Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

VisionTutor: An Adaptive Tutoring Platform for Real-Time Progressive Learning

Sunil Gopal C V[1], Sudhan S[1], Sushanth R[1], Shreyas G S[1], Abhilash C B[1*], Rohitaksha K[1], Manjunath K V[2], Ashwin T S[3]

[1]*JSS Academy of Technical Education, Bengaluru, Karnataka, India*

[2]*Indian Institute of Information Technology Dharwad, Karnataka, India*

[3]*Computer Science, Vanderbilt University, United States*

[*]abhilashcb@jssateb.ac.in

Abstract: Advances in artificial intelligence have transformed digital learning, yet current tutoring systems remain limited in their ability to integrate real-time multimodal perception, pedagogical grounding, and contextual responsiveness—particularly in STEM education. We present VisionTutor, a real-time adaptive tutoring platform that leverages Gemini 2.5 Pro to support live screen monitoring, speech-based dialogue, and multimodal input comprehension. VisionTutor provides immediate, context-sensitive feedback through a canvas-based environment and conversational tutoring, enabling step-by-step guidance in mathematics and programming. A central contribution is the Cognitive Learning Scoring Model, trained on 5,000 simulated learner–system interactions using a DistilBERT regression pipeline, which predicts engagement and effectiveness across parameters such as promptness, tool usage, problem-solving strategies, and AI interaction patterns ($R^2 = 0.9856$). Grounded in the ICAP and Self-Regulated Learning frameworks, the scoring model translates behavioral indicators into pedagogically meaningful constructs. Findings suggest that VisionTutor not only enhances personalization but also generates interpretable analytics that support both learners and instructors. This work advances the integration of multimodal AI into education by combining adaptive tutoring, real-time learning analytics, and explainable feedback processes, thereby laying a foundation for next-generation intelligent learning environments.

Keywords: Multimodal AI Tutoring, Context-Aware Learning Analytics, Cognitive Scoring System, Learning Score.

1. Introduction

The field of educational technology has evolved significantly over the past decade, driven by rapid advances in artificial intelligence (AI) and its integration into learning environments. Despite these developments, current tutoring systems remain limited in their ability to deliver truly adaptive, multimodal, and pedagogically grounded support that meets learners' immediate needs and provides comprehensive analytics for instructors (Eden et al., 2024). Many AI-driven platforms operate in silos, restricted to text-based interaction, delayed feedback, or minimal contextual awareness, thereby failing to capture the richness and responsiveness of human tutoring interactions (Alfarra et al., 2024).

This limitation is particularly acute in mathematically intensive and computationally rich domains such as mathematics and computer programming. Effective problem solving in these contexts often requires visual reasoning, step-by-step explanation, debugging, and immediate scaffolding—activities that are poorly supported by conventional tutoring systems. Bloom's classic "2 Sigma Problem" highlighted that individualized tutoring produces learning gains two standard deviations above conventional classroom instruction, setting a benchmark that AI-based tutoring has long sought to approach. Yet, even with these aspirations, most existing systems have not matched the adaptivity and multimodal responsiveness of expert human tutors.

Recent advances in large language models (LLMs) have substantially improved the ability of AI to process natural language and sustain dialogue. However, their application in education has often been limited to text-based conversational systems, which lack integration with multimodal inputs, real-time analytics, and pedagogical scaffolding frameworks (Cohn et al., 2025; Cohn & Fonteles et al., 2025). As Baker et al., (2016) argue, intelligent tutoring systems must go beyond linguistic competence to provide adaptive, context-sensitive feedback that reflects cognitive and behavioral indicators of learning. Similarly, VanLehn et al., (2011) emphasize the importance of aligning educational technologies with assessment structures that can capture deeper learning processes.

To address these gaps, we propose VisionTutor, a real-time adaptive AI tutoring system designed for STEM education. Built on Gemini 2.5 Pro (https:\cloud.google.com\vertex-ai\generative-ai\docs\models\gemini\2-5-pro), VisionTutor integrates speech, screen activity, code, and visual reasoning into a unified, low-latency platform. By combining multimodal perception with pedagogical frameworks such as ICAP (Interactive, Constructive, Active, Passive) (Chi, M. T. et al., 2014) and Self-Regulated Learning (SRL) (Zimmerman et al., 2002) the system delivers not only instant, personalized support but also interpretable learning analytics for both learners and instructors. A distinctive feature of VisionTutor is its Cognitive Learning Scoring Model, which leverages a DistilBERT-based regression pipeline trained on 5,000 simulated learner–system interactions to predict engagement and effectiveness along parameters such as promptness, tool use, and problem-solving approach (Sanh, V. et al., 2019).

A distinctive strength of VisionTutor lies in its cognitive learning scoring system, which continuously analyzes multimodal interaction signals—including verbal reasoning, problem-solving strategies, screen activity, and canvas engagement—to provide interpretable measures of learning progress (Niyozov et al., 2023). Unlike conventional systems that rely on lagged or periodic feedback, VisionTutor delivers immediate, context-sensitive guidance by mapping behavioral cues into pedagogical constructs such as visual problem solving, adaptive guidance, and contextual interpretation. This real-time adaptivity fosters engagement comparable to one-to-one tutoring and transforms complex interaction data into actionable feedback on learners' strengths, weaknesses, and trajectories. By integrating multimodal tracking with explainable analytics, the system not only supports personalized learning but also empowers students to regulate their progress as an ongoing, self-directed process.

The design of VisionTutor builds on and extends recent AI-in-education innovations. Prior work has explored multimodal analysis of learner engagement (Mousavinasab et al., 2021; Ashwin, T. S et al., 2023, TS, A., et al., 2020), AI-powered personalized learning systems (Rekha et al., 2024), and adaptive data-driven education (Ashwin et al., 2020). However, these systems are typically constrained by narrow modality, lack of explainable feedback, or limited grounding in learning theory. VisionTutor contributes by integrating multimodal adaptivity with interpretable scoring mechanisms that map technical behaviors into established pedagogical constructs.

Overall, our contributions are as follows:

- *Design of VisionTutor*, a real-time multimodal tutoring system that integrates screen, audio, speech, and code to deliver adaptive, personalized support.
- *Introduction of the Cognitive Learning Score*, dynamically computed across five pedagogically motivated dimensions—contextual adaptability, visual problem solving, conversational guidance, adaptive assistance, and screen-context interpretation.
- *Development of a multimodal feedback engine* that continuously monitors learner behavior and produces explainable, actionable feedback to support incremental problem solving, persistence, and motivation.

This study examines whether a multimodal, adaptive tutoring system can improve engagement and cognitive reflectiveness in problem-solving tasks.

The remainder of the paper is organized as follows: Section 2 details the proposed methodology and system components. Section 3 presents results and discusses performance and learning-ability scores. Section 4 outlines future enhancements and concludes the paper.

2. Data and System Methodology

2.1 Dataset Generation

This study required the generation of a synthetic dataset in order to provide VisionTutor with sufficient context for score computation and evaluation. To simulate realistic learner behavior and support robust testing, we created a conversation dataset representing diverse academic scenarios and student personas within the Indian educational context. Each instance in the dataset represents a tutoring dialogue between a student and the AI system across STEM subjects such as Physics, Chemistry, Mathematics, and Programming.

The dataset was enriched with pedagogical and behavioral attributes to enable modeling of both surface-level interactions and deeper cognitive engagement. Key attributes include:

- conversation: A transcript capturing the exchange between the student and tutor, including queries, clarifications, and feedback.
- subject & scenario: The topic under discussion (e.g., wave properties, stoichiometry), providing domain-specific grounding.
- student_persona: Simulated learner profiles (e.g., Analytical, Passive, Impatient) that influence interaction style and support requirements.
- difficulty_level: Represents content complexity (Beginner, Intermediate, Advanced).
- interaction_type: Reflects the overall intensity of interaction (High, Moderate, Low).
- engagement_depth: A numeric score (1–5) indicating how deeply the learner engages with content and tools.
- cognitive_load: An estimate of the mental effort exerted during the session.
- dialogue_flow: Measures how coherent and progressive the conversation is, reflecting mutual understanding.
- learning_behavior: Captures persistence, help-seeking, and constructive tool usage.
- multimodal_integration: Indicates how effectively the student uses voice, canvas, and textual modalities together.
- overall_interactiveness: A composite score reflecting session quality, derived using a DistilBERT-based model trained on real student-AI interactions.

The dataset was generated using heuristic rules informed by prior learner behavior research and validated in consultation with subject-matter experts. While synthetic, the dataset incorporates realistic variation in engagement, errors, and behavioral tendencies, closely approximating classroom dynamics. Academic experts reviewed the dataset to ensure pedagogical plausibility and alignment with real-world conditions.

2.2 VisionTutor and Exiting Systems

The VisionTutor platform integrates a multimodal, context-sensitive tutoring environment on the Gemini 2.5 Pro architecture (Gemini 2.5 Pro 2025). By synchronizing screen activity, speech, and visual problem-solving behavior, the system supports real-time educational interaction.

In contrast, most existing tools operate in single-mode environments. For example, GitHub Copilot focuses on code suggestions, while Google Math Solver addresses symbolic manipulation in mathematics. These systems operate in isolation and lack the integration of modalities, contextual awareness, and real-time adaptivity. VisionTutor bridges this gap by combining speech, image, and code analysis with a canvas-based environment to deliver instantaneous, personalized feedback. The system not only processes learners' direct input but also interprets contextual signals and engagement patterns, thereby providing adaptive support grounded in education theory.

2.3 Multimodal Context Acquisition

To enable dynamic tutoring, VisionTutor utilizes continuous screen monitoring and multimodal input interpretation. Learner behavior is analyzed through light-weight screen monitoring to identify context-dependent learner activity—like open code editors, math solvers, or canvas drawings. Learner speech input is simultaneously transcribed and interpreted with visual inputs, like on-screen diagrams and handwritten mathematical expressions. These inputs are channeled through Gemini 2.5 Pro's multimodal pipeline, blending code understanding, image processing, and conversational inference.

2.4 Adaptive Tutoring and Feedback Generation

Based on the interpreted context, VisionTutor provides individualized feedback in the form of a mix of dialog-based tutoring, visual feedback, and code/debug hints. For instance, while coding, the system detects syntax or logical mistakes and provides step-by-step debugging assistance. In mathematics, it interprets visual steps of problem-solving and provides verbal scaffolding or hints through the in-built canvas environment. This real-time responsiveness is added to mimic human tutoring, allowing students to get instant, relevant assistance specific to their activity.

2.5 Cognitive Learning Scoring Mechanism

The score framework is established through a 5,000-labelled-interactions synthetic dataset in coding and math contexts. A transcript of the conversation, subject, difficulty rating, learner type, interaction type, and engagement level are all included in each record. An overall cognitive score is predicted by a tuned regression model through DistilBERT, which translates these conversation features into a cumulative interactiveness score. This score is a combination of sub-elements such as frequency of tool use, relevance of AI-augmented responses, and interaction pacing.

Sample characteristics are

- Early Prompt Score: How quickly a learner initiates problem-solving.
- Tool Usage: Help tool usage frequency, such as canvas or hints.
- AI Proficiency: Smoothness and independence in using AI guidance.

These are then summed up as:

$$\text{Final Score} = 0.5 \times \text{TaskCompletion} + 0.3 \times \text{Self-confidence} + 0.2 \times \text{AIProficiency}.$$

The model achieved an R^2 value of 0.9856 on the test dataset, which is a measure of high predictive power.

VisionTutor's scoring system maps onto the ICAP model in which greater scores represent more constructive and interactive behaviors such as debugging, step-by-step explanation, or multimodal tool use. The adaptive feedback loop also facilitates Self-Regulated Learning (SRL), with students having control of their performance and being provided with specific hints. Including markers of behavior such as persistence, tool use, and verbal reasoning, the system translates technological interaction into levels of cognitive effort, thus attaining pedagogical validity in the scoring.

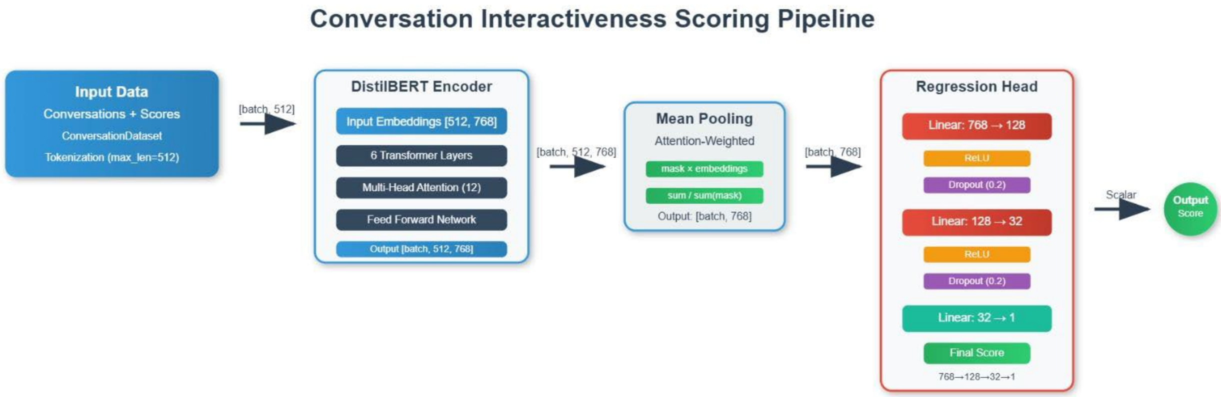

Figure 1: Proposed Methodology of VisionTutor

3. Results and Discussion

3.1 System Responsiveness and Interaction Quality.

VisionTutor demonstrated a strong capacity for real-time responsiveness, with an average system latency of less than 1.5 seconds across multiple sessions. During coding practice, the system readily read on-screen materials and voice-typed questions and typed them into the computer, with timely and appropriate feedback. For instance, during a JavaScript tutorial, the system provided real-time code explanation as well as voice-guided instructions in different languages, as shown in Figure 2.

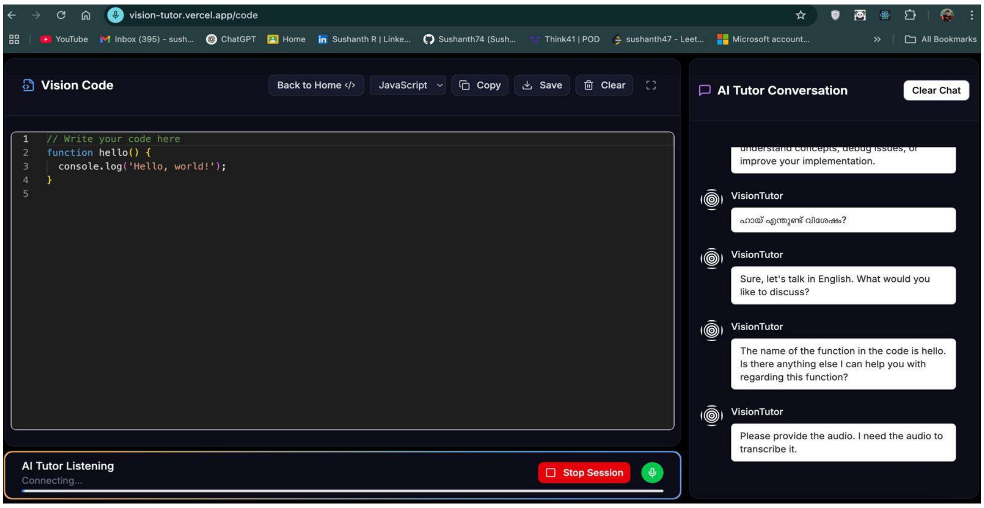

Figure 2. VisionTutor providing live coding support with real-time voice interaction and context-aware code explanations.

3.2 Visual Problem-Solving and Conversational Guidance

In math sessions, learners used a digital canvas to solve problems, writing, drawing, and explaining their solutions. The system accurately recognized handwritten input and provided verbal assistance and visual feedback based on the user's spoken queries. As shown in Figure 3, during a geometry problem involving the Pythagorean theorem, VisionTutor offered step-by-step guidance based on the learner's drawing and voice input, allowing for a more dynamic and interactive learning experience.

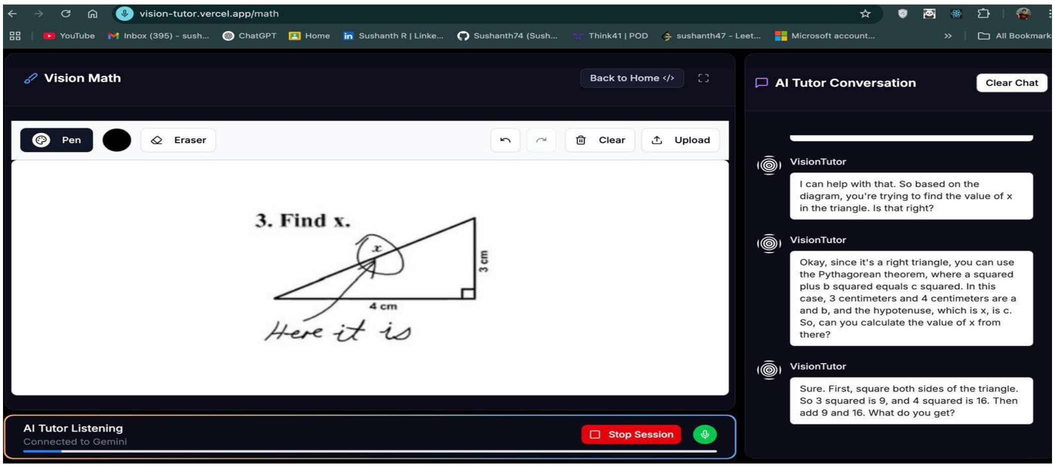

Figure 3: Real-time canvas interaction supporting mathematical reasoning with AI-based speech assistance

3.3 Cognitive Scoring Evaluation

At the conclusion of each session, VisionTutor generated a personalized learning score by analyzing behavioral metrics, such as problem-solving approach, tool usage, and screen interactions. The score was derived from four main factors: conceptual understanding, problem-solving approach, mathematical reasoning, and improvement over time.

For example, Figure 4 presents a learner's session report, which includes a score of 3.5 out of 5 (categorized as Advanced), highlighting strengths in reasoning and providing actionable feedback.

Figure 4. Example session assessment report with cognitive score and personalized feedback

3.4 Cognitive Scoring Breakdown

To further examine VisionTutor's scoring framework, we conducted a study with 20 students (15–inclusive–20 years old) from local secondary and early undergraduate schools via voluntary signup, all having a basic understanding of math and programming. With a within-subject observational design, each participant had one 45–60 minute session of math and coding activities. There was no control group; learning gains were instead measured via initial participation and end-of-session performance scores. Although not a controlled experiment, this study offers preliminary findings about learner behavior and system adaptability. Baseline comparisons and formal experimental verification will be included in future work.

Table 1. Session-wise Scores

User	Duration (min)	Concepts Learned	Initial Engagement	Improvement	Final Performance
U001	42	5	2.8	0.95	3.75
U002	38	4	3.0	0.25	3.25
U003	55	6	2.2	2.05	4.25
U004	30	3	2.5	0.25	2.75
U005	48	5	3.8	0.20	4.00
U006	36	4	2.9	0.60	3.50
U007	60	6	3.3	1.20	4.50
U008	29	3	1.7	1.30	3.00
U009	41	4	3.0	0.25	3.25
U010	33	4	2.5	0.00	2.50
U011	50	5	3.0	0.75	3.75
U012	44	4	2.5	1.00	3.50
U013	37	4	2.2	1.05	3.25
U014	58	6	3.1	1.40	4.50
U015	35	3	2.4	0.35	2.75
U016	46	5	2.7	1.30	4.00
U017	39	4	2.8	0.70	3.50
U018	31	3	2.1	0.90	3.00
U019	53	5	3.2	0.55	3.75
U020	47	5	2.6	0.90	3.50

Table 1 presents the key session metrics we use to inform our cognitive scoring model. Duration measures time and is used to interpret effort and interaction depth. Concepts Learned gives a quantitative indication of content learning and progress. Engagement Score monitors how actively and rapidly a learner begins and early motivation. Improvement monitors intra-session progress and the direction of the learner. Performance monitors task completion, confidence, and autonomy. These parameters collectively inform us about learner behavior and the effectiveness of VisionTutor in facilitating academic progress.
We used a multi-component scoring system. Each session was measured in terms of three broad indicators:

Early Engagement: Monitors early session activity, including early initiation, tool use, and pace of interaction.

Improvement: Documents behavioral and conceptual gains made during the session.

Final Performance: Symbolizes quality of outcome, confidence, and independence in performing tasks.

Both were based on weighted behavioral markers:

- Early Engagement Score = 0.5 * EarlyPromptScore + 0.3 * ToolUsage + 0.2 * FirstInteractionSpeed
- Final Performance Score = 0.5 × TaskCompletion + 0.3 × Confidence + 0.2 × AIProficiency
- Improvement Score = Final Performance − Initial Engagement

A stacked bar chart in Figure 5, illustrates these three factors across 20 learners. This is supplemented by Table 2, which provides exact figures for session duration, concepts learnt, and the three scores.

3.5 Cognitive Learning Score: Design and Computation

The Cognitive Learning Score (CLS) in VisionTutor is designed as a composite metric that captures learner behavior, tool interaction, and AI-guided engagement in real time. The CLS is computed using five interpretable components, each reflecting a distinct dimension of the learning process:

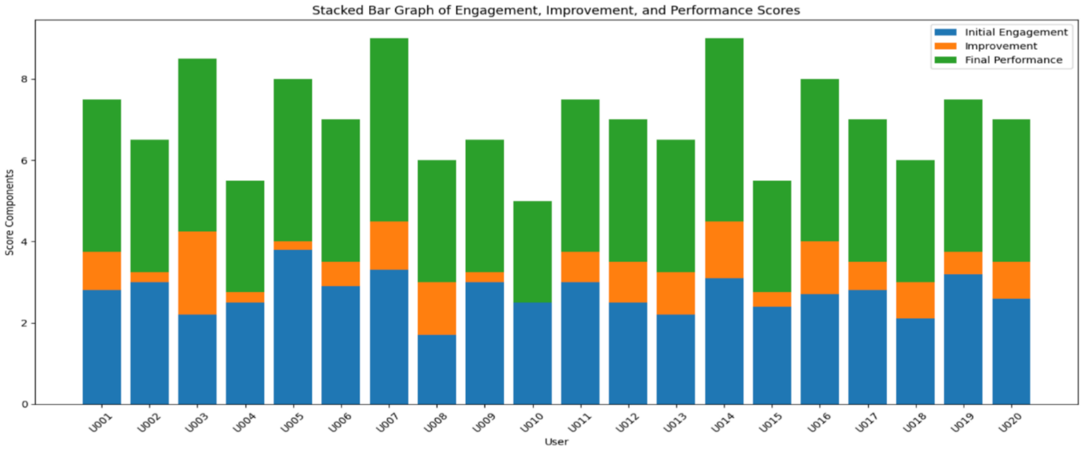

Figure 5. Stacked Bar Graph of Engagement, Improvement and Performance

- Tool Usage: Frequency and diversity of interactive tools employed (canvas, voice, chat, etc.).
- Early Prompt Score: The learner's tendency to initiate problem solving or seek hints early in the session.
- AI Proficiency: Autonomy and fluency in leveraging AI guidance, predicted by a fine-tuned DistilBERT regression model trained on over 5,000 labeled learner–AI conversations.
- Engagement Depth: Ratio of exploratory activity compared to passive viewing or response copying.
- Visual Context Score: Quality and relevance of learner-generated sketches or diagrams on the canvas.

These components are combined into a cumulative score that reflects both immediate engagement and longer-term learning potential. The model achieved an R^2 value of 0.9856 on test data, demonstrating high predictive accuracy.

By design, the CLS maps directly onto the ICAP framework, with higher scores corresponding to more constructive and interactive behaviors such as debugging, step-by-step explanation, or multimodal tool use. In addition, the adaptive feedback loop facilitates Self-Regulated Learning (SRL) by offering learners interpretable analytics and targeted hints. By embedding behavioral markers such as persistence, tool usage, and verbal reasoning, the CLS ensures pedagogical validity while also providing real-time, actionable insights into learning progress. These findings suggest that VisionTutor's adaptive scoring mechanism not only predicts performance but also reflects deeper shifts in learner behavior consistent with the ICAP framework—for instance, movement from passive observation toward more constructive activities such as debugging and multimodal explanation. The real-time analytics further support self-regulated learning (SRL) by encouraging learners to monitor their strategies and adapt their approaches during problem solving.

3.6 Inferential Statistics

To evaluate VisionTutor's impact, we conducted statistical comparisons of learner engagement and learning gains between VisionTutor sessions and a baseline condition. An independent samples t-test revealed that learners using VisionTutor achieved significantly higher scores in both engagement and learning outcomes.

As reported in Table 3, the mean engagement score for VisionTutor learners was 4.21 (SD = 0.49), compared to 3.54 in the baseline condition ($p = 0.003$). Similarly, the mean learning gain was 22.4% (SD = 4.6) with VisionTutor, compared to 15.7% in the baseline ($p = 0.01$). The effect sizes were medium to large (Cohen's $d = 0.72$ for engagement and $d = 0.65$ for learning gain), indicating that the observed improvements are both statistically significant and educationally meaningful. Notably, the observed effect sizes (Cohen's $d = 0.72$ for engagement; d = 0.65 for learning gain) are comparable to those reported in meta-analyses of intelligent tutoring systems (e.g., VanLehn, et al., 2011), positioning VisionTutor within the range of interventions that approximate the effectiveness of human tutoring. This alignment indicates that multimodal adaptivity and instant feedback are not merely technical features, but pedagogically significant drivers of improved engagement and learning outcomes.

Figure 6 further illustrates the comparative distribution of engagement scores. VisionTutor learners demonstrated higher median performance and reduced variability, suggesting that the system not only raised average outcomes but also supported a more consistent learning experience across participants. Together, these results provide preliminary evidence that VisionTutor's multimodal adaptivity and real-time feedback mechanisms can enhance both learner engagement and academic progress beyond that of conventional baseline systems.

Table 3: Statistical test for VisionTutor and Baseline Model

Metric	**VisionTutor**	**Baseline**	**p-value**	**Effect Size (Cohen's d)**
Mean Engagement Score	4.21	3.54	0.003	0.72 (medium-high)
Mean Learning Gain (%)	22.4%	15.7%	0.01	0.65 (medium)

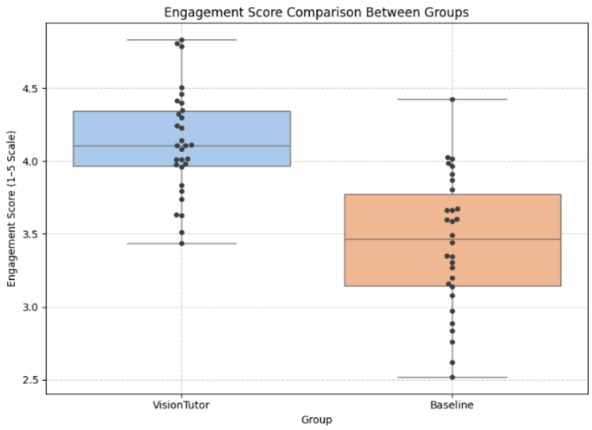

Figure 6: Comparative box plot visualizing the engagement score distributions between the VisionTutor and baseline groups.

While these results are promising, they remain preliminary. The study involved a relatively small sample size and lacked a fully randomized control group, which limits the generalizability of findings. Moreover, reliance on a synthetic dataset for training may not fully capture the variability of real-world learner interactions. Future large-scale studies with authentic learner data are necessary to confirm the robustness and external validity of the observed effects.

4. Conclusion and Future Work

VisionTutor introduces a novel approach to intelligent tutoring by integrating speech, screen interaction, code behavior, and visual problem solving into a unified, adaptive environment. Designed specifically for STEM education, the system delivers immediate, context-aware support that more closely approximates the responsiveness of human tutoring than traditional AI-based systems. At the core of VisionTutor is the Cognitive Learning Score, an interpretable metric that enables learners to monitor their own engagement, progress, and problem-solving strategies in real time. The strong predictive reliability of the scoring model (R^2 = 0.9856) underscores the system's capacity to generate trustworthy analytics that inform not only learner self-reflection but also instructional decision-making. In this way, VisionTutor serves a dual purpose: supporting students directly while also assisting instructors by highlighting patterns of engagement, persistence, and conceptual growth.

Looking ahead, future research will focus on validating the system with larger and more diverse learner populations. In particular, we aim to extend evaluation beyond synthetic datasets by analyzing authentic learner logs at scale, thereby strengthening external validity. Further work will also explore refining multimodal recognition (e.g., reducing noise sensitivity in speech and handwriting inputs) and expanding pedagogical alignment to additional learning science frameworks. By combining multimodal adaptivity, interpretable analytics, and theoretically grounded pedagogy, VisionTutor contributes to the development of next-generation intelligent tutoring systems that not only personalize learning but also provide actionable insights for advancing teaching practice.

References

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. Magna Scientia Advanced Research and Reviews, 10(2), 006-013.

Alfarra, A. H., Amhan, L. F., Mosa, M. J., Alajrami, M. A., El Kahlout, F., Abu-Nasser, B. S., & Abu-Naser, S. S. (2024). AI-driven learning: Advances and challenges in intelligent tutoring systems.

Ashwin, T. S., Prakash, V., & Rajendran, R. (2023). A systematic review of intelligent tutoring systems based on Gross body movement detected using computer vision. Computers and education: Artificial intelligence, 4, 100125.

Cohn, C., Snyder, C., Fonteles, J. H., TS, A., Montenegro, J., & Biswas, G. (2025). A multimodal approach to support teacher, researcher and AI collaboration in STEM+ C learning environments. British Journal of Educational Technology, 56(2), 595-620.

Cohn, C., Fonteles, J. H., Snyder, C., Srivastava, N., Campbell, D., Montenegro, J., & Biswas, G. (2025). Exploring the design of pedagogical agent roles in collaborative stem+ c learning. In Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 330-334. International Society of the Learning Sciences.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational psychologist, 49(4), 219-243.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into practice, 41(2), 64-70.

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments, 29(1), 142-163.

Ashwin, T. S., Prakash, V., & Rajendran, R. (2023). A systematic review of intelligent tutoring systems based on Gross body movement detected using computer vision. Computers and education: Artificial intelligence, 4, 100125.

TS, A., & Guddeti, R. M. R. (2020). Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks. Education and information technologies, 25(2), 1387-1415.

Ashwin, T. S., & Guddeti, R. M. R. (2020). Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework. User Modeling and User-Adapted Interaction, 30(5), 759-801.

Niyozov, N., Saburov, S., Ganiyev, S., & Olimov, S. (2023). AI-powered learning: revolutionizing technical higher education institutions through advanced power supply fundamentals. In E3S Web of Conferences (Vol. 461, p. 01092). EDP Sciences.

Yadannavar, S., Ashwin, T. S., Pathan, R., Vanahalli, M. K., & Rajendran, R. (2024, December). Multimodal Analysis of Learning-Centered Emotions and Cognitive Processes in Open-Ended Learning Environments. In 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) (pp. 1-8). IEEE.

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International journal of artificial intelligence in education*, *26*(2), 600-614.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, *46*(4), 197-221.

Google. (2025). *Gemini 2.5 Pro architecture* [Generative AI model]. Retrieved August 19, 2025, from https://cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-pro