

Intelligent Tutoring Systems Using Artificial Intelligence for English Language Instruction

Ni Nyoman Parmithi^{1*} and Ni Putu Sri Ratna Dewi²

^{1*}Lecturer, Faculty of Science and Technology, PGRI Mahadewa University of Indonesia, Denpasar, Indonesia. E-mail: parmithi15@gmail.com, Orcid: <https://orcid.org/0009-0004-8407-3203>

²Associate Professor, Department of Biology and Marine Fisheries, Ganesha University of Education, Singaraja, Bali, Indonesia. E-mail: ratna.dewi@undiksha.ac.id, Orcid: <https://orcid.org/0009-0008-0039-291X>

Abstract: In the field of second or foreign language learning, Intelligent Tutoring Systems (ITS) based on Artificial Intelligence (AI) have been revolutionizing the approach to teaching and learning English by providing adaptive support, personalized instruction, and immediate feedback to learners. In the realm of second or foreign language learning, Intelligent Tutoring Systems (ITS) fueled by Artificial Intelligence (AI) are transforming the way educators teach and students learn English, by adapting instruction to align with the unique needs of each learner and providing immediate feedback. This paper presents the design, implementation, and classroom evaluation of an Intelligent Tutoring System based on Artificial Intelligence for English teaching that focuses on correcting grammar, learning vocabulary, and providing pronunciation feedback. A mixed-method design using prototype development and pilot classroom evaluation was used, with forty-five intermediate-level English learners engaging in the proposed system over six weeks, and the conventional system used for the comparison group. The data collected included pre- and post-test scores, questionnaires, and short interviews, as well as the system usage logs, and were analyzed using descriptive statistics and thematic coding of the learners' responses. The results showed significantly greater improvements in grammatical accuracy, vocabulary recall, and self-reported motivation for students using the AI tutoring system compared to the control group. The incorporation of NLP elements—such as automatic error detection, learner modeling, and generation of adaptive feedback—significantly shortened the delay between learner errors and corrective feedback, maintained learner involvement, and maintained the delivery of feedback. The study also highlights key design factors in determining the effectiveness of Intelligent Tutoring Systems in language learning: learner modeling, pedagogical strategy selection, and affect-aware feedback. Issues with data privacy, teacher preparedness, algorithmic bias and the ability of AI models to adapt to different linguistic contexts are also explored. The paper points out that AI-based tutoring systems have the potential to offer significant benefits for personalized English language learning at scale, but also emphasizes the importance of retaining human pedagogical control, continued assessment, and system design that is culturally responsive in the future.

Key Words: intelligent tutoring systems, artificial intelligence, english language instruction, natural language processing, adaptive learning, personalized feedback, learner modeling

(Received: 08 December 2023; Revised: 17 January 2024; Accepted: 26 February 2024; Published: 29 March 2024)

Introduction

English remains the international language for learning, trade and digital communication, creating a need for teaching methods that can cater to students of diverse linguistic backgrounds and proficiency

levels, learning speeds and styles, and learning preferences (Schmidt & Strasser, 2022). Although the conventional classroom model of English teaching has its own merits and advantages for social communication and formal courses, it often suffers from a large classroom size, few personalized feedback opportunities, and one-size-fits-all teaching intensity, which can affect the learning rate of learners with different levels (Tafazoli et al., 2019). Artificial Intelligence (AI) has been identified as a potentially powerful solution to these limitations by making it possible to build systems that are able to monitor student actions, identify learning problems, and provide personalized learning material while the student is still engaged in learning (Fitria, 2021). Intelligent Tutoring Systems (ITS) are one of the most fully developed uses of AI in education; employ models of the learner, the subject domain, and pedagogy to provide dynamically adaptive instruction for individual students (Alfaro et al., 2020). The first research in ITS has been more or less oriented towards the structured domains, such as mathematics and programming, where correctness could be assessed with a relatively ease; the extension of ITS principles to language learning is more complex, in that language competence not only relies on a correctness of a rule-based type, but also on fluency and contextual appropriateness, as well as communicative competence (Haridas et al., 2020). Despite this, the developments of natural language processing (NLP), speech recognition, and machine learning are increasingly making it possible to create platforms for ITS that can analyze learners' text and speech, identify a large variety of error types, and provide pedagogically suitable feedback. Although Intelligent Tutoring Systems (ITSs) are increasingly popular in the field of language learning, there is still inconsistent empirical evidence on effectiveness in actual language instruction within the EFL classroom (Eryılmaz & Adabashi, 2020). Existing systems have been tested mainly by technical measures (accuracy in error detection) instead of learning outcomes (accuracy in grammar, retention of learned words, learner motivation, etc.) (Shin, 2018). Moreover, existing systematic reviews show a wide range of differences in the modeling of learner knowledge and the adaptation of content in the ITS, and very few direct comparisons of AI-tutored learners with those who are conventionally instructed under similar conditions (Xu et al., 2019). While there is a significant amount of research into the predictive and diagnostic strengths of ITS in other contexts, such as those involving instruction in other languages, there is currently less research specifically focused on scaffolding techniques and the timing of dynamic feedback in the context of English language instruction (Mousavinasab et al., 2021; Anwar et al., 2022).

This paper presents a conceptual architecture for an AI-based Intelligent Tutoring System (ITS) for English Language Learning (ELL) instruction in the form of modules that combine Learner Modeling (LM), Domain Knowledge Representation (DKR), Pedagogical Strategy Selection (PSS), and NLP-based feedback Generation (FG). Based on the empirical evidence obtained from the classroom pilot conducted for six weeks, this study compares the students who have used the learning system developed with the proposed AI-ITS with those who have received traditional learning, and analyzes the differences in grammatical accuracy, vocabulary retention, and learning motivation. Furthermore, the study reveals practical implications and implementation hurdles for institutions in adopting AI tutoring tools, such as data privacy, teacher readiness, and how to conduct affective feedback, among others. Last, the study provides structured evaluation metrics and results that can be used for future comparative studies on language tutoring systems using artificial intelligence.

The rest of this paper is structured as follows. Section II reviews the literature on Intelligent Tutoring Systems and its use in Education and in the instruction of the English language. The concept of the IT based on Artificial Intelligence is introduced in Section III, in terms of the conceptual framework and the architecture. The methodology used for the pilot evaluation is described in Section IV. The prototype system design and implementation with its feedback and assessment module based on NLP is discussed in

Section V. The evaluation results are found in Section VI, including quantitative results from the pilot study. In Section VII, the principal conclusions are summarized, and suggestions are made for further research.

Literature Review

In fact, personalization has been a key feature of all research on Intelligent Tutoring Systems to date, and is explicitly seen as the way in which AI can help to enhance learning outcomes. The use of ITS platforms has demonstrated the positive impact on the achievement of students in various subjects by tailoring the difficulty and pacing of content to the trajectory of student learning (Akyuz, 2020). For example, in the field of computer programming in low-resource settings, adaptive guidance that provides step-by-step instruction has been challenging to design for because of infrastructure constraints (Sharma & Harkishan, 2022). The effectiveness of dialogue-based tutoring systems has likewise been shown in a remedial instructional context, where the dialogue between the system and the learner facilitates the progressive elimination of misconceptions, which is directly relevant to language instruction in which dialogic interaction plays a pivotal role in the development of language skills (Pai et al., 2021).

In addition to text-based interaction, there are some research prototypes that include physiological or behavioral signals to identify learner frustration/disengagement, so that the ITS can control the tone of the feedback or the level of the task (Alqahtani et al., 2020). Emotional responsiveness is of special relevance to language learners who are prone to anxiety in producing spoken and written output in a second language, and would indicate that emotional responsiveness is as important as content adaptivity for maintaining learner motivation.

More specifically, focusing on English language teaching, recent scholarship has focused on the potential of new and emerging technologies such as fifth generation (5G) connectivity and artificial intelligence (AI) in facilitating low-latency interaction between the learner and the cloud-based AI teaching system to facilitate more responsive and situationally relevant English language instruction (Yu & Nazir, 2021). Specifically, there are studies exploring the use of negative knowledge, or explicit models of common learner errors, which have been used to enhance the diagnostic precision of automated feedback in English language education and shown to improve learner academic performance (Demir, 2020). Although the larger body of ITS literature has focused on mathematics teaching with younger learners where systems have been developed for math tutoring at the primary level, the general principles of architecture are easily applied to teaching English language in a tutoring setting (King et al., 2021).

Implementation of ITS has also been made supplementary and in blended learning settings, and there have been reports that the use of adaptive tools can be successfully integrated into an existing instructional program, with positive outcomes reported when systems are used to reinforce and not replace classroom instruction (Phillips et al., 2020). More recent advances in pedagogical intervention generation that leverage automatic, data-driven techniques to automatically generate individualized hints and prompts show how the burden on instructors in developing differentiated materials can be alleviated using data from learner interactions (Kochmar et al., 2022). Finally, research on the use of ITSs by adult learners with poor literacy skills is informative with regard to issues of usability that can be applied directly to designing English language ITS platforms for a range of literacy skill levels (Fang et al., 2022).

As a group, these studies indicate that AI tutoring is well researched in subjects like mathematics and general reading, but not in English language learning, where more evidence-based design and evaluation is needed, which is the purpose of this study. Based on the results of this study, it can be concluded that AI-based ITS is a meaningful tool in English language teaching that embeds features of learner modeling,

pedagogically relevant feedback sequencing, and NLP-based error detection. The advantages of immediate, individual correction (which is not easily achieved by the conventional classroom environment at scale) are suggested by the greater improvements in grammatical accuracy and vocabulary retention for experimental group learners. Simultaneously, the qualitative results suggest that the anonymity and privacy afforded by AI feedback can reduce affective obstacles that prevent anxious language learners from producing language beyond just the content of the message.

The shorter time lag between system uses and feedback, as seen in system use logs, is consistent with the pedagogical theory that giving timely feedback helps learners to resolve errors, especially in procedural areas of language like grammar and pronunciation. The teachers' recollection of a period of adaptation when using the new diagnostic dashboards, however, indicates that the effectiveness of use will not only rely on the robustness of the technical implementation but also on good preparation of teachers for implementation, otherwise even a well-designed system will not be fully used. Combined, the observations suggest that AI-based ITS should be used in conjunction with high quality classroom teaching, and the best advantage seems to come when the human teacher and the adaptive AI are strategically integrated into a common instructional strategy.

Conceptual Framework For AI-Based English Language Its

3.1 Architecture of the Proposed Model

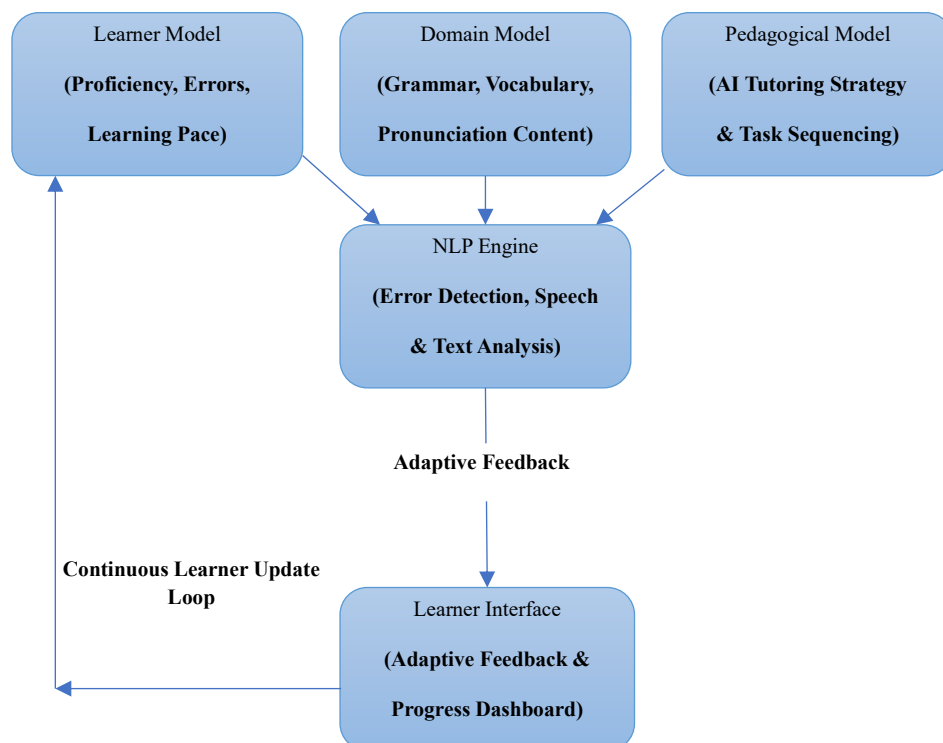


Figure 1: Architecture of the Proposed AI-Based Intelligent Tutoring System for English Language Instruction

The proposed system is based on the classical four-component ITS architecture, and tailored to the needs of the English language teaching field as shown in figure 1.

The Learner Model keeps track of the learner's level of mastery, the common error patterns, the time it takes for the learner to make the response, and preferred rate of learning, so that the system can

differentiate between a learner who has trouble with verb tense and a learner who has trouble with vocabulary range. The Domain Model represents the structured content of English learning, such as grammar rules, vocabulary lists arranged by rank and subject, pronunciation targets and more, and is extensible to accommodate new content as course curricula change.

The Pedagogical Model acts as the decision-making layer in the system, which chooses the pedagogical strategies, sequencing and timing for introducing corrective feedback, further practice and progression to new content as a result of the current state of the Learner Model. The layer builds on proven strategies used by tutors, including the use of scaffolded hinting, spaced repetition for vocabulary retention, and graduated difficulty progression. The NLP Engine connects to and produces output of all three of these parts and automatically parses and analyzes learners' texts and speech input, categorizes these errors according to semantic nature (e.g., syntactic error, morphological error, lexical error, etc.), and provides feedback in natural language based on the level of proficiency of the learner. Finally, the Learner Interface provides the student with feedback and progress information adaptively, and communicates feedback to the Learner Model, creating a feedback loop that allows the Learner Model to be continually updated as the student progresses, enabling the system to become more at ease with the learner's knowledge over time.

Methodology

The study was a mixed-method quasi-experimental prototype system development and pilot evaluation in the classroom, which enabled both quantitative measurement of the learning outcomes and qualitative insights into the learner experience.

4.1 Participants

The participants were 90 intermediate-level English learners (ages 14-16) from two concurrent sections in an English program in a secondary school. The proposed AI-based Intelligent Tutoring System was administered to learners in the experimental group ($n = 45$) while learners in the comparison group ($n = 45$) were exposed to the conventional teacher-led instruction and the regular program for the same duration of 6 weeks. Both groups were supervised by two English language teachers (L1 teachers), with over 5 years of experience teaching in the classroom, and engaged in structured interviews after the pilot.

4.2 Data Collection Instruments

Four instruments were used for data collection in order to provide triangulation. The grammar, vocabulary, and reading comprehension of both groups were evaluated with pre-tests and post-tests to determine learning gains, which were standardized. Only the experimental group was asked to complete a structured Likert scale questionnaire at the end of this study to gain insights into learner perceptions of engagement, feedback usefulness and confidence in the use of English. The experimental group's system usage logs included details of the frequency of interaction, types of errors, and time spent on corrective actions. Thirdly, a small number of semi-structured interviews were carried out with a small group of eight learners and the two teachers involved to obtain qualitative views on the tutoring experience.

4.3 Data Analysis

The quantitative data obtained from pre and post-tests and the Likert scale questionnaire data was analysed by descriptive data such as percentage gain and group comparison to characterise the overall performance trend. Thematic coding was used to identify patterns in qualitative data gathered from interview questions and open-ended questionnaire questions in relation to motivation, feedback perception

and usability problems. This integration of quantitative and qualitative analysis gave a more holistic understanding of the measurable learning outcomes and of the subjective experience of the learner.

4.4 Ethical Considerations

All necessary institutional approvals were sought before data collection, and informed consent was requested from school administrators, participating teachers and parents/guardians of student participants. Students were not required to participate, and students were able to drop out at any time without penalty. Participant codes were used to anonymize all collected data and these data were kept securely for only the purposes of this research.

System Design and Implementation

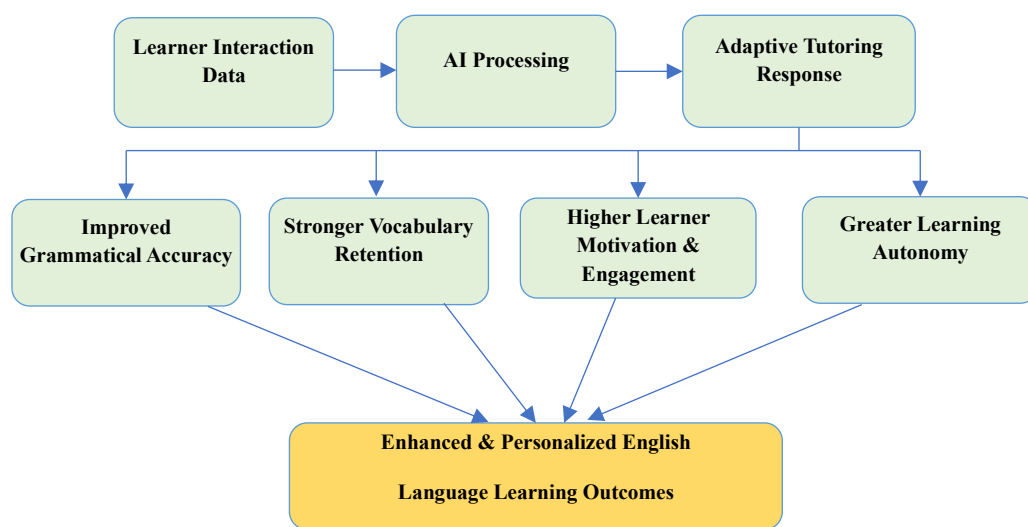


Figure 2: Conceptual Model of AI-Enabled Learning Outcomes in English Language Instruction

The conceptual model of English language learning with AI is illustrated in figure 2. AI-driven NLP and error categorization of learner interaction data such as text, speech and response time. Adaptive tutoring provides individualized feedback and grammar, vocabulary, motivation and learner autonomy improvements, which ultimately contribute to enhanced and individualized English language learning outcomes.

In the first stage, the entered text or spoken speech is broken down into words and sentences, and analyzed for deviations from the target grammar, with the use of a combination of rule-based pattern matching for recurrent error categories (such as subject-verb agreement, articles, etc.) and statistical language models for more context-dependent errors (such as collocations, awkwardness, etc.). In the second stage, errors detected are categorized according to severity and type, and are matched to a set of pre-stored feedback models that the Pedagogical Model chooses from, depending on the level of the learner and error history, so that the feedback is not too difficult for novice learners or too easy for more advanced learners.

The system also includes automatic speech recognition (ASR) to compare the pronunciation of learner's utterances with the pronunciation patterns from the target language, highlighting the phonemes that differ significantly from the expected pronunciation, and allowing the learner to play a model pronunciation and try again. Words are reinbated via a Spaced Repetition Scheduling algorithm that brings words that were secreted or responded incorrectly in earlier sessions back at greater intervals depending

upon the word's retention as shown by student responses. In all of the modules, interaction data such as accuracy of responses, response time and self-corrections are recorded and incorporated into the Learner Model outlined in Section III so that the system recommendation will be increasingly more learner-specific throughout the 6-week pilot period.

Evaluation and Results

The learning outcomes were assessed through pre-test to post-test comparisons of the experimental group with the comparison group and with the post-study questionnaire for assessment of the experimental group self-perceptions. The experimental group showed a difference of 24 percentage points for grammatical accuracy in posttest as compared to 11 percentage points in the comparison group, and a difference of 21 percentage points in vocabulary retention as compared to 9 percentage points in the comparison group. Table 1 shows that the responses to the learners' questionnaire are also overwhelmingly positive regarding the system in various aspects of the learning experience.

Table 1: Learners' Perceptions of AI-Enabled English Language Learning System

Statement	Agree (%)	Neutral (%)	Disagree (%)
The system's feedback helped me understand grammar mistakes	85	9	6
I felt more confident practicing English using the system	79	14	7
The pace of exercises matched proficiency level	74	17	9
Feedback was easier to understand than written corrections	76	15	9
I would recommend this system to other English learners	81	12	7

The results of the learners' perceptions of the AI-based English language learning system are summarized in table 1. The improvement of grammar understanding, the boost of confidence, the appropriate exercises for different levels of skills and the feedback were the skills which most participants agreed that the system helped them in. The high recommendation rate signifies a high satisfaction level from users as well as the success of tailored language instruction through AI assistance.

System usage logs revealed that learners in the experimental group received corrective feedback within an average of 18 seconds of error, much quicker than the timescales involved in traditional written corrective feedback in the comparison classroom, which typically take days. Thematic analysis of the interview data showed that students were happy to receive AI feedback in a non-judgmental, private environment, and several students reported that that did not feel as though being judged when using AI to receive feedback on mistakes in front of others. Teachers, on the other hand, stated that able to detect struggling students more quickly than would during classroom observation, as a result of the usage logs and progress dashboards the system provided. Teachers also commented on needing a period to get used to the system in order to interpret the system's diagnostic reports effectively.

Conclusion and Future Work

The purpose of this study was to investigate the application of Artificial Intelligence in the context of Intelligent Tutoring Systems (ITS) for the teaching of English and the results are promising. The pilot assessment demonstrated that students receiving the AI-based system had more improvement in grammatical accuracy, word retention, and self-reported motivation compared to students in the traditional

instruction group, with most students giving a positive rating for the ease and clarity of use of the system. Central to reducing the gap between learner errors and corrective feedback were the NLP aspects of the system, especially the automatic error detection, and the adaptive generation of feedback. The system's learner modeling component enabled instructional pacing to be adapted to the proficiency level of each learner. Meanwhile, the study raised some questions that should be addressed if AI is to be used in the future: How will learner data be handled transparently? Will teachers receive adequate training to use AI in teaching practice? How will the feedback be calibrated so it is encouraging rather than discouraging for lower-proficiency learners? Moreover, the evidence indicates that AI tutoring is best when it serves as a tool to complement rather than supplant teacher-led instruction, as teachers are still needed to provide a breadth of support that goes beyond the limitations of current AI models, such as sociocultural nuances and motivation. Future research should continue and build on this work by using longitudinal studies that capture learning outcomes over a longer period of time, larger and linguistically more diverse samples, and comparative studies of various AI architectures and feedback methods. In addition, implementation of speech AI tutors for pronunciation instruction, teachers' learning analytics dashboard, and ethical standards for the application of AI tutors in the classroom should be emphasized. The findings of this study support the general opinion that, when carefully planned and well managed, Intelligent Tutoring Systems can be effectively used to improve the personalization, efficiency and inclusiveness of English language learning in a variety of educational contexts.

References

- Akyuz, Y. (2020). Effects of intelligent tutoring systems (ITS) on personalized learning (PL). *Creative Education*, 11(6), 953-978. <https://doi.org/10.4236/ce.2020.116069>
- Alfaro, L., Rivera, C., Castaneda, E., Zuniga-Cueva, J., Rivera-Chavez, M., & Fialho, F. (2020). A review of intelligent tutorial systems in computer and web-based education. *International Journal of Advanced Computer Science and Applications*, 11(2), 755-763. <https://doi.org/10.14569/ijacsa.2020.0110295>
- Alqahtani, F., Katsigiannis, S., & Ramzan, N. (2020). Using wearable physiological sensors for affect-aware intelligent tutoring systems. *IEEE Sensors Journal*, 21(3), 3366-3378. <https://doi.org/10.1109/jsen.2020.3023886>
- Anwar, A., Haq, I. U., Mian, I. A., Shah, F., Alroobaea, R., Hussain, S., ... & Umar, F. (2022). Applying Real-Time Dynamic Scaffolding Techniques during Tutoring Sessions Using Intelligent Tutoring Systems. *Mobile Information Systems*, 2022(1), 6006467. <https://doi.org/10.1155/2022/6006467>
- Demir, Ü. (2020). The effect of using negative knowledge based intelligent tutoring system evaluator software to the academic success in English language education. *Pedagogies: An International Journal*, 15(4), 245-261. <https://doi.org/10.1080/1554480x.2019.1706522>
- Eryılmaz, M., & Adabashi, A. (2020). Development of an intelligent tutoring system using bayesian networks and fuzzy logic for a higher student academic performance. *Applied Sciences*, 10(19), 6638. <https://doi.org/10.3390/app10196638>
- Fang, Y., Lippert, A., Cai, Z., Chen, S., Frijters, J. C., Greenberg, D., & Graesser, A. C. (2022). Patterns of adults with low literacy skills interacting with an intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 32(2), 297-322. <https://doi.org/10.1007/s40593-021-00266-y>

- Fitria, T. N. (2021). The use technology based on artificial intelligence in English teaching and learning. *ELT Echo: The Journal of English Language Teaching in Foreign Language Context*, 6(2), 213-223. <https://doi.org/10.24235/eltecho.v6i2.9299>
- Haridas, M., Gutjahr, G., Raman, R., Ramaraju, R., & Nedungadi, P. (2020). Predicting school performance and early risk of failure from an intelligent tutoring system. *Education and Information Technologies*, 25(5), 3995-4013. <https://doi.org/10.1007/s10639-020-10144-0>
- King, C. L., Vincent, Kelvin, Warnars, H. L., Nordin, N., & Utomo, W. H. (2021). Intelligent Tutoring System: Learning Math for 6th-Grade Primary School Students. *Education Research International*, 2021(1), 5590470. <https://doi.org/10.1155/2021/5590470>
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2022). Automated data-driven generation of personalized pedagogical interventions in intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 32(2), 323-349. <https://doi.org/10.1007/s40593-021-00267-x>
- Mousavinasab, E., Zarifsanaiy, 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. <https://doi.org/10.1080/10494820.2018.1558257>
- Pai, K. C., Kuo, B. C., Liao, C. H., & Liu, Y. M. (2021). An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. *Educational Psychology*, 41(2), 137-152. <https://doi.org/10.1080/01443410.2020.1731427>
- Phillips, A., Pane, J. F., Reumann-Moore, R., & Shenbanjo, O. (2020). Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development*, 68(3), 1409-1437. <https://doi.org/10.1007/s11423-020-09745-w>
- Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. *Anglistik: International Journal of English Studies*, 33(1), 165-184. <https://doi.org/10.33675/angl/2022/1/14>
- Sharma, P., & Harkishan, M. (2022). Designing an intelligent tutoring system for computer programing in the Pacific. *Education and Information Technologies*, 27(5), 6197-6209. <https://doi.org/10.1007/s10639-021-10882-9>
- Shin, M. H. (2018). How to use artificial intelligence in the English language learning classroom. *Indian Journal of Public Health Research & Development*, 9(9), 557-561. <http://dx.doi.org/10.5958/0976-5506.2018.01058.6>
- Tafazoli, D., María, E. G., & Abril, C. A. H. (2019). Intelligent language tutoring system: Integrating intelligent computer-assisted language learning into language education. *International Journal of Information and Communication Technology Education (IJICTE)*, 15(3), 60-74. <https://doi.org/10.4018/ijicte.2019070105>
- 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>
- Yu, H., & Nazir, S. (2021). Role of 5G and artificial intelligence for research and transformation of English situational teaching in higher studies. *Mobile Information Systems*, 2021(1), 3773414. <https://doi.org/10.1155/2021/3773414>