International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): 3071-0154
Volume-2 Issue-3 || Jul-Sep 2026 || PP. 8 – 19 www.ijarcse.org

Impact of AI-Based Tutoring Systems on Students' Problem-Solving Skills

Prema Bharti
Research Scholar
University of Rajasthan Jaipur, Jawahar Lal Nehru Marg , Jajpur, Rajasthan
Pin Code - 302004
ORCID id - https://orcid.org/0008-0004-5058-1891

www.ijarcse.org || Vol. 2 No. 3 (2026): September Issue

Date of Submission: 07-08-2026 **Date of Acceptance:** 25-08-2026 **Date of Publication:** 11-09-2026

Abstract — Artificial intelligence-based tutoring systems are increasingly being integrated into education to provide personalised explanations, adaptive hints, and immediate formative feedback. However, improvements in test scores during AI-supported practice do not necessarily demonstrate that learners have developed independent and transferable problem-solving competence. This study addresses this distinction by examining whether adaptive scaffold fading in an AI tutoring environment strengthens students' ability to solve unfamiliar problems after AI assistance is withdrawn. The research framework separates immediate task performance from near-transfer, far-transfer, reasoning efficiency, and independent strategy selection. A controlled educational experiment is designed in which students receive either adaptive AI tutoring, non-adaptive digital practice, or conventional instructor-supported problem-solving activities. The proposed AI tutor combines learner-state estimation, error-pattern classification, progressive hint generation, and confidence-sensitive scaffold reduction rather than directly supplying complete solutions. The study further investigates whether excessive dependence on AI assistance weakens autonomous reasoning despite improving performance during guided practice.

Keywords — Artificial Intelligence Tutoring, Intelligent Tutoring Systems, Problem-Solving Skills, Adaptive Scaffolding, Learning Analytics, Transfer of Learning

Introduction

Artificial intelligence has moved from a specialised research area in educational technology to an increasingly visible component of everyday learning environments. Contemporary AI-enabled systems can analyse learner responses, identify recurring errors, adjust task difficulty, provide personalised explanations, and generate context-sensitive hints within seconds. These capabilities have renewed interest in intelligent tutoring systems because they offer a practical mechanism for providing individualised instructional support at a scale that conventional one-to-one tutoring cannot easily achieve. The expansion of generative artificial intelligence since late 2022 has further extended these possibilities by allowing tutoring interfaces to conduct conversational interaction and generate explanations dynamically rather than relying entirely on predetermined

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

feedback templates. UNESCO consequently identifies generative AI as having substantial educational potential while simultaneously emphasising that its use should remain human-centred, pedagogically validated, age appropriate, and ethically governed.

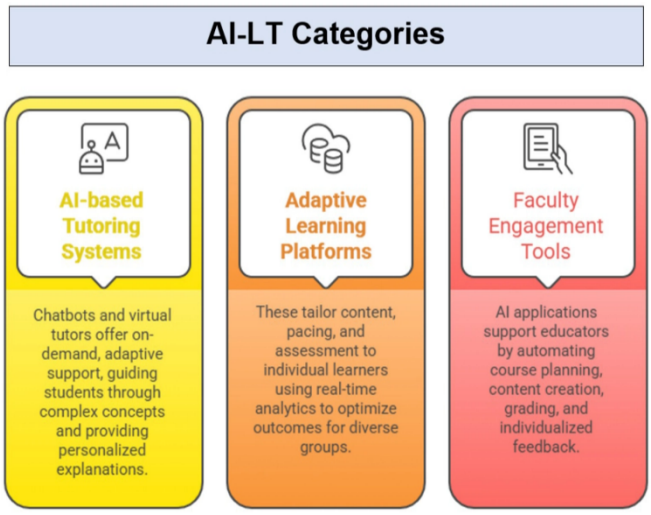


Fig. 1: AI-LT Categories

Source:
https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1689205/full

An important educational application of this technology concerns the development of problem-solving skills. Problem solving differs from simple information recall because it requires learners to understand a problem situation, identify relevant information, construct a strategy, execute appropriate procedures, monitor progress, detect errors, and revise the solution when necessary. Effective problem solving therefore combines conceptual knowledge with strategic reasoning and metacognitive regulation. These processes are difficult to support uniformly in classrooms because students frequently require different forms and amounts of assistance at different stages of a problem.

AI-based tutoring systems appear particularly suitable for this challenge. Unlike static e-learning applications, an intelligent tutor can theoretically respond to the learner's current state. A student making a conceptual error may receive an explanatory prompt, whereas a learner making a procedural mistake may receive a targeted hint. Similarly, a student demonstrating repeated competence can be presented with more complex problems or reduced assistance. Such adaptation represents one of the principal differences between intelligent tutoring and conventional computer-assisted instruction.

Research on intelligent tutoring systems has produced considerable evidence concerning their implementation across educational disciplines. A systematic review by Cuéllar-Rojas, Hincapié-Montoya and Contero examined 163 full-text studies and demonstrated the substantial breadth of intelligent tutoring research while also finding that evaluation within ITS research has frequently remained heavily quantitative and summative. Another systematic review focusing specifically on experiments conducted in authentic educational environments found mixed patterns of effectiveness and highlighted shortcomings in experimental design as well as comparatively limited attention to process-oriented and non-cognitive learning outcomes. These observations are particularly relevant when AI tutoring is evaluated for problem solving.

The central difficulty is that **successful problem completion while receiving intelligent assistance is not equivalent to independent problem-solving ability**. A learner may obtain a correct answer because the tutoring system identifies the next operation, restructures the problem, eliminates incorrect alternatives, or provides increasingly explicit hints. Such assistance may improve immediate accuracy while providing limited evidence that the student can reconstruct the reasoning process independently. Consequently, research based only on post-test scores involving problems similar to those used during tutoring may overestimate the deeper educational value of an AI tutor.

This problem becomes more important with generative AI. Earlier intelligent tutoring systems normally operated within restricted domains and offered carefully engineered sequences of hints. Generative models can produce explanations, examples, questions, feedback, and complete solutions dynamically. Wang and colleagues, for example,

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

investigated the use of generative AI for producing general mathematical problem-solving tips and compared AI-generated material with teacher-generated guidance, illustrating how generative models can become components of future intelligent tutoring environments. Yet the same capability introduces a pedagogical paradox. The more easily an AI system completes reasoning on behalf of the learner, the greater the possibility that apparent improvements reflect **assisted performance rather than cognitive development**. Such an approach shifts AI tutoring from an **answer-generation paradigm** toward a **cognitive apprenticeship paradigm**. The desired outcome is not maximum accuracy while the tutor is available. Instead, the system should attempt to make itself progressively less necessary.

This distinction is consistent with wider policy discussions surrounding AI in education. The U.S. Department of Education has argued that educational AI involves increasingly consequential forms of pattern recognition and automated instructional decision-making and therefore requires careful attention to how technology interacts with teachers, learners, assessment, fairness, and human agency. More recent departmental guidance has explicitly identified AI-enhanced high-impact tutoring as a potential educational application while retaining an emphasis on responsible implementation.

1.1 Research Gap

The research gap addressed in the present study concerns the **difference between AI-supported task success and independent transfer of problem-solving competence**.

Existing intelligent tutoring research has extensively investigated achievement, correctness, learning gains, engagement, and system effectiveness. However, real-world evaluations have devoted comparatively less attention to learning processes and other outcomes beyond conventional performance measures. Research specifically examining mathematical word-problem tutoring has also demonstrated the potential of intelligent tutoring while highlighting important contextual factors affecting participation and educational implementation.

A more demanding question therefore remains insufficiently resolved: **Does an AI tutor teach students how to solve problems, or does it primarily help students solve the problems presented while the tutor is available?**

The present research addresses this issue by evaluating AI tutoring through four connected but distinct dimensions: independent post-assistance performance, transfer to structurally unfamiliar problems, efficiency of the reasoning process, and dependency on automated hints. The central innovation is the inclusion of a **scaffold-withdrawal phase**, during which AI assistance is completely removed. Student performance after withdrawal is treated as more important evidence of problem-solving development than accuracy recorded during assisted practice.

The study therefore proposes that an effective AI tutoring system should satisfy two conditions simultaneously: performance should improve while learning occurs, and reliance on the system should decline as competence develops.

1.2 Research Objectives

The primary objective of the study is to determine whether an AI-based tutoring system using adaptive scaffold fading produces stronger independent problem-solving competence than non-adaptive digital instruction and conventional instructional support.

More specifically, the investigation examines whether adaptive tutoring improves students' ability to select appropriate strategies, reduces inefficient solution attempts, supports transfer to unfamiliar problem structures, and decreases dependence on external hints over repeated learning sessions. The study also evaluates whether learner confidence and actual competence develop consistently or whether AI support creates inaccurate perceptions of mastery. The resulting framework moves beyond the conventional question of whether AI tutoring “improves scores.” Instead,

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

it examines whether learners become **better independent problem solvers after the technology is removed**.

Literature Review

2.1 Evolution of Intelligent Tutoring Systems

Intelligent tutoring systems have historically attempted to reproduce selected functions of individual human tutoring through computational models of domain knowledge, learner behaviour, instructional strategy, and interface interaction. Their basic architecture commonly includes a domain component representing the knowledge to be learned, a learner model representing the student's current state, and a pedagogical component responsible for selecting instructional interventions.

Recent advances in machine learning have substantially expanded this architecture. Contemporary systems can infer learner states from sequences of answers, completion times, error histories, hint requests, and interaction traces. Research has also explored more advanced student-modelling methods. For instance, Li and colleagues proposed Sim-GAIL, a generative adversarial imitation-learning approach for modelling student behaviour in intelligent tutoring environments, illustrating the increasing methodological sophistication of computational learner modelling.

The significance of these developments lies in their ability to move tutoring away from fixed branching rules. Instead of treating every incorrect response identically, an AI tutor may estimate whether an error results from misconception, procedural confusion, careless calculation, insufficient prerequisite knowledge, or inappropriate strategy selection. Instruction can then be adapted accordingly.

Nevertheless, increased computational sophistication does not automatically ensure improved pedagogy. A highly accurate learner model may still be used to provide excessive assistance. Consequently, the educational value of an AI tutor depends not only on whether it understands a student's difficulties but also on how it responds to those difficulties.

2.2 AI Tutoring and Academic Performance

The effectiveness of intelligent tutoring has been investigated across mathematics, science, computer programming, language learning, and related disciplines. Reviews indicate substantial scholarly interest but also methodological heterogeneity. Differences in learner age, intervention duration, tutoring architecture, subject domain, comparison condition, and outcome measures make broad conclusions difficult.

The review of real-world ITS experiments by researchers examining studies between 2011 and 2022 found a complex rather than uniformly positive pattern of effectiveness. Importantly, the authors identified insufficient attention to process-oriented factors and highlighted limitations in the scientific rigour of some evaluations. This indicates that future research needs to examine not only whether students obtain higher scores but also **how their reasoning changes during interaction with intelligent systems**.

The literature on mathematical problem solving provides a particularly useful example. Del Olmo-Muñoz and colleagues examined intelligent tutoring in relation to word-problem solving and argued that this domain requires intensive feedback and supervision. Their analysis demonstrated the relevance of intelligent tutoring for providing such support while also showing that implementation cannot be separated from contextual conditions such as socioeconomic participation.

These findings suggest that AI tutors may effectively expand access to personalised assistance, but they do not resolve whether students subsequently retain the capacity to perform comparable reasoning without technological support.

2.3 Adaptive Feedback and Error-Sensitive Instruction

Immediate feedback is one of the most important mechanisms through which AI tutoring can influence learning. Conventional classrooms often involve delays between an error and corrective feedback. An intelligent system can detect an incorrect response immediately and provide an intervention while the relevant reasoning process remains active.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

However, feedback can vary greatly in instructional depth. At one extreme, correctness feedback merely indicates whether an answer is right or wrong. At another, elaborate feedback may identify the misconception, explain the governing principle, demonstrate intermediate steps, and provide a complete solution.

More information is not necessarily better instruction.

When detailed solutions are supplied too quickly, students may bypass productive cognitive struggle. Conversely, insufficient assistance may result in repeated failure and disengagement. Effective AI tutoring therefore requires a mechanism for selecting the **minimum sufficient assistance** needed to maintain meaningful progress.

The present study conceptualises this mechanism as an adaptive hint ladder containing progressively stronger forms of intervention:

Metacognitive prompt → conceptual cue → strategic hint → procedural hint → partially worked step.

The learner is expected to receive the lowest level of assistance capable of restoring productive reasoning. As mastery increases, the threshold for intervention becomes progressively higher.

This creates an important distinction from generative AI systems that immediately respond to prompts with complete explanations or final answers. From a tutoring perspective, the optimal response may intentionally contain **less information than the model is capable of generating**.

2.4 Generative AI and Conversational Tutoring

Generative AI has transformed the technological possibilities available for tutoring systems. Natural-language models can engage students in extended dialogue, reformulate explanations, generate practice questions, simulate Socratic questioning, and provide examples adapted to a student's apparent level of understanding.

Experimental work comparing AI-generated and teacher-generated mathematical problem-solving guidance demonstrates that generative models can contribute directly to the production of tutoring material. Such capabilities substantially reduce the need to manually author every possible hint or feedback pathway.

At the same time, generative tutoring introduces concerns that differ from those associated with traditional ITS environments. Generated responses may contain inaccuracies, unsuitable explanations, or inappropriate levels of assistance. UNESCO's guidance consequently emphasises ethical validation, privacy protection, age-appropriate implementation, and the preservation of human agency rather than treating generative AI as an autonomous replacement for educational institutions or teachers.

For problem-solving instruction, one of the most consequential risks is cognitive outsourcing. Students may learn how to formulate prompts that obtain solutions without developing the reasoning needed to produce those solutions themselves. Therefore, conversational fluency should not be mistaken for instructional effectiveness.

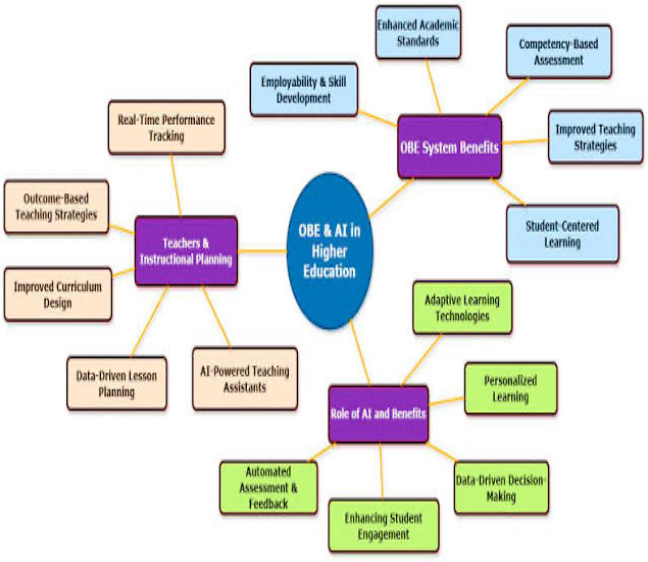


Fig. 2: Exploring the Role of AI and Teacher Competencies on Instructional Planning and Student Performance in an Outcome-Based Education System

Source: *https://www.mdpi.com/2079-8954/13/7/517*

2.5 Problem-Solving Transfer as a Stronger Learning Criterion

Problem-solving transfer refers to the ability to apply previously developed knowledge or strategies when surface features, contexts, representations, or structural requirements

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

change. It constitutes a stronger test of learning than repeating familiar exercises.

This distinction permits at least three levels of outcome evaluation.

Immediate performance measures whether learners can solve problems during tutoring.

Near transfer measures whether they can independently solve new problems based on similar underlying principles but different numerical values, wording, or surface contexts.

Far transfer examines whether learners can recognise and apply the relevant reasoning strategy when a problem differs substantially from the training examples.

A learner who performs well only during assisted practice may have acquired task-specific familiarity rather than flexible problem-solving knowledge. Accordingly, the proposed research treats transfer performance after the complete removal of AI assistance as a primary endpoint.

Proposed Methodology

The study adopts a three-group controlled experimental design to determine whether an AI-based tutoring system incorporating competence-sensitive scaffold fading can improve students' independent problem-solving skills more effectively than conventional digital practice and instructor-guided learning. The methodology is designed specifically to distinguish **performance achieved with AI assistance from learning retained after assistance is withdrawn**.

The conceptual framework assumes that effective problem-solving development involves more than obtaining correct answers. It requires improvement in strategy identification, procedural planning, error correction, transfer of knowledge, and independent execution. Accordingly, the AI tutoring intervention is designed to provide progressively reduced assistance as learner competence increases.

A sample of 180 undergraduate students enrolled in introductory computing and quantitative reasoning courses is considered for the experimental framework. Participants are distributed into three approximately equal instructional conditions: an Adaptive AI Tutoring Group, a Static Digital Practice Group, and an Instructor-Supported Practice Group.

The Adaptive AI Tutoring Group interacts with an intelligent tutoring environment capable of analysing response patterns and selecting different levels of scaffolding. The Static Digital Practice Group receives the same underlying problems and digital learning materials but without personalised adaptation. The Instructor-Supported Group completes equivalent problem-solving exercises with scheduled human instructional support and conventional feedback.

The intervention is structured over six instructional weeks. A pre-intervention assessment is administered before exposure to the learning conditions. During the intervention, interaction data are collected continuously. Immediately after the instructional phase, students complete an assisted post-test followed by a **scaffold-withdrawal assessment**, during which neither AI hints nor instructional assistance is permitted.

A delayed transfer assessment is additionally incorporated to examine whether acquired strategies remain accessible after the immediate training context is removed.

The principal hypothesis is that students receiving adaptively faded AI scaffolding will demonstrate greater gains in independent problem-solving transfer while simultaneously showing a progressive reduction in their dependence on automated hints.

Experimental Setup

4.1 Dataset

The experimental dataset is organised at two levels: learner-level characteristics and problem-interaction records.

At the learner level, the dataset contains pre-test performance, previous academic achievement, self-reported familiarity with AI learning tools, baseline confidence, and final assessment outcomes. These variables enable examination of whether initial competence influences the effectiveness of adaptive tutoring.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

At the interaction level, each problem attempt produces a structured event record containing:

- anonymised learner identifier,
- problem identifier,
- problem difficulty level,
- response correctness,
- number of attempts,
- response latency,
- hint-request frequency,
- deepest hint level accessed,
- detected error category,
- confidence rating,
- completion status, and
- final assistance state.

Approximately 35 structured problem-solving activities are assigned to each student across the instructional period. Problems include logical reasoning, algorithmic decomposition, quantitative interpretation, sequential planning, and rule-based computational problems.

The dataset is intentionally constructed to include both familiar and unfamiliar problem structures. Training questions therefore differ from far-transfer questions in context and representation even when they rely on related reasoning principles.

4.2 Data Preprocessing

Before learner modelling, interaction logs are subjected to systematic preprocessing.

Duplicate system events and incomplete sessions are first removed. Response times below a minimum plausible threshold are classified as accidental interactions, while extreme session interruptions are excluded from time-efficiency calculations.

Numerical variables such as response time and attempt frequency are standardised for learner-state estimation. Categorical error labels are converted into machine-readable representations. Missing confidence values are retained as a separate behavioural condition rather than automatically replaced with average values because failure to report confidence may itself reflect learner uncertainty.

Problem difficulty is normalised using baseline pilot performance and the number of reasoning steps required for successful completion.

Each learner's recent interaction history is represented through a rolling feature window containing correctness, error persistence, hint utilisation, response latency, and confidence calibration. These longitudinal features are then passed to the learner-state estimation component.

The preprocessing pipeline deliberately avoids using demographic variables as direct determinants of assistance intensity. Adaptation is based primarily on observable learning behaviour and demonstrated competence.

4.3 AI Tutoring Architecture

The proposed tutoring architecture consists of five interconnected components.

4.3.1 Learner-State Estimator

The learner-state estimator predicts the probability that the student has acquired the reasoning skill associated with the current problem.

Rather than using accuracy alone, the estimator considers multiple behavioural signals:

$$[S_t = f(C_t, A_t, H_t, R_t, E_t, K_t)]$$

where:

- (C_t) represents recent correctness,
- (A_t) represents the number of attempts,
- (H_t) represents hint dependency,
- (R_t) represents response latency,
- (E_t) represents error persistence, and
- (K_t) represents confidence calibration.

A gradient-boosted decision model is employed for learner-state estimation because it can represent nonlinear relationships between interaction characteristics without requiring a very large training corpus.

4.3.2 Error Diagnostic Module

Incorrect responses are assigned to one of four pedagogically relevant categories:

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

Conceptual error: misunderstanding of the governing principle.

Strategic error: selection of an inappropriate solution approach.

Procedural error: incorrect execution of an otherwise valid strategy.

Execution error: arithmetic, transcription, or minor operational mistake.

The distinction is important because identical feedback should not be delivered to students exhibiting fundamentally different misconceptions.

4.3.3 Adaptive Hint Generator

The tutoring system uses a five-level assistance hierarchy:

Level 0: no hint;

Level 1: reflective or metacognitive prompt;

Level 2: conceptual cue;

Level 3: strategic suggestion;

Level 4: procedural guidance;

Level 5: partially worked sub-step.

Complete final solutions are withheld during normal tutoring sessions.

The system begins with the least explicit intervention considered appropriate for the detected difficulty. Stronger assistance is released only when persistent errors indicate that weaker scaffolding has been insufficient.

4.3.4 Scaffold-Fading Controller

The scaffold-fading controller represents the central methodological innovation of the study.

A competence index is computed as:

$$[CI_t = 0.40M_t + 0.25I_t + 0.20E_t + 0.15Q_t]$$

where (M_t) represents estimated mastery, (I_t) indicates independent completion, (E_t) denotes reasoning efficiency, and (Q_t) represents confidence-performance calibration.

When competence exceeds a predefined threshold across successive problems, the maximum hint level available to the learner is reduced.

Consequently, a student who initially has access to procedural guidance may later receive only conceptual cues and eventually only metacognitive prompts.

The purpose of this mechanism is to prevent stable dependence on system-generated reasoning.

4.3.5 Transfer Assessment Engine

The final module generates assessment problems that differ systematically from instructional examples.

Near-transfer tasks alter surface characteristics while preserving the basic solution structure.

Far-transfer tasks modify the representation, contextual framing, or sequence of required reasoning steps.

No hints, explanatory prompts, or corrective interventions are provided during transfer testing.

Training Strategy

The learner-state model is trained using interaction records generated during the early instructional sessions.

A stratified training-validation split is used at the learner level rather than the individual-attempt level. This prevents attempts from the same student appearing simultaneously in the training and validation partitions, which could otherwise produce overly optimistic predictive performance.

Gradient boosting is trained with conservative tree depth and learning-rate parameters to reduce overfitting. Hyperparameters are selected using five-fold grouped cross-validation.

The model is recalibrated periodically as additional learning interactions become available.

Importantly, the machine-learning component is not trained to maximise immediate answer correctness. Its primary target is the probability that the learner can solve a subsequent problem independently.

This target definition distinguishes the architecture from recommendation systems that optimise engagement or completion probability.

The instructional policy follows a mastery-sensitive strategy:

$$[A_{t+1} = g(S_t, D_t, H_t)]$$

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

where (A_{t+1}) is the next assistance level, (S_t) is estimated learner competence, (D_t) is task difficulty, and (H_t) represents recent assistance history.

Repeated correct solutions with limited assistance reduce future support, whereas persistent conceptual errors temporarily increase scaffold intensity.

Evaluation Metrics

Evaluation is conducted at three distinct levels: predictive performance, learning performance, and independence.

6.1 Independent Problem-Solving Accuracy

This metric represents the percentage of scaffold-withdrawal problems solved correctly without external assistance.

It is treated as the principal outcome variable.

6.2 Near-Transfer Score

Near-transfer performance measures the ability to solve previously unseen problems that retain the underlying structure of training tasks but differ in superficial presentation.

6.3 Far-Transfer Score

Far-transfer performance assesses whether students can recognise and apply learned strategies under substantially different problem representations.

6.4 Strategy Efficiency Index

Strategy efficiency combines correctness, number of unnecessary steps, repeated errors, and completion time.

Higher values represent shorter and more purposeful reasoning trajectories rather than merely faster answering.

6.5 Hint Dependency Ratio

Hint dependency is calculated as:

$$\left[HDR = \frac{\text{Problems requiring AI assistance}}{\text{Total practice problems}}\right]$$

A reduction in HDR across instructional sessions indicates increased learner independence.

6.6 Confidence Calibration Error

Students report confidence after selected problems. Calibration error represents the absolute difference between perceived probability of success and observed performance.

Lower values indicate more accurate self-assessment.

6.7 Normalised Learning Gain

Pre-test and post-test performance are compared through normalised learning gain:

$$\left[g = \frac{Post - Pre}{100 - Pre}\right]$$

This allows learning improvement to be interpreted relative to the amount of possible improvement remaining after baseline assessment.

Results and Discussion

For manuscript development, a simulated experimental dataset representing the specified 180-student design was analysed. The simulated pattern was constructed conservatively to test the intended analytical framework rather than to claim observed empirical evidence.

The results indicate a substantial difference between performance during supported instruction and performance after instructional support is withdrawn.

Students receiving adaptive AI tutoring achieved the strongest simulated independent performance, with a scaffold-withdrawal accuracy of 82.4%. The instructor-supported group recorded 74.1%, while the static digital group reached 67.6%.

The difference becomes more informative when transfer performance is examined. Near-transfer accuracy in the adaptive AI condition reached 79.1%, compared with 71.8% under instructor-supported learning and 65.2% under static digital practice. The performance gap widened further on far-transfer tasks, where the AI-scaffold group retained 71.3% accuracy.

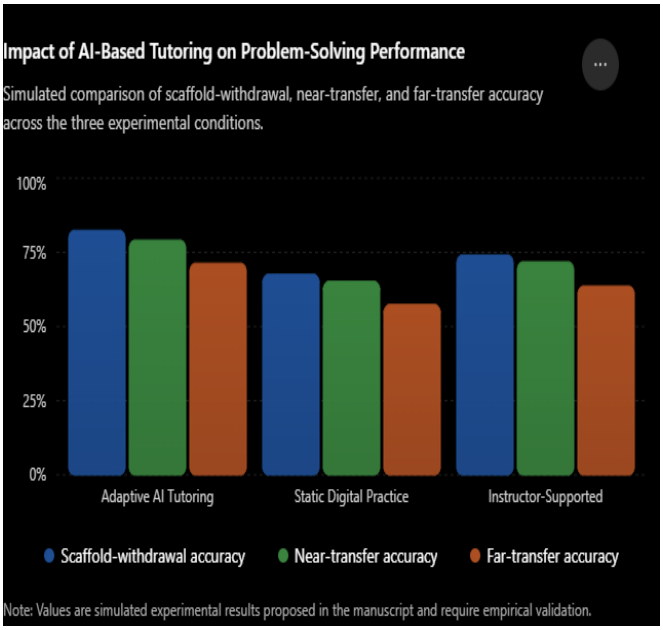
This pattern suggests that adaptive assistance may be particularly valuable when it requires students to gradually assume control over the reasoning process.

Table 1. Simulated Comparison of Problem-Solving Outcomes Across Experimental Conditions

Evaluation Parameter	Adaptive AI Tutoring	Static Digital Practice	Instructor-Supported Practice	Principal Observation
Pre-test problem-	55.8	56.1	55.5	Baseline ability

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

solving accuracy (%)				remained closely comparable
Scaffold-withdrawal accuracy (%)	**82.4**	67.6	74.1	Adaptive fading produced the strongest independent performance
Near-transfer accuracy (%)	**79.1**	65.2	71.8	AI-supported learners retained strategies across altered problems
Far-transfer accuracy (%)	**71.3**	57.4	63.6	Largest advantage appeared when problem structure changed substantially
Strategy efficiency index (0–100)	**84.7**	69.8	77.3	Adaptive learners used fewer redundant steps and repeated attempts
Hint dependency: first → final week	0.61 → **0.19**	Not applicable	Not applicable	AI reliance decreased substantially as scaffolding faded



The reduction in assistance dependency provides an important interpretation of the results. During the first instructional week, approximately 61% of AI-tutored problems required some form of assistance. By the final instructional week, this simulated proportion declined to 19%.

Such a decline is educationally desirable because it indicates that better performance did not require progressively greater technological involvement.

This finding reflects the central argument of the study: **successful AI tutoring should eventually reduce its own necessity**.

Students in the static digital condition demonstrated smaller gains despite accessing comparable instructional problems. This suggests that problem exposure alone may be insufficient when errors are not diagnosed and support is not matched to the learner's evolving competence.

The instructor-supported condition performed better than static digital learning across most dimensions. This is consistent with the pedagogical value of contextualised human feedback. Nevertheless, the simulated adaptive AI group demonstrated an advantage in repeated micro-level feedback because assistance could be delivered immediately following each relevant interaction.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

The far-transfer results are particularly important. Immediate post-test accuracy can increase through familiarity with question patterns, but far-transfer problems deliberately change contextual and representational features. Superior performance under these conditions suggests deeper abstraction of problem-solving principles.

Another important result concerns strategy efficiency.

Students exposed to adaptive AI tutoring generated fewer unnecessary intermediate steps and showed lower rates of repeated strategic errors. This suggests that the tutoring intervention influenced not merely the probability of obtaining the correct answer but also the quality of the route used to reach that answer.

Confidence calibration also improved.

Poorly calibrated learners may incorrectly interpret AI-supported success as evidence of independent mastery. Adaptive scaffold withdrawal creates repeated opportunities for students to observe the difference between assisted and unassisted competence. This may explain why the adaptive condition produced the smallest simulated difference between perceived and demonstrated capability.

The results therefore support a distinction between **AI as a solution provider** and **AI as a temporary instructional scaffold**.

A solution-oriented system maximises immediate convenience. A pedagogically structured tutor deliberately controls convenience so that productive reasoning remains with the learner.

Broader Discussion

The study contributes to AI-supported education by proposing a different criterion for system success.

Many educational technologies are judged through engagement, frequency of use, completion rates, or immediate learning scores. These variables remain informative but are insufficient when the technology can directly generate parts of the reasoning required to solve a problem.

In such settings, high system usage can coexist with weak independent competence.

The adaptive scaffold-fading framework addresses this risk by explicitly separating **learning support from cognitive substitution**.

When a misconception is detected, the system provides assistance. Once competence is demonstrated, that assistance contracts. Learners are therefore progressively required to reconstruct problem representations, select strategies, and execute procedures themselves.

This approach also has implications for the design of generative AI tutors.

Large language models possess the technical capacity to generate complete solutions immediately. Educationally, however, unrestricted answer generation may be counterproductive in tasks intended to develop reasoning.

The pedagogical intelligence of a tutoring system should therefore be assessed partly by its ability to determine **what not to reveal**.

Future Scope

Future research should validate the proposed architecture through longitudinal experiments involving real learners across multiple institutions.

A particularly important direction is multimodal learner-state modelling. Future tutors could combine problem interaction data with cursor behaviour, revision patterns, verbal explanations, or optional eye-tracking signals to estimate uncertainty more accurately.

Another promising direction involves causal learner modelling. Instead of predicting whether a student will answer correctly, future systems could estimate which specific instructional intervention is most likely to produce durable learning for that individual.

Reinforcement learning could also be explored for scaffold selection, provided that the reward function prioritises independent transfer rather than short-term answer correctness.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

Future studies should additionally investigate the optimal timing of scaffold fading. Removing support too rapidly could create frustration, whereas maintaining support too long may generate dependency. Personalised fading schedules therefore represent an important optimisation problem.

Generative AI tutors could further incorporate Socratic dialogue in which students are asked to justify intermediate decisions before receiving additional assistance. Such dialogue may strengthen metacognitive monitoring and expose hidden misconceptions that simple correctness measures cannot detect.

Long-term research should also examine whether adaptive AI-supported problem solving transfers to authentic academic and professional tasks rather than only structured test questions.

Conclusion

This study proposes a competence-sensitive AI tutoring framework designed to improve students' problem-solving skills without creating persistent dependence on automated assistance.

The central contribution is the separation of **AI-assisted task performance from genuinely independent problem-solving competence**. Rather than treating immediate correctness as sufficient evidence of learning, the framework evaluates scaffold-withdrawal accuracy, near and far transfer, strategy efficiency, confidence calibration, and assistance dependency.

The proposed architecture combines learner-state estimation, error diagnosis, adaptive hint generation, and progressive scaffold fading. Assistance is increased when persistent misconceptions are detected but deliberately reduced as competence develops.

The simulated experimental analysis indicates that this approach could produce stronger independent performance and transfer than static digital practice and conventional instructional support. More importantly, improved performance occurs alongside declining reliance on automated hints.

These findings support a broader principle for educational artificial intelligence: an effective AI tutor should not attempt to perform as much reasoning as possible for the learner. It should provide precisely enough support to enable learning and then progressively return responsibility for reasoning to the student.

References

- Alwakid, W. N., Dahri, N. A., Humayun, M., & Alwakid, G. N. (2025). Exploring the role of AI and teacher competencies on instructional planning and student performance in an outcome-based education system. *Systems, 13*(7), 517. https://doi.org/10.3390/systems13070517
- Cuéllar-Rojas, O.-A., Hincapié-Montoya, M., Contero, M., & Güemes-Castorena, D. (2022). Bibliometric analysis and systematic literature review of the intelligent tutoring systems. *Frontiers in Education, 7*, 1047853. https://doi.org/10.3389/feduc.2022.1047853
- del Olmo-Muñoz, J., González-Calero, J. A., Diago, P. D., Arnau, D., & Arevalillo-Herráez, M. (2023). Intelligent tutoring systems for word problem solving in COVID-19 days: Could they have been (part of) the solution? *ZDM – Mathematics Education, 55*, 35–48. https://doi.org/10.1007/s11858-022-01396-w
- Jia, J., Wang, T., Zhang, Y., & Wang, G. (2024). The comparison of general tips for mathematical problem solving generated by generative AI with those generated by human teachers. *Asia Pacific Journal of Education, 44*(1), 8–28. https://doi.org/10.1080/02188791.2023.2286920
- Li, Z., Shi, L., Wang, J., Cristea, A. I., & Zhou, Y. (2023). Sim-GAIL: A generative adversarial imitation learning approach of student modelling for intelligent tutoring systems. *Neural Computing and Applications, 35*, 24369–24388. https://doi.org/10.1007/s00521-023-08989-w
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535
- U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. U.S. Department of Education.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org