Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD38734

Application Patterns and Educational Effects of Intelligent Tutoring Systems for Memory Learning in K-12 Education

Weijie Zhang

School of Computer Science and Technology, Anhui University of Technology, Maanshan, China
zhangwj@ahut.edu.cn

Abstract. It is crucial for K-12 students to find a way to remember the information they have learned, particularly when working with an intelligent tutoring system. In order to learn more about how intelligent tutoring systems can assist them in this process, the literature review was conducted in relation to documents that were published between the dates of January 1st 2018 and August 31st 2026. The methods that were studied include spaced practice, AI-based question creation, visual flashcards, and gamification. Various outcomes were obtained from the studies; the textbooks that were studied on the subject of spaced practice and visual flashcards offered information on retaining vocabulary, while studies examining the art of question formulation provided information about the practice of retrieval and feedback technique. However, investigations into the gamification of intelligent tutoring systems mainly examined how effective the system is in terms of how engaged learners were with the learning material. Therefore, all of the studies were conducted with different samples and settings. In the general sense, it can be said that using computer-assisted instruction properly is one of the most effective methods of memory retention learning in K-12 education, which provides the strongest effectiveness in the area of vocabulary retention.

Keywords: intelligent tutoring systems, memory learning, spaced repetition, AI question generation, K-12 education

1. Introduction

1.1. Research background

K-12 memory tasks could comprise anything from dates and terminology to specialist knowledge. Intelligent tutoring systems (ITS) utilize learner input to personalize learning experiences through provision of practice, feedback, and assessment. Xu et al. [1] indicated that the impact of ITS on K-12 reading comprehension has been always very positive. Bayesian knowledge tracing (BKT) seeks to predict achievement of learners on the basis of their answers [2, 3], while generative AI offers the chance to enhance tutoring personalization [4].

There are important variations between different memory processes because of the differing amount of prerequisite knowledge needed for the memory process and also due to the lag between the moment of acquiring that knowledge and using it in practice. It is possible for a student to be

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD38734

able to say a term in class but then to be unable to recall it later. In addition to providing practice activities, memory support systems must deliver other methods of helping learners remember including the timing of practice, assistance provided through prompts, and management of errors.

1.2. Research questions and methodology

Three research questions (RQs) guided the study:

RQ1: What are the common applications of intelligent tutoring systems with respect to memory learning in K-12 education?

RQ2: How do these applications compare regarding memory retention and achievement?

RQ3: What are the necessary conditions needed for effective use of these applications?

1.3. Research objectives and significance

The current article explores the ways by which intelligent tutoring systems can meet memory-retention needs through K-12 learning processes. An important aspect of memory-retention has to do with a learner's ability to study any subject and produce knowledge about the subject at some time later.

The article is about applications and does not concern itself with the entire framework of intelligent tutoring systems as made use of by anyone. Rather, the article discusses the characteristics of these various applications in terms of how they are used in learning, the evidence available for their being effective, and the nature of conditions that render this effect meaningful. Understanding this difference is key to knowing that while there may be an increase in the participation and achievement of results, there might be no permanent memory retention.

1.4. Paper structure

Section 2 deals with theoretical assumptions. In Section 3, the search and encodings are defined. Section 4 deals with the evidence. In Section 5, the conditions and the implications of the design are explained. Finally, Section 6 summarizes the paper and its limitations.

2. Literature review

2.1. Theoretical foundations of memory learning

Memory learning refers to how much information students will be able to retrieve after encountering it. Certain terms, dates, formulas, and concepts make the timing of reviews and the manner of retrieval important design considerations for learners.

Retention is dependent on time in that immediate and delayed measuring will yield different results. Recognition does provide clues while recall requires learners to recollect the information; these two are therefore not synonymous.

Spacing distributes practice over time while retrieval practice makes reconstructing the information necessary. However, both of them do support learning for a long time [5, 6].

2.2. Intelligent tutoring systems: an overview

An intelligent tutoring system collects data about students to modify their exercise sequence, feedback, and difficult level [1].

DOI: 10.54254/2755-2721/2026.LD38734

An intelligent tutoring system usually has models related to the domain, students, teaching methodology, and interface. The BKT model revises its estimate of mastery level every time [2, 3] and it needs to be determined whether the learning effect comes from the process itself or from the material and feedback given.

The four patterns of learning are spaced repetition, AI-generated questions, visual flashcards, and gamification. Due to different functions performed by these methods, one general measure will not be applicable in these cases.

2.3. Four core application patterns

Retention can be measured in any of the three ways mentioned previously and through achievement as an indicator of instructional effectiveness. Engagement can be determined through coursework and use of the materials available through the system, while motivation can be assessed in several ways, including through the Motivated Strategies for Learning Questionnaire (MSLQ), which assesses a variety of motivational items.

SyncStudy is a prototype platform developed by the authors. It is presented only as an illustrative design case and is not part of the reviewed literature.

2.3.1. Spaced repetition pattern

The different types of spaced-repetition systems include, but are not limited to, the Leitner system, the SM-2 algorithm, the half-life regression, and the Adaptive Control of Thought-Rational (ACT-R) models [7]. More recent studies have looked into the different spaced-repetition techniques geared toward second-language learning, Kanji learning, and mobile-assisted vocabulary learning [8-10]. The success of the different spaced repetition systems depends on such factors as vocabulary knowledge transfer, cognitive load, adherence to a schedule, and learner motivation. In the case of SyncStudy, Bayesian knowledge tracing is used to ascertain weak knowledge states and determine an optimal time to review any content.

2.3.2. AI question generation pattern

Creating questions through AI offers possibilities for retrieval, but the quality of the generated questions varies with respect to appropriateness and applicable instructional utility. Each question type has to be evaluated according to its intended purpose [11, 12]. The questions created by AI in SyncStudy serve purposes when they reflect the knowledge required and provide an explanation that helps a user who did not previously understand.

The automation of this process opens up new practice opportunities, but it does not eliminate the need to verify content accuracy and difficulty.

2.3.3. Visual flashcard pattern

Visual flashcards help serve as cues and connect ideas but do not help ensure learning. Previous studies have shown that learner-made flashcards and digital flashcards can help the learners in word learning [13-15]. In SyncStudy, retrieval and recognition tasks can be attached to visual representations of ideas. Flashcards work well for specific types of cues but mind maps and diagrams are better suited for tasks that require making sense of relationships between ideas.

DOI: 10.54254/2755-2721/2026.LD38734

2.3.4. Gamification pattern

Gamification involves using points, badges, and levels to encourage learners to participate more and stay engaged; however, the outcomes of gamification can vary based on its design and the targeted students [16, 17]. Furthermore, gamification is only recommended as supplementary practice; it should not replace effective teaching strategies.

Rewarding is justified in gamification, but only when the activity has led to productive learning and correction of mistakes.

2.4. Summary

The included evidence indicates four distinct memory-learning patterns with different pedagogical functions. No single tool is universally appropriate because vocabulary retention, reading comprehension, participation, and motivation are distinct outcomes.

3. Methodology

3.1. Literature search strategy

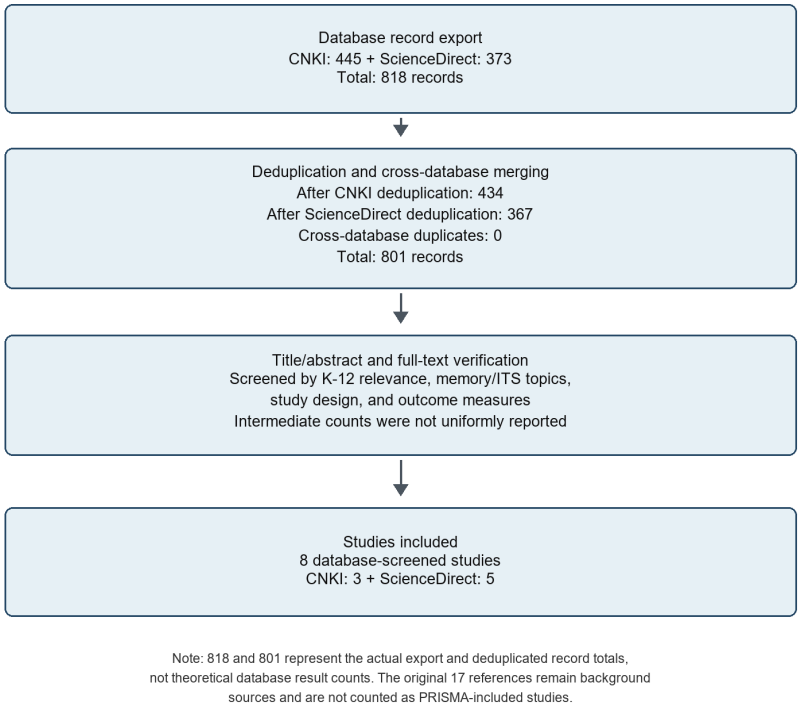


Figure 1. Literature screening flowchart

Note: 818 records were exported and 801 remained after deduplication. Intermediate counts were drawn from database export and screening records.

The research strategy employed CNKI and ScienceDirect from January 2018 to August 2026. The use of Google Scholar was avoided because it is not able to provide advanced filtering criteria

DOI: 10.54254/2755-2721/2026.LD38734

appropriate to complete systematic reviews; the use of Web of Science was avoided due to relevant search findings. Records were screened for descriptive data of the population, treatment, and outcome. After deduplication and screening, there were eight records kept – three from CNKI and five from ScienceDirect. Figure 1 shows the details of the screening and search process.

The included records were those involving K-12 learning, memory tasks, intelligent tutoring, etc., and provided results on retention, achievement, etc. Excluded were theoretical studies, not relevant studies, etc. The year of publication was checked during the screening for inclusion as was a review of background and review materials but retained not for data collection but for context.

The final database included eight studies that had undergone screening. Background and review articles were included solely for contextual purposes due to the different databases utilized.

3.2. Coding and classification

Records were examined in terms of application categories, participant characteristics, educational level, subject matter, technology usage, outcomes, and direction of effect. The application categories are spaced repetition, AI generation of questions, visual flashcards, gamification, and adaptive assessment and tutoring. The coding for retention time, performance, engagement, and motivation was done separately as these reflect a variety of educational outcomes.

The studies examined language, Arabic writing skills, mathematics, reading, and person intelligence information with different methods for reporting. The decisions regarding screening were compared with database export information and record with data that were retained. The coding differed between the study of delayed retention and that of immediate performance or motivation and academic achievement. The method chosen was structured narrative synthesis; however the differences in duration of the intervention and the characteristics of each study's subjects made it impossible to pool effect sizes into one value.

4. Results

Table 1 summarizes the eight included studies by participant population, intervention type, outcome, and reported effect.

Due to a variety of the educational evidence designs and measurements, no actual synthesis in a quantitative way is possible. Therefore, the known differences among interventions are based on the goal of the intervention and the surrounding conditions that provide significance of reported outcomes. When talking about spaced repetition, it deals with the timing of recall; AI question generation is concerned with the quality of retrieval and feedback; visual tools are applied for encoding and organizing information; while gamification is associated with the influence on persistence and engagement. Even though intelligent tutoring studies can be the evidence for adaptive assessment, they cannot offer supportive conclusions for the long-term retention of information. LLMs are considered to be useful for note-taking and increasing the level of learner engagement process. Notably, the effect of gamification can be estimated based on recognition, motivation, anxiety, performance, and delayed recall outcomes. Such differences imply the importance of measuring outcomes for each pattern of the gamification process. In one case, the use of ARCS principles for vocabulary instruction was found to be very efficient.

DOI: 10.54254/2755-2721/2026.LD38734

Table 1. Included studies and reported outcomes (ARCS = Attention, Relevance, Confidence, and Satisfaction; LLM = large language model)

Pattern	Study	N	Level	Task	Outcome	Effect
Gamification	ARCS vocabulary teaching	80	High school	English vocabulary	Retention; motivation	Positive
ITS	Assessment system	NR	Middle school	Mathematics	Outcomes; workload	Preliminary positive
ITS	Big-data ITS design	NR	Middle school	Mathematics	Assessment	Limited evidence
AI/LLM	LLM note-taking	405	Middle school	Reading/memory	3-day retention; comprehension	Combined better
Gamification	Complexity game	NR	High school	English vocabulary	Performance; anxiety	Better; more anxiety
Visual tool	Arabic letters	20	Primary	Arabic letters	Recognition; recall	Higher visual scores
Flashcards	Grammar cards	31	School age	English grammar	Delayed retention	Higher gain; less loss
Flashcards	Video/flashcards	4	School age	Global words	Recognition	Positive for some

5. Discussion

• Applicable conditions and moderating factors.

Table 2 depicts the variety of processes available, related duties, outcome metrics, and constraints of the four models. It elaborates on the mechanisms with no claim made of superiority. Previous studies suggest that the four models work in disparate ways in various contexts.

The different aspects of the memory process play the part of pacing in the case of spaced repetition, support retrieval with AI-generated questions, point out key concepts in association with visual information, and motivate engaged learning via gamification.

In terms of task comparison and outcome indicators, a straightforward measure of retention would be the delayed recall. As far as AI-generated questions are concerned, the accuracy of objects and comprehension of the text would be adequate metrics. The use of visual components necessitates vocabulary retention, while the use of gaming requires motivation to engage users.

The third research question addresses the following limitations. For the use of AI-generated questions, information on treatment provision and verification, as well as difficulty level and error types is required. In terms of visual components, the information must be clear. As for gaming technologies, the emphasis needs to be laid on achieving educational aims rather than gaming for entertainment purposes.

• Design recommendations for K-12 memory tutoring systems.

The SyncStudy prototype can incorporate all four designs as an illustration. The prototype is able to detect potential knowledge gaps by employing BKT and help learn in time, guide towards prompt questions, visualize dependent concepts, and gamify activities to stimulate participation. These design aspects only serve as examples of what can be done, not a demonstration of the efficacy of the prototype.

Assessment should be used to tell immediate retention from storage retention instead of relying on visit frequency or assignment scores. For example, assessment could include recall, accuracy,

DOI: 10.54254/2755-2721/2026.LD38734

timing of task completion, or engagement or motivation as demonstrated through age-appropriate behavior. Overall, in the case of the four designs, the success or quality of the implementation of tasks is more significant than merely the presence of another particular resource.

Table 2. Comparison of four application patterns

Pattern	Main mechanism	Applicable learning tasks	Main outcome indicators	Main limitations
Spaced repetition	Schedules review by mastery	Long-term retention	Delayed retention tests	Workload or forgetting
AI question generation	Retrieval with feedback	Retrieval and correction	Item quality and reading	Errors, difficulty, duplicates
Visual flashcards	Encodes relationships	Terms, concepts, relations	Vocabulary retention	Visual burden or form focus
Gamification	Feedback and incentives	Low engagement; sustained practice	Engagement and motivation	Shallow strategies; competition

6. Conclusion

With respect to RQ1, the analysis indicates four ways in which K-12 memory learning occurs: all types of spaced repetition, AI-generated questions, visual flashcards, and gamification. Rather than being interchangeable, these types of memory learning describe different phases of cognition and memory. The components of spaced repetition provide temporality for various kinds of information retrieval; the use of AI-generated questions provides opportunities for retrieval effort and feedback; visual flashcards help facilitate coding and concept organization; and gamification encourages participants to continue participating. Therefore, the evidence presented in the literature supports a pattern-based approach to IT-based learning, whereby features are deemed useful depending on the memory task and contextual aspects of instruction.

As for RQ2, there was no single program that consistently outperformed the other programs. Spaced repetition proved to have the most compelling evidence when it comes to vocabulary retention as long as the reviews fell within the time frame. The use of AI for providing questions was most valuable in the area of retrieval practice and recovering from errors made during learning but its value in the educational process depends on the correctness, level of difficulty, and how well feedback is provided. Using visual flashcards helped people to learn vocabulary and the relationships between the subject matter and the pictures. Gamification seemed to stimulate engagement and motivation; however, it is vital not to interpret growth in participation as an indicator of increased retention in the process. Hence, retention, motivation, participation, and comprehension need to be identified and tested separately.

As to RQ3, successful implementation requires the following considerations: matching patterns to task types, ensuring compliance with reviewing schedules, and ensuring reward systems align with the objectives of learning. At the system's level, a distinction should be made between immediate performance vs. retention with the outcomes measurement appropriate for the intervention. Thus, for instance, delayed recall will be used in scheduled review, whereas the accuracy of questions and quality of explanations will be employed in AI question-driven reviews. The visual tool of this program should clarify the relations contained within, and the gaming features of the program should provide an environment where learners are engaging in meaningful practice rather than rapid guessing.

DOI: 10.54254/2755-2721/2026.LD38734

The limitations of this study stem from the relatively small evidence base: it contained only eight studies from CNKI and ScienceDirect after database screening, with heterogeneous characteristics including study designs, participants, and measures. This means that the findings promote cautious design implications instead of broad causal conclusions. Future research needs to implement the comparison of personalized review schedules, retrieval-practice formats, and the combination of these patterns with consistent delayed-retention measures in K-12 settings. It should be noted that the approach that is most effective for long-term K-12 memory learning is the combination of scheduled review and retrieval practice with visual and motivational input where needed.

References

[1] Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology, 50*(6), 3119–3137. https://doi.org/10.1111/bjet.12758

[2] Šarić-Grgić, I., Grubišić, A., & Gašpar, A. (2024). Twenty-five years of Bayesian knowledge tracing: A systematic review. *User Modeling and User-Adapted Interaction, 34*(4), 1127–1173. https://doi.org/10.1007/s11257-023-09389-4

[3] Cui, S., Xu, S., Wang, M., & Zhou, Y. (2026). Research progress on deep learning knowledge tracing for intelligent education. *Computer Engineering, 52*(4), 39–61. https://doi.org/10.19678/j.issn.1000-3428.0069653 (in Chinese)

[4] Xu, S., Tong, J., & Hu, X. (2024). Next-generation personalized learning: Generative artificial intelligence-enhanced intelligent tutoring systems. *Open Education Research, 30*(2), 13–22. (in Chinese). https://doi.org/10.13966/j.cnki.kfjyyj.2024.02.002

[5] Davis, D., Kizilcec, R., Hauff, C., & Houben, G. J. (2018). Scaling effective learning strategies: Retrieval practice and long-term knowledge retention in MOOCs. *Journal of Learning Analytics, 5*(3), 21–41. https://doi.org/10.18608/jla.2018.53.3

[6] Corral, D., & Carpenter, S. K. (2025). Effects of retrieval practice on retention and application of complex educational concepts. *Learning and Instruction, 100*, 102219. https://doi.org/10.1016/j.learninstruc.2025.102219

[7] Wu, Y., Luo, F., Tao, J., Wang, C., & Zhang, Z. (2025). An ACT-R-based cognitive spaced-repetition learning method. *Journal of East China University of Science and Technology (Natural Science Edition), 51*(3), 371–379. https://doi.org/10.14135/j.cnki.1006-3080.20240617001 (in Chinese)

[8] Seibert Hanson, A. E., & Brown, C. M. (2020). Enhancing L2 learning through a mobile assisted spaced-repetition tool: An effective but bitter pill? *Computer Assisted Language Learning, 33*(1-2), 133–155. https://doi.org/10.1080/09588221.2018.1552975

[9] Zhang, R., Zou, D., & Xie, H. (2022). Spaced repetition for authentic mobile-assisted word learning: Nature, learner perceptions, and factors leading to positive perceptions. *Computer Assisted Language Learning, 35*(9), 2593–2626. https://doi.org/10.1080/09588221.2021.1888752

[10] Zhong, Z., & Zhang, Q. (2026). Optimizing Japanese Kanji learning through Anki-assisted spaced repetition: The mediating role of cognitive engagement and visual-verbal integration efficiency. *Innovation in Language Learning and Teaching, 1*–29. https://doi.org/10.1080/17501229.2026.2682439

[11] Das, B., Majumder, M., Phadikar, S., & Sekh, A. A. (2021). Automatic question generation and answer assessment: A survey. *Research and Practice in Technology Enhanced Learning, 16*(1), Article 5. https://doi.org/10.1186/s41039-021-00151-1

[12] Steuer, T., Filighera, A., Tregel, T., & Miede, A. (2022). Educational automatic question generation improves reading comprehension in non-native speakers: A learner-centric case study. *Frontiers in Artificial Intelligence, 5*, Article 900304. https://doi.org/10.3389/frai.2022.900304

[13] Yasar, S., & Kocoglu, Z. (2025). Enhancing vocabulary mastery: The impact of learner-created digital flashcards on L2 vocabulary learning and self-regulation. *Australasian Journal of Educational Technology, 41*(4), 17–31. https://doi.org/10.14742/ajet.10402

[14] Najafi Karimi, S., & Kheradmandi Amiri, R. (2025). The effects of using digital flashcards versus paper flashcards on vocabulary learning and retention of Iranian intermediate EFL learners. *Cogent Education, 12*(1), Article 2548353. https://doi.org/10.1080/2331186X.2025.2548353

[15] Teymouri, R. (2024). Recent developments in mobile-assisted vocabulary learning: A mini review of published studies focusing on digital flashcards. *Frontiers in Education, 9*, Article 1496578. https://doi.org/10.3389/feduc.2024.1496578

DOI: 10.54254/2755-2721/2026.LD38734

[16] Kalogiannakis, M., Papadakis, S., & Zourmpakis, A. I. (2021). Gamification in science education: A systematic review of the literature. *Education Sciences, 11*(1), 22. https://doi.org/10.3390/educsci11010022

[17] Lampropoulos, G., & Kinshuk. (2024). Virtual reality and gamification in education: A systematic review. *Educational Technology Research and Development, 72*(3), 1691–1785. https://doi.org/10.1007/s11423-024-10351-3