliver better results while respecting learners' time constraints.

From a theoretical lens, the results offer some encouragement. The nonsignificant interaction with prior knowledge indicates that the avatar's benefits were not confined to novices, hinting that a broad range of learners may find at least some value in a visible speaker. Given the professional sophistication of the teacher sample, it is noteworthy that any incremental gain emerged at all; more heterogeneous or less experienced audiences might exhibit a larger advantage. While technically advanced, the avatar used did not utilize any sophisticated gesturing. Social-presence research shows that richer nonverbal signals (e.g., signaling gestures) amplify learners' sense of partnership and can translate into stronger cognitive outcomes. The small but positive trend we observed could therefore be viewed as an early indicator of what more expressive, adaptive avatars might achieve.

From a practical perspective, these findings suggest that AI avatars represent a viable alternative to traditional video instruction, particularly for professional development. The 5-6% advantage in posttest scores for avatar instruction, while modest, could translate to meaningful improvements when scaled across large professional organizations. The cost-effectiveness of AI avatar generation compared to traditional video production makes this approach particularly attractive for organizations needing to rapidly develop and update training materials. Organizations should consider AI avatars as one component of a comprehensive instructional design strategy rather than a standalone solution for enhancing learner engagement.

The study's results both align with and diverge from previous research in important ways. The overall positive effect of AI avatars on learning

($g = +0.26$) falls within the range reported in recent meta-analyses of pedagogical agents; for example, Castro-Alonso et al. (2021) reported an average gain of $g = +0.20$, with larger effects for simpler, 2-D agents. However, the absence of engagement benefits contrasts with Guo et al.'s (2014) MOOC research showing 30% higher engagement with visible instructors. This discrepancy may reflect differences between AI-generated and human instructors, the specific content domain, or measurement approaches. The finding that audio narration alone provided benefits over PowerPoint-only instruction underscores the continued importance of the modality principle, regardless of visual social presence.

Future research should address several questions raised by this study. First, investigations should examine how learner awareness of AI versus human instruction affects outcomes, as transparency may influence trust and processing strategies. Second, studies should explore optimal AI avatar characteristics across different content domains and learner populations, including potential matching effects between avatar and learner demographics. Third, longitudinal research is needed to understand whether novelty effects influence short-term responses to AI avatars and how effectiveness changes with repeated exposure. Additionally, process-oriented research using eye-tracking, physiological measures, and think-aloud protocols could illuminate the cognitive and affective mechanisms underlying AI avatar effects. As AI technology continues to advance, enabling more sophisticated and personalized avatar behaviors, research must keep pace to guide evidence-based implementation in professional learning contexts.

References

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review, 33*, 989–1015.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the First ACM Conference on Learning@ Scale* (pp. 41–50). Association for Computing Machinery.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27–54.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.

Kerac, J., Golubović, G., Milić Keresteš, N., & Ilić, T. (2025). The effects of avatar design on e-learning: A review. In N. Martins & D. Brandão (Eds.), *Advances in Design and Digital Communication V* (pp. 673-686). Springer.

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology, 21*(6), 747–764.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*(2), 419–425.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100.

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, Article 104564.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89.

Author note

Kelly Puzio, PhD, is the Director of Research at Creatium. Dr. Puzio is a learning scientist whose work focuses on personalized learning and artificial intelligence. He earned a PhD in Learning, Teaching, and Diversity from Vanderbilt University, a master's degree in education from DePaul University, and a bachelor's degree from the University of Notre Dame. He has been a certified teacher in New Zealand and the United States and brings experience as a faculty member in the Department of Teaching and Learning at Washington State University. Dr. Puzio has served as Principal Investigator on more than $5 million in funding from the National Institutes of Health and the National Science Foundation, and he has been recognized as a Young Scholar by the International Literacy Association.