CREATIUM

Research

Transforming instructional video production:

A comprehensive study of Creatium's AI platform Versus Traditional Methods.

Dr. Kelly Puzio
Director of Research, Creatium

Transforming Instructional Video Production: A Comparative Study of Creatium's AI Platform Versus Traditional Methods

Kelly Puzio, PhD

Abstract

This investigation examined the impact of Creatium's AI platform on instructional video production efficiency and quality through a controlled experiment. Sixty teachers living in the United States were randomly assigned to either create instructional videos using traditional methods or using the AI-powered Creatium platform, with both groups working from an identical article on digital twins and simulation technology. Results revealed large differences in both production efficiency and output quality. Participants using Creatium's platform completed their video presentations in approximately one-fifth the time required by the traditional approach (10.63 vs. 49.27 minutes) while simultaneously achieving significantly higher quality scores across all dimensions measured (visual, audio, and information quality). Statistical analysis revealed large effect sizes for all variables ($\eta^2 = 0.18$-0.87), with the most substantial advantages observed in production time and information quality. Regression modeling further predicted that participants using traditional methods would require an additional 38.3 minutes on average to achieve quality parity with Creatium's AI-assisted outputs, resulting in a total production time more than eight times longer than the AI approach. These findings have significant implications for workforce development, PK-20 learning, and corporate learning contexts, where producing high-quality instructional content is essential. The substantial efficiency gains demonstrated by Creatium's AI platform could translate to thousands of saved labor hours annually while simultaneously establishing more consistent quality standards across learning resources.

Background

The landscape of instructional design has evolved significantly in recent decades, with video-based instruction becoming increasingly central to educational practices across contexts (Mayer & Fiorella, 2021). A systematic review by Poquet et al. (2022) spanning a decade of research found that effective instructional videos have become essential components of modern pedagogical approaches, with impact on engagement, knowledge retention, and skill development. However, educators face mounting pressure to produce these high-quality instructional videos despite constraints in time, resources, and content expertise (Martin et al., 2019). Traditional approaches to instructional video creation typically involve multiple labor-intensive stages—from content research and script development to visual design and recording—often requiring days or weeks to complete (Guo et al., 2014).

This time-intensive process presents a significant challenge for organizations that support learning and professional development, especially ones that seek to maintain rigorous quality standards. As Clark and Mayer (2016) established in their research on multimedia learning principles, effective instructional videos must balance cognitive load considerations, engagement factors, and content accuracy—a complex set of requirements that traditionally demands both pedagogical and technical expertise.

Artificial intelligence tools have begun transforming numerous aspects of educational practice, from assessment to personalization (Holstein et al., 2020). The integration of AI into instructional design workflows represents a potentially disruptive innovation that could fundamentally alter how educational content is created (Zawacki-Richter et al., 2019). Recent advances in generative AI systems suggest these technologies may dramatically compress production timelines while potentially enhancing quality through built-in pedagogical frameworks (Kasneci et al., 2023).

However, empirical research directly comparing AI-assisted and traditional approaches to instructional video creation remains limited. While theoretical frameworks suggest potential benefits, questions persist regarding the comparative quality of outputs, efficiency gains, and the relationship between production time and final quality. The proposed study is particularly timely as educational institutions increasingly seek evidence-based approaches to resource allocation in content development.

By quantifying both the time efficiency and quality differences between AI-assisted and traditional instructional video creation methods, this study will significantly advance our understanding of AI's practical impact on educational content development. The findings will provide critical insights for instructional designers, educators, and administrators making strategic decisions about technology adoption in educational contexts where both efficiency and quality are important concerns.

Research Methods

This investigation focused on examining the comparative effectiveness of Creatium's AI platform against traditional methods for creating instructional videos. We sought to understand how the AI platform affects production efficiency and output quality across multiple dimensions. Specifically, the following questions guided this investigation:

1. To what extent does the use of Creatium's AI platform affect the efficiency of instructional video production compared to traditional methods?
2. How does the quality of instructional videos created using Creatium's AI platform compare to those created using traditional methods across visual, audio, and information dimensions?
3. How much additional time would be required for videos created using traditional methods to reach quality parity with those created using AI assistance?

This study employed a between-subjects experimental design with 60 US-based teachers recruited through Prolific, randomly assigned to either a control group (traditional method, n=30) or experimental group (AI-assisted using Creatium, n=30). Participants in both conditions received the same informational text (500 words) on digital twins and simulation technology, a technical topic that would be unfamiliar to most teachers. We chose this topic to simulate the workflow of professional instructional designers, who are often tasked with creating instructional videos on topics they know very little about.

In both conditions, participants were asked to create an instructional video that 1) expressed the main ideas in the informational text, 2) included supporting images on each slide, and 3) narrated the slide presentation. Control condition participants could use any presentation software (e.g., PowerPoint, Google Slides, Canva) along with a voice- and screen-capturing software (e.g., Loom). Experimental condition participants completed these tasks using the Creatium AI platform. As an introduction to using a new software application, participants in the experimental condition also were provided with a brief video (2.5 minutes) to help locate and use key aspects of the Creatium AI content-creation platform. Participants submitted all their materials through an online survey along with background information about themselves (e.g., years/level of teaching). Time tracking was implemented through SurveyMonkey with triangulation from the Prolific platform.

Each video was evaluated for quality on visuals, audio, and information. For information quality, the evaluation focused on how well the video represented the claims, evidence, and academic language of the article by comparing the language used in the slide presentation and narration against the information in the article itself. For visual quality, the assessment checked if at least one visual was provided for each slide and whether the visual connected clearly to the slide information. Lastly, for audio quality, the evaluation involved listening to the presenter and qualitatively assessing whether the audio was clear, emotionally expressive, and free of background noise and/or static.

For videos scoring below production quality standards (below 8 on the quality scale), an adjusted time metric was calculated through regression analysis to estimate the additional time needed to bring the video to presentation-ready standards. This approach involved developing three separate regression models—one for each quality dimension (visual, audio, and information)—with slide creation time and narration time as predictor variables. The resulting equations were then used to calculate dimension-specific time deficits for each participant based on their current quality scores, which were summed to provide individualized estimates of the total additional time investment required to achieve the minimum quality threshold of 8 across all dimensions.

To analyze the results, we conducted a Multivariate Analysis of Variance (MANOVA) with condition (traditional vs. AI-assisted) as the independent variable and six dependent measures: three time metrics (slide creation time, narration time, and total time) and three quality metrics (visual quality, audio quality, and information quality). MANOVA was selected to account for potential correlations among the dependent variables while controlling for Type I error inflation that would occur with multiple separate tests. Following a significant MANOVA result (using Wilks' lambda criterion), we conducted follow-up univariate ANOVAs to determine which specific variables showed significant differences between conditions. Effect sizes were calculated using partial eta squared (η^2) to quantify the magnitude of differences for each outcome variable.

Participants

Participants included 60 currently employed US teachers randomly assigned to either traditional or AI-assisted instructional video creation conditions. As shown in Table 1, both groups were well-matched across key demographic and professional characteristics. Participants were predominantly female, with an average age in the mid-to-late 30s. Most identified as White (63-67%) and working as full-time teachers (70-80%). Both conditions included similar distributions of teaching experience, with slightly over half being veteran teachers (12+ years experience), and comparable representation across grade levels from elementary through postsecondary education.

Table 1
Participant Characteristics

Characteristic	Traditional (n=30)	AI Creatium (n=30)
Age (mean)	36.8 years	38.4 years
Gender	57% female	63% female
Ethnicity	63% White	67% White
Teaching Experience	57% Experienced (12+ years) 43% Early/Mid-career	53% Experienced (12+ years) 47% Early/Mid-career
Grade Level	47% K-8 23% High School 30% Postsecondary	47% K-8 27% High School 27% Postsecondary
Employment	80% Full-time	70% Full-time
Technology Experience	73% Moderate to High	70% Moderate to High
Prior Knowledge	7% Some familiarity 93% None	10% Some familiarity 90% None

Note. Demographic data on age, ethnicity, gender, and employment status were collected from Prolific; information on technology experience, prior knowledge, grade level taught, and teaching experience collected via survey.

Most participants reported moderate to high experience with instructional technology (70-73%), though very few had prior knowledge of digital twins and simulation technology (7-10%), confirming that this topic was indeed unfamiliar to the vast majority of participants as intended in the study design. No participants reported having average, above average, or expert knowledge of the topic (digital twins and simulation).

Results

Table 2 presents the means and standard deviations for the outcome variables by condition.

Between the traditional approach and Creatium AI approach, the results demonstrate differences across all metrics. Participants using AI Creatium completed their narrated presentations in approximately one-fifth the time required by those using traditional methods (10.63 vs. 49.27 minutes), representing an efficiency gain of nearly 80%. This substantial time advantage was consistent across both slide creation (8.77 vs. 42.43 minutes) and narration (1.87 vs. 6.83 minutes) phases.

Despite this dramatic reduction in production time, the AI-assisted approach yielded significantly higher quality scores across all dimensions. Visual quality showed the most dramatic improvement (8.33 vs. 4.80), followed by information quality (8.33 vs. 6.05) and audio quality (8.00 vs. 6.77). Notably, the AI Creatium condition also demonstrated much greater consistency in outcomes, as evidenced by the substantially smaller standard deviations across all metrics, particularly in information quality (0.31 vs. 1.69) and audio quality (0.00 vs. 2.71), suggesting that the AI-assisted approach not only improves efficiency and quality but also standardizes the production process.

Table 2

Descriptive Statistics, by Condition

	Traditional (n=30)	AI Creatium (n=30)
Time		
Slide Creation Time	42.43 (12.78)	8.77 (2.64)
Narration Time	6.83 (1.95)	1.87 (0.82)
Total Time	49.27 (13.21)	10.63 (3.28)
Quality		
Visual Quality	4.80 (3.32)	8.33 (1.47)
Audio Quality	6.77 (2.71)	8.00 (0.00)
Information Quality	6.05 (1.69)	8.33 (0.31)

Note. Values represent means with standard deviations in parentheses. All time metrics are measured in minutes. Quality metrics are scored on a scale of 1-10, with higher scores indicating better quality. The 0.00 standard deviation for audio quality in the Creatium condition reflects the AI platform's consistent audio generation capability, which produces standardized, high-quality narration for every presentation regardless of individual user variations.

Multivariate Analysis

Preliminary tests indicated violations of MANOVA assumptions, including multivariate normality (Mardia's test: skewness = 156.32, kurtosis = 498.17, $p < .001$) and homogeneity of variance-covariance matrices (Box's M = 312.64, $p < .001$). Given these conditions, Pillai's trace was selected as the most robust multivariate test statistic. Pillai's trace is a positive-valued statistic ranging from 0 to 1 that measures the proportion of variance in the dependent variables explained by the group factor, with values closer to 1 indicating stronger effects while maintaining robustness against assumption violations. MANOVA results revealed a significant multivariate effect of condition on the combined dependent variables, Pillai's trace = 0.95, $F(6, 53) = 162.43$, $p < .001$, indicating substantial differences between traditional and AI-assisted approaches.

Univariate Analyses

Follow-up ANOVAs with Welch's correction for unequal variances were conducted to examine the impact of condition on each dependent variable separately (Table 3).

All variables showed significant differences between conditions ($p < .05$). The largest effects were observed for time metrics, particularly slide creation time ($\eta^2 = 0.85$) and total time ($\eta^2 = 0.87$), indicating that 85% and 87% of the variance in these variables, respectively, could be attributed to the condition. Among quality metrics, information quality showed the largest effect ($\eta^2 = 0.64$), followed by visual quality ($\eta^2 = 0.40$) and audio quality ($\eta^2 = 0.18$).

Participants in the AI-assisted condition completed their presentations in approximately one-fifth the time required by participants in the traditional condition (10.63 vs. 49.27 minutes) while achieving significantly higher quality scores across all dimensions. The most dramatic improvement was observed in visual quality, with AI-assisted participants scoring 3.53 points higher on average (8.33 vs. 4.80).

Table 3

ANOVA Results and Effect Sizes

Dependent Variable	F	df	p	Partial η^2
Slide Creation Time	187.43	1, 33.24	< .001	0.85
Narration Time	162.97	1, 39.29	< .001	0.76
Total Time	211.76	1, 33.38	< .001	0.87
Visual Quality	29.75	1, 44.18	< .001	0.4
Audio Quality	6.45	1, 29.00	0.017	0.18
Information Quality	53.08	1, 30.41	< .001	0.64

Note. Welch's correction was applied to address the violation of homogeneity of variance assumption. The decimal points in the degrees of freedom (df) reflect this adjustment, which provides more accurate p-values when group variances differ substantially, as was the case in this study.

Predicted Time for High-Quality Presentation

Due to a variety of factors, participants using traditional methods to create instructional videos often did not produce a high-quality video production. For example, some presentations did not include any supporting images; in other cases, there was significant background noise in the environment. Due to this, the investigation predicted how long it might take a user to bring their video up to a standard of quality that would be worthy of publishing to an audience of potential learners.

Table 4

Predicted Additional Time Required by Quality Dimension

Quality Dimension	% Below Thres-hold	Avg. Predicted Additional Time	Avg. Time Across All Participants	Total Predicted Time
Visual	63%	25.9	16.3	
Audio	40%	12.2	4.9	
Information	77%	22.3	17.1	
Total			**38.3**	**87.6**

Note. "Average Time Across All Participants" represents each dimension's weighted contribution to the total average additional time. "Total Predicted Time" combines the current average time (49.27 min) with the regression predicted additional time needed.

To predict the time required for participants to achieve quality scores of 8 or higher across all dimensions, we employed a multiple regression approach using the traditional condition dataset (n=30). Three separate models were constructed for visual, audio, and information quality, with slide creation time and narration time as predictor variables. For each participant, we calculated the quality deficit for each dimension and generated regression-estimates for each individual on each quality indicator. This approach accounted for participants' current performance levels while providing both dimension-specific and aggregate additional time requirements.

Based on this individualized regression analysis, participants in the traditional condition would need to invest an average of 38.3 additional minutes to achieve quality scores comparable to those in the AI-assisted condition. Most participants (77%) required improvements in information quality, contributing 17.1 minutes to the overall average additional time. Visual quality improvements were needed by 63% of participants, contributing 16.3 minutes, while audio quality required the least additional time overall. The total predicted production time would be 87.6 minutes—more than eight times longer than the 10.63 minutes required using the AI-assisted approach. These estimates should be interpreted as conservative approximations of the additional time required to achieve quality parity with the AI-assisted approach.

Limitations

Despite the significant findings, several important limitations must be acknowledged. Our sample size of 60 participants (30 per condition), while sufficient for detecting large effects, limits our ability to conduct more nuanced subgroup analyses or account for individual differences in content creation aptitude and technical proficiency. Our study utilized a single technical article on digital twins as content material, which may not represent the full spectrum of topics addressed in workforce development, particularly in humanities, arts, or highly specialized domains. Perhaps most significantly, our experimental design used a written article as the starting point for content creation, which represents just one of many possible workflows in instructional design—in practice, diverse inputs such as learning standards, existing slide decks, or collaborative planning documents may be the starting point. Additionally, the regression models used to predict additional time requirements had relatively low R^2 values (0.04-0.11), indicating that factors beyond simple time likely influence quality outcomes. Despite these constraints, the magnitude of effects observed suggests that the core findings regarding efficiency and quality advantages are likely relevant across various training and development contexts.

Discussion

The results of this study provide empirical evidence of a transformative shift in instructional video production enabled by Creatium’s AI platform, with particularly profound implications for workforce development and adult learning contexts. The substantial reductions in production time (approximately 80%) coupled with significant improvements in quality across all measured dimensions address a critical challenge in corporate training and professional development: rapidly creating high-quality instructional content to meet evolving skill demands. These findings align with theoretical predictions by Kasneci et al. (2023) regarding AI's potential to compress production timelines while maintaining pedagogical quality and extend previous work by demonstrating that these benefits are achievable by individuals without specialized backgrounds in instructional design or content knowledge—a common scenario in workforce training environments.

From an organizational perspective, our regression analysis predicts that achieving quality parity through traditional methods would require approximately 87.6 minutes per video—over eight times longer than using AI assistance. For workforce development programs, this translates to potential time savings of nearly 77 minutes per video, which at enterprise scale could represent thousands of labor hours annually that could be redirected toward other high-value activities such as personalized coaching, skills assessment, or curriculum customization. The standardization evident in the AI condition's reduced variability across quality metrics indicates that AI tools may help establish consistent quality standards across diverse training materials—a particular advantage for large organizations with distributed content creation teams or industry associations developing standardized certification materials.

While our study utilized teachers as participants, the findings likely extend to corporate trainers, subject matter experts, and instructional designers in workforce settings where there is a regular need to develop high-quality instructional videos on technical topics. Creatium’s AI-assisted approach appears particularly valuable for addressing common challenges in professional training: creating engaging visuals, accurately translating complex content into accessible presentations, and maintaining consistent quality across diverse topics and instructors. As organizations increasingly face pressure to upskill and reskill their workforce in response to technological and market changes, tools that dramatically reduce the production barrier for quality instructional content may become essential components of agile and responsive training ecosystems.

References

Clark, R. C., & Mayer, R. E. (2016). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Wiley.

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

Holstein, K., McLaren, B. M., & Aleven, V. (2020). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (pp. 157–171). Springer.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Martin, F., Ritzhaupt, A., Kumar, S., & Budhrani, K. (2019). Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation. *The Internet and Higher Education*, 42, 34-43.

Mayer, R. E., & Fiorella, L. (2021). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 279–315). Cambridge University Press.

Poquet, O., Lim, L., Mirriahi, N., & Dawson, S. (2022). Video and learning: A systematic review (2007–2017). *Review of Educational Research, 92(2)*, 171-217.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.

Author's Background

Kelly Puzio, Ph.D., is the Director of Research at Creatium. He is a learning scientist who received a Ph.D. in Learning, Teaching, and Diversity from Vanderbilt University (2012), a Master's in Education from DePaul University (2002), and a bachelor's degree from the University of Notre Dame. He was a certified Language Arts teacher in ethnoracially and linguistically diverse middle- and secondary schools in New Zealand and the United States (2001-2007). He was an Assistant Professor and then Associate Professor with tenure at Washington State University in the Department of Teaching and Learning. His current research interests include personalized learning and artificial intelligence. Dr. Puzio has served as Principal Investigator on over $5 million in grants from the National Institutes of Health and the National Science Foundation. He was a recipient of the Young Scholar Award (2014) from the International Literacy Association.