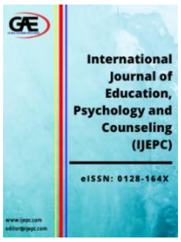
International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 63 (June 2026) PP. 588-600



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
https://gaexcellence.com/ijepc

GLOBAL ACADEMIC EXCELLENCE

AI-ENHANCED EDUCATIONAL SHORT VIDEOS: KNOWLEDGE DISSEMINATION AND LEARNING EFFICIENCY

Liang Jingjing[1], Suraya Md Nasir[2]*

[1]Faculty Of Art, Sustainability, and Creative Industry, Sultan Idris Education University, Malaysia
P20241000508@siswa.upsi.edu.my https://orcid.org/0009-0008-0109-6020

[2]Faculty Of Art, Sustainability, and Creative Industry Sultan Idris Education University, Malaysia
suraya.mn@fskik.upsi.edu.my https://orcid.org/0000-0002-1700-5049

*Corresponding Author

Article Info:

Article history:
Received date: 02.04.2026
Revised date: 19.04.2026
Accepted date: 25.05.2026
Published date: 18.06.2026

To cite this document:

Liang, J., & Nasir, S. M. (2026). AI-Enhanced Educational Short Videos: Knowledge Dissemination And Learning Efficiency. *International Journal of Education, Psychology and Counselling, 11*(63), 588-600.

DOI: 10.35631/IJEPC.1163034

Abstract:

Current educational short videos generally suffer from problems such as loosely structured knowledge presentation, excessive learner workload, and low knowledge retention. To address these issues, this study proposes an Artificial Intelligence (AI) enhanced educational short-video framework based on multimodal analysis and dynamic AI adaptation. This framework utilizes pre-trained VideoMAE, Whisper, and fine-tuned Contrastive Language–Image Pre-training (CLIP) models to parse and align video content across visual, audio, and text modalities, generating fine-grained semantic indexes. Furthermore, a subject-specific knowledge graph is constructed based on structured segments, enabling adjacency relationships and cross-module association recommendations. A Long Short-Term Memory (LSTM) network is designed for real-time diagnosis of learners' attention states and knowledge mastery levels. Finally, a rule-driven dynamic content adaptation engine triggers personalized intervention strategies based on diagnostic results and the graph topology. This method significantly improves learning outcomes. In the reported learning experiment, the experimental group demonstrated greater improvement in post-test scores, with a 31.6% reduction in cognitive load, a 40.2% increase in fixation time in important domains, and high knowledge retention. This effectively promotes long-term memory and the comprehensive application of knowledge.

Keyword:
Dynamic Content Adaptation; Learning Efficiency; Educational Short Videos; Knowledge Graph; VideoMAE

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 63 (June 2026) PP. 588-600

Introduction

With the in-depth integration of the "short video + education" model, educational short videos have become an essential carrier of knowledge dissemination and an immersive experience. Their personalized and social uses promote the reconstruction of learning scenarios. However, most existing short-video education often suffers from problems such as linear, rigid content presentation, a lack of real-time interactive feedback, and difficulty accommodating individual differences in cognitive patterns. This issue leads to bottlenecks, such as low knowledge dissemination efficiency and deep learning failures.

This paper designs and empirically tests a comprehensive Artificial Intelligence (AI) enhanced educational short-video framework. Furthermore, this paper constructs a fine-grained multimodal corpus and parsing pipeline for educational videos, achieving cross-modal semantic alignment and structured indexing of video content. A dual recommendation mechanism integrating knowledge graph logical relationships and embedded semantic similarity was proposed, providing a foundation for knowledge association and multi-perspective interpretation. A lightweight, real-time cognitive-state diagnostic model based on behavioral sequences was developed, enabling non-intrusive assessment of learners' attention and mastery. Additionally, a rule-driven dynamic content adaptation engine was constructed that triggers personalized learning intervention strategies based on real-time diagnostic results. Through controlled experiments, this paper systematically verifies the effectiveness of this framework in improving knowledge mastery, reducing cognitive load, optimizing attention allocation, and enhancing long-term knowledge retention.

Literature Review

As a typical product of the integration of "short video + education", educational short videos have demonstrated the core characteristics of an immersive experience, personalized expression, and socialized dissemination. They have become a crucial force in reconstructing knowledge transmission scenarios. Zhang (2025) proposed a path to enhance dissemination power through content optimization, leveraging social media platforms and integrating language characteristics. This approach aimed to address problems of uneven content quality, incorrect video placement strategies, and non-standard language use in the dissemination of educational short videos. Moreover, Song (2022) used the video account of the magazine "Primary School Chinese Teaching" as an example to explore the demand positioning, content planning, and promotion strategies for short videos. Educational short videos are a product of the era of "short video + education". They have the characteristics of immersive experience, personalized expression, and socialized dissemination. Thus, Gou et al. (2022) proposed that the design of educational short videos should be guided by micro-learning theory, multimedia

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 63 (June 2026) PP. 588-600
learning cognition theory, and immersion theory, with text expression and audiovisual language design as the core. Meanwhile, Xie (2023) explored the specific strategies and practical applications of short-video education dissemination in colleges and universities, further underscoring the important role of short videos in disseminating education. In particular, Xu and Fan (2023) extracted 275 videos as samples. They evaluated their content quality from four dimensions: context quality, expression quality, access quality, and dissemination quality to help improve the quality of short videos in the field of international Chinese education.

On the other hand, Khlaif and Salha (2021) combined educational theory frameworks to analyze the characteristics of educational content on the TikTok platform and its fit with micro-learning and nano-learning theories. Fidan and Debbag (2023) asserted that explanatory videos performed best in terms of learning effectiveness and cognitive load management. However, while situational drama videos could enhance interest, they were easy to distract from. YouTube is a commonly used teaching resource in middle school classrooms. Bakla and Mehdiyev (2022) compared the effects of videos from different sources on the learning experience, participation, and cognitive processes. Meanwhile, Polat (2023) analyzed the impact of teachers' appearance in teaching videos across three dimensions: cognition, emotion, and social presence. Pure video teaching often lacks interaction and feedback mechanisms. Additionally, Fidan and Gencel (2022) combined AI feedback with peer evaluation to improve the effectiveness of online learning. Existing research lacks a systematic empirical analysis of the relationship between the knowledge-dissemination path and learning efficacy under AI-enhancement conditions. Therefore, this paper focuses on AI-enhanced educational short videos, aiming to overcome the limitations of existing frameworks. Based on empirical data, it analyzes the impact mechanism on knowledge dissemination efficiency and deep learning efficacy, in line with the research trend of digital education evolving towards intelligence and precision.

Methods

Construction of an Educational Video Corpus

To support model training and evaluation, a dedicated educational video corpus was constructed. The corpus was compiled from mainstream online education platforms and covers two major disciplines: basic information technology and cognitive psychology. After screening, 500 independent videos totaling approximately 125 hours were included. Annotation was completed by three graduate students with relevant academic backgrounds. First, based on curriculum standards and textbooks, the video content was decomposed into the smallest knowledge units, resulting in 5000 standard knowledge points. Specifically, each knowledge point's annotation included: precise start and end timestamps, core keywords, the knowledge module it belonged to, and its logical relationship with preceding and subsequent knowledge points (e.g., "premise," "parallel," "deepening"). To control quality, cross-validation and expert review mechanisms were adopted. Annotation consistency was measured using Cohen's kappa (κ).

$$\kappa=\frac{p_o-p_e}{1-p_e}, (1)$$

where p_o is the proportion of consensus among observers, and p_e is the proportion of consensus among random expectations. The proportion of random expectation consistency is required to

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 63 (June 2026) PP. 588-600

be κ>0.85 for acceptance. The corpus was stored in structured JavaScript Object Notation (JSON) format, retaining the original video, audio separation, and SubRip Subtitle (SRT) subtitle files, forming the foundation dataset for subsequent multimodal model training and evaluation.

Multimodal Content Parsing and Structuring

Content parsing employs a pipeline based on a pre-trained model and achieves fine-grained semantic indexing of video content through multi-stage feature extraction and alignment.

For visual features, VideoMAE (ViT-Base) is utilized to uniformly sample and encode the video at 1 frame per second. Let the total video duration be T seconds, and let the number of sampled frames be $G=\lfloor T \rfloor$. Each frame is encoded into a d_v-dimensional feature vector through a visual transformer, resulting in a visual feature sequence:

$$V=\{v_1,v_2,\ldots,v_N\},v_i\in R^{d_v}, (2)$$

where v_i is the visual feature vector corresponding to the i-th frame.
For audio, the Whisper (medium) model is used to automatically generate high-precision transcripts, and the last hidden state is extracted as the speech representation, yielding the text content and corresponding timestamp. Let the number of speech segments be M, and the dimension of the hidden feature corresponding to each speech segment be d_a, the speech feature sequence is:

$$A=\{a_1,a_2,\ldots,a_M\},a_j\in R^{d_a}. (3)$$

Moreover, for subtitle text, sentence-transformers (all-MiniLM-L6-v2 model) are used directly for encoding to obtain the text feature vector $t_k\in R^{d_t}$ (d_t is the dimension of the text feature vector), which corresponds one-to-one with the speech segments (k and j are aligned).

The key alignment step was achieved through the improved CLIP model. Taking the visual frame and the corresponding Automated Speech Recognition (ASR)-transcribed sentence pair as input, CLIP (ViT-B/32) was fine-tuned on a subset of the constructed corpus to optimize its video-text contrastive learning objective. After fine-tuning, the model can map visual and text features to a shared semantic space, and the aligned visual features $\tilde{v}_i$ and text features $\tilde{t}_k$ satisfy a similarity metric:

$$s(\tilde{v}_i,\tilde{t}_k)=\frac{\tilde{v}_i^{\top}\tilde{t}_k}{\|\tilde{v}_i\|\|\tilde{t}_k\|}, (4)$$

where ⊤ represents vector transpose.

Finally, the video was structured into a sequence of triples $\langle t_s,t_e,f\rangle$, where t_s is the start time of the segment, t_e is the end time of the segment, and the multimodal feature vector f is a weighted concatenation of aligned visual, speech, and text features:

$$f=\alpha\tilde{v}+\beta a+\gamma\tilde{t},\alpha+\beta+\gamma=1, (5)$$

where α,β, and γ are the weight coefficients of visual, audio, and text features, respectively, and ṽ, a, and t̃ are the aligned visual feature vector, the original audio feature vector, and the aligned text feature vector (from the same time segment), respectively.

This process achieves fine-grained semantic indexing at the timestamp level, supporting cross-modal content retrieval and analysis.

Knowledge Graph Construction and Association Recommendation

Based on the structured knowledge fragments, a subject knowledge graph is constructed. Each knowledge fragment serves as a graph node, and its attributes include its multimodal feature vector and text description. The directed edges between nodes are defined by the labeled logical relationships ("premise", "parallel", etc.). Furthermore, semantic associations are calculated using the cosine similarity of node feature vectors, and "related" edges from the k nearest neighbors (k=5) are dynamically added to each node to fill in potential connections not covered by explicit logic. Notably, the graph is stored and retrieved using the Neo4j graph database.

Figure 1 illustrates the accuracy and response time of direct adjacency recommendation and cross-module association recommendation in different subject knowledge modules:

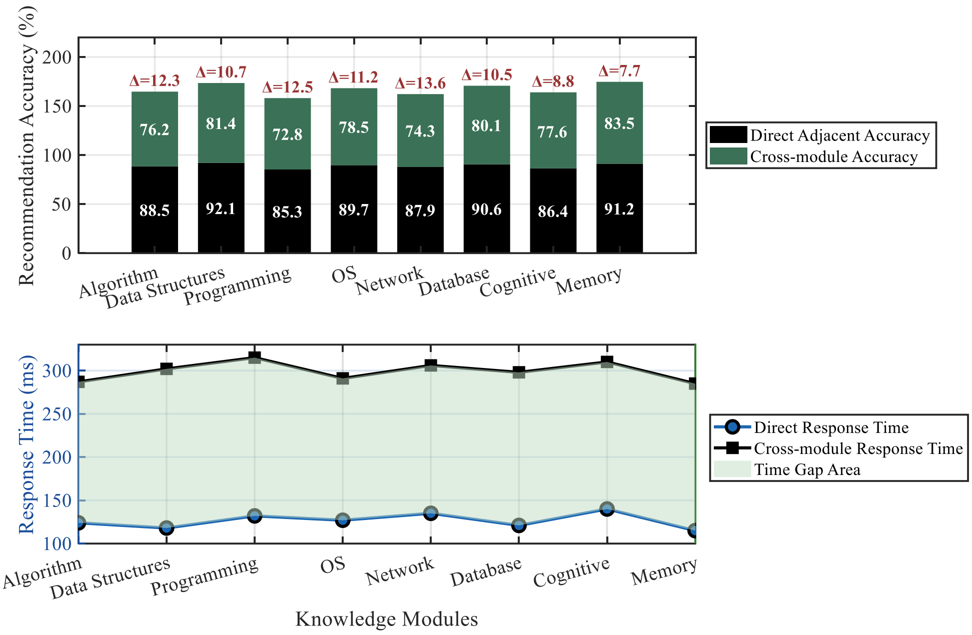


Figure 1. Knowledge Graph Recommendation Performance

Source: By Researcher

Direct adjacency recommendations generally outperform cross-module association recommendations in accuracy, especially in modules involving data structures and memory mechanisms. Although cross-module recommendations have a longer response time, they still maintain high explanatory applicability in complex knowledge associations such as cognitive models.

Additionally, regarding recommendation logic, the system supports two types of queries. One type provides linear or branching recommendations based on the direct adjacency relationship of the node currently viewed by the user. The other is cross-module association

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 63 (June 2026) PP. 588-600

recommendations based on embedding similarity, triggered when the diagnostic module detects that the learner has difficulty understanding. This query retrieves semantically similar but differently expressed perspectives from the graph to provide multi-perspective explanations. Note that all recommendations are accompanied by confidence scores, which serve as one of the decision criteria for the adaptation engine.

Learner Cognitive State Diagnosis

The diagnostic module collects interactive behavior streams non-intrusively through front-end event tracking. Specific captured events include video playback (play, pause, stop), seek operations (including rewind and fast forward), and the results of responses to system-inserted interactive quizzes (correct or incorrect, response time). These events are encoded into time series in real time. The module employed a lightweight Long Short-Term Memory (LSTM) network (128 hidden units) to model the sequence. The network input consisted of the event type, the time offset of the event relative to the knowledge segment's start, and a combined encoding vector of the test response. The output consisted of two real-time estimates: an attention score (0-1) and a knowledge mastery probability (0-1), both updated internally via a gating mechanism. Notably, the LSTM updated formula for the memory unit c_t at time step t is as follows:

$$c_t=f_t \odot c_{t-1}+i_t \odot \tilde{c}_t \quad (6),$$

where f_t is the forget gate, i_t is the input gate, and $\tilde{c}_t$ is the candidate's memory. This process allows the model to selectively remember or forget long-term information.

The LSTM model was trained in a supervised manner on independent behavioral data collected in the early stages of the experiment, using cross-entropy loss and weakly supervised labels for the learner's subsequent test scores as mastery levels. For the binary classification task of knowledge mastery, its loss function L is defined as:

$$L=-\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i)+(1-y_i)\log(1-p_i)\right], \quad (7)$$

where N is the number of samples, y_i is the binary true label (mastery or non-mastery) of the i-th sample, and p_i is the model's mastery probability.

During the inference phase, the module operates in a sliding window manner (window size = the last 10 events). For each new event, the window slides forward one step, and the model performs forward computation based on the event sequence within the window, providing low-latency real-time state updates.

Dynamic Content Adaptation Engine

The adaptation engine is rule-based, based on considerations, including the knowledge graph topology, the characteristics of the current node, and the real-time estimate of the cognitive state output by the diagnostic module. The policy library in the adaptation engine contains several actions the motor can take. For instance, if the "knowledge mastery probability" suggested by the graph is below the threshold of 0.6 for more than 30 seconds, a "prerequisite

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 63 (June 2026) PP. 588-600

knowledge review” strategy is employed. The adaptation engine queries the graph, finds core prerequisite nodes of the current node, and generates a path for review - a collection of chunks along the path. However, if the “attention concentration score” is below the threshold of 0.4, the “insert interactive quiz” strategy is triggered, selecting a question from the medium difficulty question bank and displaying it. This can be formalized in finding the question that would minimize the evaluation function J:

$$j^{*}=\arg\min_{j\in Q}|d_j-\hat{\theta}_l|, (8)$$

where j^{*} represents the index of the selected optimal question, and Q represents the set of candidate questions that match the current knowledge point. Moreover, d_j represents the preset difficulty value of the j-th question, and $\hat{\theta}_l$ represents the learner's current ability level, estimated in real time by the engine based on data such as the diagnostic module's output.

For clusters with more complex knowledge, a "generate summary" strategy is automatically applied after cluster learning is detected. The text generator merges the text descriptions of multiple (or many) nodes in the current cluster to generate a summary, which is then displayed as an overlay of text and image cards. The execution process of all strategies is recorded for subsequent review and improvement.

Experimental Design

To test the effectiveness of the framework, a randomized controlled trial design was employed. 240 undergraduate students from different majors at the same university participated in the experiment. They were randomly assigned to the experimental and control groups using a random number table. In particular, the experimental materials were selected from a constructed corpus covering two independent topics (Information Technology Fundamentals and Cognitive Psychology). Note that each topic included a video of the same length. The videos in the experimental group were augmented with a complete AI system, while those in the control group were the original videos. Furthermore, the experiment was conducted in a controlled laboratory using standardized equipment. Measurement tools included: 1) pre- and post-test questionnaires to assess knowledge mastery and transferability; 2) the NASA Task Load Index (NASA-TLX) Subjective Cognitive Load Scale, completed immediately after learning; 3) a Tobii eye tracker to record fixation points and heatmaps during the learning process; and 4) a system background log to fully record all interaction events. The experimental procedure was as follows: pre-test -> learning phase (eye-tracking recording) -> post-test and questionnaire completion -> delayed post-test one week later.

Results and Discussion

Knowledge Dissemination Effect

In a controlled environment, participants learned both the experimental version equipped with the AI system and the original control version of video materials on basic information technology and cognitive psychology. A quantitative analysis of intra-subject and inter-group comparisons of post-test scores was conducted, and the results are illustrated in Figures 2 and 3:

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 63 (June 2026) PP. 588-600

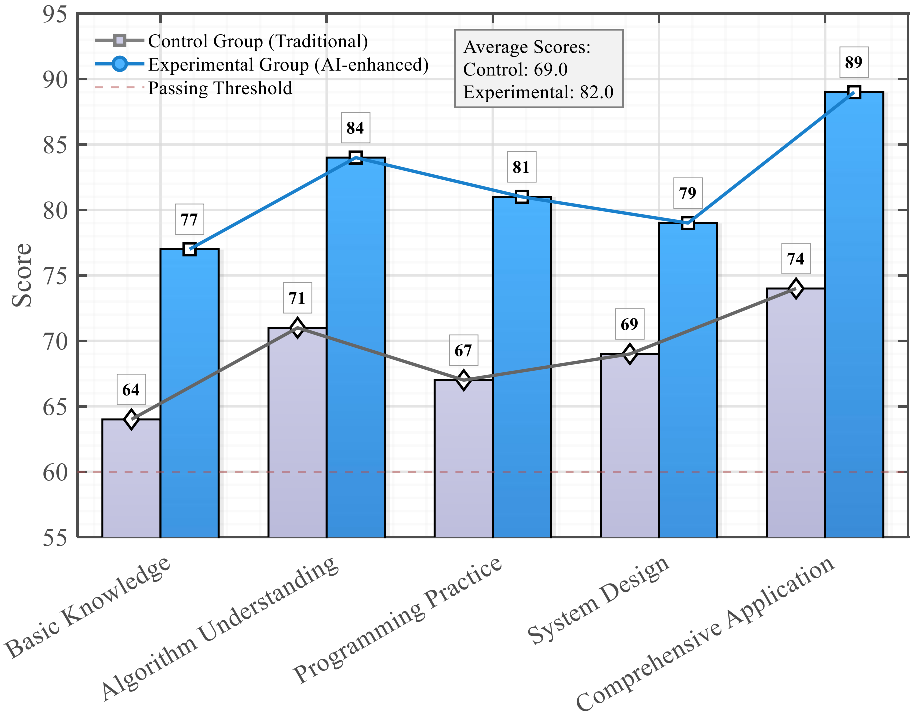


Figure 2. Post-test scores in Information Technology

Source: (By Researcher)

From the results analysis, it can be seen that the post-test scores of the experimental group on the five test items in the Information Technology course were significantly higher than those of the control group, resulting in an overall average improvement of 18.8%. However, the improvement was not equally effective for all kinds of knowledge modules. In the second test item, “Algorithm Understanding,” and in the third, “Programming Practice,” which focus on logical reasoning and skill application, the experimental group scored 84 and 81, respectively. This result represented a significant improvement over the control group, demonstrating that the dynamic path recommendation feature and the AI system's immediate feedback could promote learning for some knowledge types that involve gradual learning and practical applications. Conversely, the item scores for “Basic Knowledge” and “System Design” were 77 and 79, respectively, indicating that the AI system's enhancement effect still differs somewhat across knowledge types, with more facts and design capabilities that require synchronizing comprehensive contexts. The item score for the experimental group is highest in the “Comprehensive Application” and is fairly high at 89. This score also exhibited the largest gap relative to the control group, indicating that this feature effectively promotes learners’ integration of scattered knowledge points and their higher-order transfer application ability through cross-knowledge-point association recommendations and summary generation.

International Journal of
Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 63 (June 2026) PP. 588-600

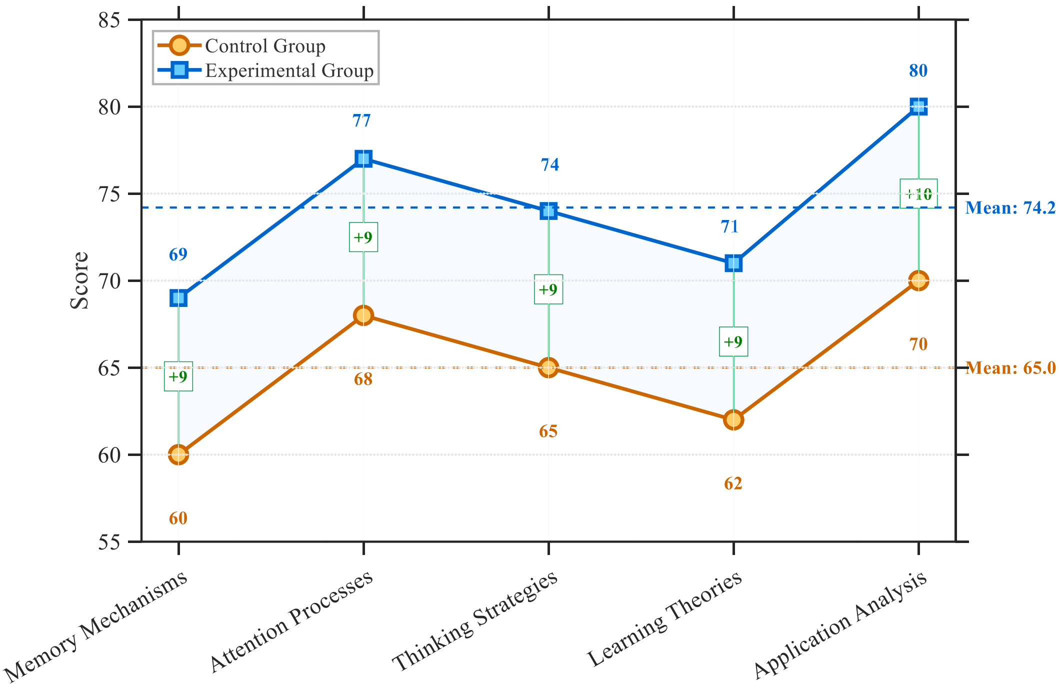


Figure 3. Post-test scores in Cognitive Psychology

Source: By Researcher

The results analysis revealed that the experimental group outperformed the control group in all areas of cognitive psychology, with an average improvement of 14.2%, and its improvement pattern differed from that of information technology. Moreover, trend lines indicated that the experimental group's improvement across all knowledge modules was relatively stable. In "Memory Mechanisms" and "Learning Theory," the experimental group scored 69 and 71 points, respectively. On the other hand, the control group scored 60 and 62 points, a difference of 9 percentage points. These areas involve abstract mechanisms and models, and the AI system improved students' learning outcomes and lowered the learning threshold through multifaceted explanations and key-point reviews. Similarly, the improvement was relatively stable in the areas of "Attention Processes" and "Thinking Strategies", indicating that the system effectively helped students learn strategies for cognitive process regulation.

In the most challenging "Application Analysis" section, the control group achieved the highest score (70 points), while the experimental group still significantly improved by 10 points, reaching 80 points. Therefore, the AI system did not interfere with the performance of high-ability learners and achieved greater progress within the user's framework, thanks to its rigorous analytical reasoning capabilities and the system information ontology built based on knowledge graphs and integrated summaries.

Learning Process and Cognitive Load

Participants completed the NASA-TLX scale immediately after the learning process. Data obtained from eye-tracking was filtered using an Velocity-Threshold Identification (IV-T) Fixation filter and sampled at 120Hz. Consequently, the video frame was sliced into critical knowledge areas (KAR) and non-critical areas (NAR). Total time of fixation (sec), retracements, and average pupil diameter (mm) were recorded. Data processing was performed through Python. The inter-group comparisons are depicted in Table 1:

Table 1. Key Indicators of Learning Process and Cognitive Load

Metric	Experimental Group (Mean)	Control Group (Mean)	Change Rate
NASA-TLX Total Cognitive Load Score	42.3	61.8	Decrease 31.6%
Total Fixation Duration on KAR (sec)	358.7	255.9	Increase 40.2%
Number of Saccades Back (count)	12.4	18.6	Decrease 33.3%
Average Pupil Diameter (mm)	3.81	4.12	Decrease 7.5%

Source: (By Researcher)

The experimental group had a 31.6% lower total cognitive load than the control group, along with a 7.5% lower mean pupil diameter, indicating that the dynamically responsive AI system reduces learners' psychological effort and stress during learning. Likewise, the experimental group recorded a 40.2% increase in fixation time for key knowledge areas and a 33.3% reduction in rescans. These findings reflected that learners' visual attention was better stewarded by the system and focused on core content, minimizing ineffective backtracking due to comprehension issues or inattention.

Knowledge Retention Rate

To evaluate knowledge retention, all subjects underwent delayed post-tests at one, two, three, and four weeks, and the percentage of retention at each time point relative to the immediate post-test score was calculated. The results after data processing are depicted in Table 2:

Table 2. Knowledge Retention Rate at Different Time Points

Time Point	Experimental Group (IT)	Control Group (IT)	Experimental Group (Psychology)	Control Group (Psychology)
1 Week Later	92.4	85.1	90.7	82.3
2 Weeks Later	88.6	76.5	86.2	74.8
3 Weeks Later	83.2	68.9	81.5	67.4
4 Weeks Later	79.8	62.4	77.1	60.9

Source: By Researcher

As illustrated in Table 2, the experimental group had noticeably higher retention rates than the control group at all time points. The difference in retention rate between the two groups

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 63 (June 2026) PP. 588-600
widened over time (by one week, the experimental group was 90% or higher, while the control group was 85.1% or lower. After four weeks, the experimental group was close to 80%, and the control group was 62.4% or lower. Consequently, this result indicated that the AI-enhanced system's use of dynamic content adaptation and knowledge graph-based recommendations improved learning outcomes, as expected. It also effectively promoted the consolidation and transfer of knowledge from working memory to long-term memory, thereby reducing forgetting.

Conclusions

Based on multimodal parsing and the construction of a structured knowledge graph, this system achieves intelligent knowledge association and dynamic path planning. Based on a dynamic adaptive engine employing real-time cognitive diagnosis, it significantly improves learners' knowledge mastery, optimizes attention allocation, reduces cognitive load, and promotes long-term knowledge retention. Notably, the limitations of this study include the relatively controlled experimental environment, the relatively homogeneous participant group, and the need for further observation of its long-term application effects. Therefore, future studies could explore more complex affective computing and adaptive recommendation algorithms. It could also validate the system's generality and scalability across a broader range of educational scenarios and populations, thereby promoting the in-depth development of intelligent education services.

Acknowledgements:	The authors would like to express their sincere gratitude to Shandong Huayu University of Technology for providing the necessary resources and support throughout the course of this research. Special appreciation is extended to colleagues and peers who contributed valuable insights and constructive feedback, which greatly enhanced the quality of this paper.
Funding Statement:	No Funding
Conflict of Interest Statement:	The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Education, Psychology and Counseling (IJEPC).
Ethics Statement:	This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards.
Author Contribution Statement:	All authors contributed significantly to the development of this manuscript. Suraya Md Nasir was responsible for the conceptualization, methodology, and overall supervision of the study. Liang Jingjing handled data collection, analysis, interpretation of results， the literature review, drafting, and critical revision of the manuscript. All authors read and approved the final version of the manuscript prior to submission.

References

Bakla, A., & Mehdiyev, E. (2022). A qualitative study of teacher-created interactive videos versus YouTube videos in flipped learning. *E-Learning and Digital Media, 19*(5), 495–514.

Fidan, M., & Debbag, M. (2023). Comparing the effectiveness of instructional video types: An in-depth analysis on pre-service teachers for online learning. *International Journal of Human–Computer Interaction, 39*(3), 575–586.

Fidan, M., & Gencel, N. (2022). Supporting instructional videos with chatbot and peer feedback mechanisms in online learning: Effects on learning performance and intrinsic motivation. *Journal of Educational Computing Research, 60*(7), 1716–1741.

Fyfield, M., Henderson, M., & Phillips, M. (2022). Improving instructional video design: A systematic review. *Australasian Journal of Educational Technology, 38*(3), 155–183.

Gou, R., Chen, X., & Yang, X. (2022). Educational short videos in the era of intelligent media: Connotation, types, and application approaches. *Journal of Distance Education, 40*(6), 103–110.

Khlaif, Z. N., & Salha, S. (2021). Using TikTok in education: A form of micro-learning or nano-learning? *Interdisciplinary Journal of Virtual Learning in Medical Sciences, 12*(3), 213–218.

Krumm, I. R., Miles, M. C., Clay, A., et al. (2022). Making effective educational videos for clinical teaching. *Chest, 161*(3), 764–772.

Polat, H. (2023). Instructors' presence in instructional videos: A systematic review. *Education and Information Technologies, 28*(7), 8537–8569.

Rahman, M. R., Koka, R. S., Shah, S. K., et al. (2024). Enhancing lecture video navigation with AI-generated summaries. *Education and Information Technologies, 29*(6), 7361–7384.

Simon, F., Peer, S., Michel, J., et al. (2021). IVORY guidelines (Instructional videos in otorhinolaryngology by YO-IFOS): A consensus on surgical videos in ear, nose, and throat. *The Laryngoscope, 131*(3), E732–E737.

Song, Y. (2022). Exploration of demand positioning, content planning, and promotion strategies for educational short videos: A case study of the video account of *Primary School Chinese Language Teaching. Journal of Journalism Research, 13*(1), 112–114.

Vogt, F., Koechlin, A., Truniger, A., et al. (2021). Teaching assistants and teachers providing instructional support for pupils with SEN: Results from a video study in Swiss classrooms. *European Journal of Special Needs Education, 36*(2), 215–230.

Xie, Y. (2023). Research on dissemination strategies and practices of short video education in universities based on the Douyin platform. *Journal of Journalism Research, 14*(7), 180–183.

Xu, D., & Fan, J. (2023). An exploration of quality evaluation and optimization strategies for short video content in international Chinese education: A case study of *Chinese Bridge. Journal of Liaoning Normal University (Social Sciences Edition), 46*(5), 25–33.

Zhang, C. (2025). Research on enhancing the dissemination power of educational short videos. *Journal of Hubei Normal University (Philosophy and Social Sciences Edition), 45*(2), 151–156.