International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Improving the Effectiveness of Chatbots by Incorporating Self-Assessed User Knowledge into the Question-Answering Process

Yuqi Wang and John Leddo

MyEdMaster, LLC

Virginia, USA

DOI: 10.46609/IJSSER.2025.v10i08.035 URL: https://doi.org/10.46609/IJSSER.2025.v10i08.035

Received: 12 August 2025 / Accepted: 28 August 2025 / Published: 5 September 2025

ABSTRACT

Chatbots, personal assistants, and large language models (LLMs) have become pervasive in our world. A major function they serve is to provide information and answer questions. Collectively, these technologies typically have two major weaknesses: they provide general answers to questions they are asked without regard to the specific knowledge needs of the user and they do not assess whether the user actually understood the answers or information provided. The present paper addresses the first weakness by creating a self-assessment chatbot that helps a user to self-assess what s/he knows about a topic and then uses that knowledge when answering the user's questions with an eye toward filling in the knowledge gaps. This self-assessment chatbot was presented to college students learning math and compared to Chat GPT, which simply answers questions without such self-assessment information. Results showed that students using a self-assessment chatbot scored 10 points or the equivalent of a full letter grade higher on a posttest than those using the standard Chat GPT.

Introduction

Across more than forty years of studies, researchers have shown that individualized instruction can improve learning outcomes far more effectively than whole-class teaching. Bloom's well-known "2-sigma" finding showed that one-to-one tutoring can raise student performance by as much as two standard deviations (Bloom, 1984). Based on this foundation, research on intelligent tutoring systems (ITS) shows that well-designed computer tutors can closely approximate the effectiveness of human tutoring. Van Lehn (2011) showed that such systems work by modeling learners' cognitive states and adjusting both feedback and problem selection. For example, cognitive tutor frameworks use domain modeling and step-level feedback. According to Koedinger and Corbett (2006), such designs contribute to gains in both procedural

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

fluency and conceptual understanding. More recently, the advent of large language models (LLMs) has renewed interest in natural language–based tutoring tools. These tools are valued for their conversational flexibility. Reviews, however, emphasize that their educational impact depends on alignment with learner needs, along with transparency and accuracy (Kasneci et al., 2023). Building on this concern, our study tests whether a self-assessment-informed conversational agent can outperform a general-purpose LLM in supporting undergraduate mathematics learning.

Adaptive learning systems—particularly ITS—have demonstrated consistent effectiveness in improving learner outcomes by tailoring instruction to individual cognitive states (Létourneau, 2025). Létourneau's (2025) systematic review of AI-driven ITS across K–12 schools found mostly positive results. Students using ITS showed stronger learning gains than those in non-intelligent environments, though the extent of improvement depended on design and duration. Central to these systems are mechanisms such as model tracing and knowledge tracing. These techniques allow real-time monitoring of students' problem-solving steps and estimation of skill mastery, which in turn support dynamic task selection and just-in-time feedback (Koedinger & Corbett, 2006). Moreover, modern AI-based ITS structures often incorporate natural language processing modules and real-time assessment pipelines. Through these components, the system can adjust how it delivers content in response to students' ongoing performance (Villegas-Ch et al., 2025).

Researchers are now exploring how LLMs, including GPT-4, might be applied in adaptive learning. Early findings suggest that such models could act as tutoring systems that are both flexible and sensitive to context. Although LLMs have demonstrated strong capabilities in generating explanations and scaffolding problem-solving (Kasneci et al., 2023), concerns remain about their tendency to produce overly general responses when not anchored in student-specific data. Emerging work in personalized educational agents suggests that coupling LLMs with diagnostic or self-assessment modules may provide a pathway toward more learner-sensitive feedback (Zawacki-Richter et al., 2019). Nevertheless, rigorous controlled experiments validating the effectiveness of such hybrid systems in real classroom settings remain scarce. This gap emphasizes the need for empirical studies to test the effectiveness of personalization mechanisms built on LLM infrastructure. The present experiment addresses this need. It tests whether such mechanisms improve learning outcomes more effectively than conventional chatbot interactions.

One way to achieve personalization is to adjust instruction to what the learner already knows. Indeed, the traditional ITS model contains a student model for that very purpose (Greer, 1995; Brna, Ohlsson and Pain, ref). The lack of a student model represents a fundamental weakness in mainstream LLMs, which are geared towards answering questions without regard to who is

ISSN: 2455-8834
Volume:10, Issue:08 "August 2025"

asking them. This makes sense since LLMs are, by their very nature, language models not teaching models. Therefore, they are not constructed to strategically assess what knowledge learners have and what they are missing, so that these gaps can be used in the process of generating answers.

Once solution is to create an independent assessment system and link it to an LLM. This is labor intensive. Another solution is to allow a learner to enter his or her own existing subject matter knowledge into LLM and have the LLM use that information when answering a learner's questions. Given that learners may not be skilled in assessing their own knowledge, a self-assessment LLM-based chatbot needs a reliable and easy to use self-assessment method.

Our previous work has been devoted to developing such a method. Given that the goal of the proposed self-assessment chatbot is to fill in knowledge gaps, traditional assessment methods that focus on whether users can correctly answer questions are inadequate since these methods do not diagnose knowledge but performance. The assessment method used in the present project is called Cognitive Structure Analysis (Leddo et al., 1990).

Cognitive Structure Analysis or CSA It is based on decades of cognitive psychology research that have illustrated that people possess various knowledge types, each of which is organized and used differently in problem-solving. Since people possess different types of knowledge, our framework integrates several prominent and well-researched formalisms. These include semantic nets, which organize factual information (Quillian, 1966); production rules, which organize concrete procedures (Newell and Simon, 1972); scripts, which are general goal-based problem-solving strategies (Schank and Abelson, 1977; Schank, 1982); and mental models, which explain the causal principle behind concepts (de Kleer and Brown, 1981). Because our framework integrates these four knowledge types, it is called INKS for the INtegrated Knowledge Structure.

The INKS framework is based on research by John Leddo (Leddo et al., 1990) which showed that true mastery of a topic or subject requires all four knowledge types. The framework also brings helpful implications for instruction. For example, in John Anderson's ACT-R framework, people initially learn factual/semantic knowledge that is later operationalized into procedures (Anderson, 1982). Research by Leddo takes this one step further showing that expert knowledge is organized around goals and plans (referred to in the literature as "scripts" – Schank and Abelson, 1977; Schank, 1982) and abstracted into causal principles (referred to in the literature as "mental models" – cf., de Kleer and Brown, 1981) that allow people to construct explanations and make predictions/innovations in novel situations.

To identify the root cause of the mistake, the query-based assessment framework, CSA, incorporates principles from the INKS knowledge representation framework. CSA is chosen

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

because previous research describes a strong correlation between user knowledge — as assessed by CSA — and performance practical problem-solving. In one previous research project, we found that using an automated multiple-choice CSA system to access student learning produced measures of knowledge that correlated .88 with student problem-solving performance and measures of change of knowledge as a result of the instruction that correlated .78 with change in performance from pretest to post test. Moreover, at risk students who had their learning needs diagnosed using CSA performed at a mainstream level three grades higher than their own after a 25-hour tutoring program in science (Leddo and Sak, 1994). Leddo et al. (2022) extended these findings. Students were given open ended questions to assess their factual (semantic), strategic (script-based), procedural, and rational (mental model) concept, knowledge of Algebra 1. The total INKS knowledge and individual component knowledge scores were correlated with the total number of correctly solved problems. Results showed correlations of .966 between problem-solving and total knowledge, .819 between problem-solving and strategic knowledge, .866 between problem-solving and factual knowledge, .937 between problem-solving and procedural knowledge and .788 problem-solving and rational knowledge. These findings were extended to pre-calculus (Zhou and Leddo, 2023), biology (Ahmad and Leddo, 2023), and elementary school math (Bekkari and Leddo, 2023). In two other projects, assessments produced using the CSA methodology produced assessments of students. Learning agreed with teachers' assessments, approximately 95% - 97% of the time which was statistically equal to teachers' assessments with each other (Leddo et al., 1998, Liang and Leddo, 2020).

Our previous work in CSA shows that CSA can be a powerful tool in helping educators assess what students do and do not know. CSA has been presented as an alternative to the classical test theory approach of measuring learning as a function of the number of correct answers students give. However, it could be reasonably argued that the purpose of education is to improve student performance, and, therefore, replacing an assessment system with one that directly measures underlying knowledge but does not raise student performance would be less appropriate. Leddo and Ahmad (2024) addressed that issue directly. In that study, high school and college students were initially assessed in their knowledge of logarithms. Half were assessed using CSA and the other half were assessed by asking them to solve problems and show all work. After each problem, students received remediation on either their knowledge concepts (in the CSA condition) or in their problem-solving steps (the "show all work" condition). Results showed that remediating problem-solving steps raised student performance from an average of 68% on the pretest to 75% on the post-test, a statistically significant increase. However, those who had their knowledge assessed and remediated scored 85% on the post-test, a statistically significant, full-letter grade higher performance than those in the "show all work" condition. The Leddo and Ahmad (2024) was replicated in a follow-up study with middle schoolers that also showed that students who were assessed using CSA and had their knowledge remediated performed, on

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

average, a full letter grade higher than those whose step-by-step procedures were assessed and remediated (Challagulla and Leddo, in press).

Showing that assessing and remediating INKS-based knowledge improves performance addresses only half the issue. The previously-cited research involved learners being assessed using external means. For a self-assessment chatbot to work, the question remains whether learners can be taught to reliably assess their own knowledge and, equally importantly, whether learning to self-assess can be done quickly and easily so as to be practical to implement.

It turns out the answers to these questions is yes (Cynkin and Leddo, 2023; Dandemraju, Dandemraju and Leddo, 2024). In these two studies, we showed that learners can be trained to accurately assess what they do and do not know and that this process takes about 10 minutes. To train a person to self-assess, s/he is shown a sample of what a self-assessment for a topic area looks like. The learner is then asked to use the sample as a model for generating a self-assessment for a new topic. A template is provided for filling in the factual (semantic), strategic (scripted-based), procedural (production rule) and rational (mental model) knowledge.

To ensure that remediation of self-assessed knowledge also leads to improvement in performance, we have also taken the next logical step in that area to see if students can not only assess their knowledge gaps but also then remediate these gaps. It turns out that students can do so very successfully. To address this issue, Ravi and Leddo (2024) conducted a study in which high school students learned an advanced topic in chemistry by watching a video. Half the students were told to rewatch the video to fill in any knowledge gaps, while the other half were taught to self-assess their knowledge using CSA and then told to rewatch the video to fill in any assessed knowledge gaps. The group that was taught to self-assess scored 15 points or 1.5 letter grades higher on a post-test than students who simply rewatched the video without self-assessment. Nehra and Leddo (2024) replicated the Ravi and Leddo study to the learning of Spanish. They found that high school students performing self-assessment plus remediation scored, on average, 25 percentage points or 2.5 letter grades higher than those re-reading the material without performing a self-assessment. Prakash and Leddo (2025a) extended the Ravi and Leddo (2024) and Nehra and Leddo (2024) findings to another subject area: high school reading comprehension. The results revealed a mean post-test score of 8.3 out of 12 (69.17%) for the control group and 11.2 out of 12 (93.33%) for the experimental group. This difference in averages was statistically significant ($t = 3.75$, $df = 11.07$, $p < .01$). Notably, individual scores further illustrated the disparity: the lowest score in the control group was 41.67%, whereas the lowest in the experimental group was 83.33%. This is the difference between an F letter grade and B letter grade. Following this, another study conducted by Prakash and Leddo (2025b) examined CSA's effectiveness in teaching math, specifically, the topic of Bayes' Theorem, and found a 27-point improvement. Statistical analysis yielded a t-value of 4.38 ($df = 18$, $p =$

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

0.0004), confirming the significance of the difference. Individual scores also highlighted the disparity. The control group's lowest score was 6/20 (30%), whereas the experimental group's lowest score was 15/20 (75%). Following this, a history assessment revealed that students who utilized CSA for self-assessment and remediation significantly outperformed their peers in the control group (Prakash and Leddo, 2025c). Post-test results demonstrated that the experimental group achieved an average score of 87.5%, whereas the control group scored 65.8%, indicating a substantial difference in comprehension and retention of historical concepts.

These results on high school students were further extended by Leddo, Clark and Clark (2025) in their investigation of middle school math. Leddo, Clark and Clark found that middle school students who self-assessed using CSA and then remediated their knowledge gaps scored 18 percentage points higher on a posttest than those who relearned material without first performing a self-assessment.

Following this, Prakash and Leddo (2025d) conducted a study on middle school students' reading comprehension, specifically through an analysis of To Kill a Mockingbird, a novel that explores complex themes of ethics and social structure. Students in the experimental group were trained to evaluate their own knowledge gaps and use targeted remediation strategies, while those in the control group engaged with the text without structured self-assessment. Results showed that students in the self-assessment group scored 16 points higher on a post-test than those who re-read the material without self-assessment. This was followed up with a study on middle school science (Prakash and Leddo, 2025e), in which students learned about topics in ecology. Results showed that students who used the self-assessment technique plus remediation scored on average 98% on a post-test, while those who simply reread the material without self-assessment scored on average 77.5%.

Finally, Sathiyamoorthy and Leddo (2025) showed that college students who used CSA to self-assess and then remediate knowledge performed 13 percentage points higher on a college psychology post-test than those who simply reread the material after initially learning it. Taken together, these results suggest that regardless of whether the students self-assess and remediate knowledge or the assessment and remediation is mediated by technology, assessing and remediating knowledge greatly improves student performance compared to traditional methods of assessment. This indicates that student achievement could be increased systemically and cheaply by introducing CSA-based knowledge assessment into educational practices.

The present project seeks to incorporate CSA-based self-assessment into a chatbot, so that the chatbot can use the results of that assessment when answering learner questions. This will be compared to a traditional chatbot/LLM that simply answers questions without knowing the learner's knowledge needs.

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Methodology

Participants

Participants were 24 undergraduate students from Boston College. They were from diverse academic backgrounds, including STEM, Education, and Liberal Arts majors. All participants were proficient in English and fluent in English reading and writing. Among them, 13 were male and 11 were female. Participants were randomly allocated to one of two groups: the control group (n = 12) or the experimental group (n = 12).

Self-assessment Chatbot Technology

In this study, we developed a lightweight AI-powered personal agent designed to provide adaptive guidance based on students' inputs. The agent is built upon the GPT-4.1 large language model which incorporates a reasoning framework that supports dynamic prompt chaining and reflection processing. Meanwhile, through integrating a self-assessment step, the agent evaluates and captures the student's current understanding level of the subject before initiating personalized tutoring conversations. This pre-processing step allows the agent to generate instructional responses that are more aligned with each student's cognitive level. Although the infrastructure utilizes the common LLM API, the customized personal agent built on it can be more adaptable to the individual level of students than the common chatbot.

The self-assessment data was collected through a self-assessment form for math that was used in the previous work cited in the Introduction and that can be accessed for free from the link www.myedmaster.com/ways-to-improve-learning/.

Procedure

The goal of the experiment was to evaluate the effectiveness of a personalized tutoring agent compared to the general-purpose LLMs (Large Language Model) chatbot in helping academy of students. In this experiment, the composite function of math was selected as the experiment subject.

The experiment consisted of four phases, including the pre-test phase, instruction phase, interaction phase, and post-test phase.

In the pre-test phase, both groups first completed a 10-question multiple-choice pre-test using Google Forms. The test measured baseline knowledge of the composite function. All participants were selected from the experiment who scored below 50% accuracy to ensure minimal prior knowledge and a comparable starting point.

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

In the instruction phase, researchers provided the same instructions of the composite function for both groups via Zoom. The instructions covered key concepts of the composite function, including composite function notation, graph transformations, domain and range analysis, and effects of inner and outer function changes.

Then, in the interaction phase, each group conducted different manipulations. For the control group, participants interacted with general LLMs (e.g., ChatGPT, Gemini) to ask questions about the experiment topic. For the experiment group, participants first completed the self-assessment form about the composite function within the application. Then, participants submitted the self-assessment to the personalized tutoring agent, which adapted each participant's current ability. Finally, the participants interacted with the personalized tutoring agent in the same way as the experiment group did.

Finally, in the post-test phase, both groups finished the same 10-question multiple-choice post-test, which was more challenging than the pre-test of composite function to test the result after interacting with different LLMs.

Results

All responses from the pre- and post-tests were collected via Google Forms. Each participant's responses were graded, and the number of correct answers was recorded. Results showed that Participants using the standard chatbot scored, on average, 69.17% on the post-test, while those using the self-assessment chatbot scored, on average, 80%. This difference was statistically significant, $t(22) = 2.31$, $p = .03$.

Discussion

The results showed that Participants who used the self-assessment chatbot, scored on average, the equivalent of a full letter grade higher than those using a chatbot/LLM by itself. This suggests that providing chatbot with learners' knowledge needs and directing them to use those needs when answering questions (as opposed to giving general answers) can improve learning. Moreover, implementing the knowledge assessment and transfer to the chatbot through the self-assessment paradigm allows this enhancement to be implemented in a way that involves minimal updating to chatbots (and without changing their fundamental makeup) and minimal intrusion on the user. This provides the best of both worlds. Of course, this study examined just one subject (math) and one population (college students). Additional research should be conducted with other topic areas, both academic and non-academic, and other user groups.

Given that this one intervention can dramatically improve chatbot effectiveness, the question arises as to whether there are other interventions that can also improve learning. (We use the term

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

learning because it seems that chatbots focus on delivery of information, but information is useless until it is learned.) Variables that could be examined include learner variables (e.g., learning style, age), characteristics of the material (how advanced it is, is it abstract or sensory based), type of question (is it asking for facts, procedures, reasons?) and types of answers (informational, analogical). In a previous study, we found that how one answers a question can double how well the questioner learns the material (Leddo et al., 2021).

A final area of investigation is one that appears to be universally neglected by chatbots and even search engines. Chatbots, LLMs and search engines dutifully respond to queries by providing relevant (in most cases) information. However, information is useless until it is learned. Current chatbots, LLMs and search engines do not check to see if the recipient of the information understood the information/answer that was delivered. This may not be as easy as it seems. While humans frequently as each other "Did you understand what I just said?", research by Leddo, Clark and Clark (2021) suggests that people are not always accurate in knowing whether they did or did not understand something. In their study, both middle schoolers and adults were given an algebra lesson and then asked if they understood what they were taught. They were then given problems to solve based on the taught topic. Both middle schoolers and adults missed a third of the problems that tested what they said they understood. Interestingly, while adults proved relatively accurate in determining when they did not understand something, correctly answering only 10% of the problems that tested what they said they did not understand, middle schoolers actually correctly answered a third of the problems that tested what they said they did not understand. In cases such as these, (self)CSA might serve as a means to measure how much people understand the answers given by chatbots.

Even then, checking for understanding will not matter unless the chatbot can adjust how it answer questions to improve that understanding. A solution to this may be to create a feedback loop in which answers to questions are given, users are assessed for understanding and then feedback from the assessment is used in a machine learning program to update the effectiveness of types of answers to types of questions for types of users.

References

Ahmad, M. & Leddo, J. (2023). The Effectiveness of Cognitive Structure Analysis in Assessing Students' Knowledge of the Scientific Method. International Journal of Social Science and Economic Research, 8(8), 2397-2410

Anderson, J.R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-405.

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Bekkari, V., & Leddo, J., (2023). Cognitive Structure Analysis: Assessing Elementary School Students in Math to Determine the Types of Knowledge They Have. International Journal of Social Science and Economic Research, 8(10)

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, *13*(6), 4-16.

Brna, P., Ohlsson, S., & Pain, H. (1993) (Eds.). Proceeding of Artificial Intelligent in Education '93. Charlottesville, VA: Association for the Advancement of Computing in Education.

Challagulla, M & Leddo, J. (2025). Assessing and Remediating Knowledge Improves Problem-solving Performance in Middle School Math Compared to Remediating Problem-solving Steps. International Journal of Social Science and Economic Research, in press.

Cynkin, C. and Leddo, J. (2023). Teaching Students to Self-Assess Using Cognitive Structure Analysis: Helping Students Determine What They Do and Do Not Know. International Journal of Social Science and Economic Research, 8(9), 3009-3020.

Dandemraju, A., Dandemraju, R. and Leddo, J. (2024). Teaching students to self-assess their own chemistry knowledge. International Journal of Social Science and Economic Research, 9(2), 541-549.

De Kleer, J. and Brown, J.S. (1981). Mental models of physical mechanisms and their acquisition. In J.R. Anderson (Ed.), Cognitive Skills and their acquisition. Hillsdale, NJ: Erlbaum

Greer, J. (1995) (Ed.) Proceeding of Artificial Intelligent in Education '95. Charlottesville, VA: Association for the Advancement of Computing in Education.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Koedinger, K. R., & Corbett, A. (2006). *Cognitive tutors: Technology bringing learning sciences to the classroom* (pp. 61-77). Na.

Leddo, J., Chen, T., Menachery, A., Agarwal, J. and Agarwal, T. (2021). Towards Improving Personal Assistants and Educational Software: How Questions Are Answered Affects Learning. International Journal of Social Science and Economic Research, 6(2), 696-705.

Leddo, J. Clark, A. and Clark, E. (2021). Self-Assessment of Understanding: We Don't Always Know What We Know. International Journal of Social Science and Economic Research, 6(6), 1717-1725.

Leddo, J. Clark, E. and Clark, A. (2025). Using self-assessment and remediation to raise middle school student achievement in math. International Journal of Social Science and Economic Research, 10(3), 1083-1092.

Leddo, J., Cohen, M.S., O'Connor, M.F., Bresnick, T.A., and Marvin, F.F. (1990). Integrated knowledge elicitation and representation framework (Technical Report 90-3). Reston, VA: Decision Science Consortium, Inc..

Leddo, J., Li, S. & Zhang, Y. (2022). Cognitive Structure Analysis: A technique for assessing what students know, not just how they perform. International Journal of Social Science and Economic Research, 7(11), 3716-3726.

Leddo, J. and Sak, S. (1994). Knowledge Assessment: Diagnosing what students really know. Presented at Society for Technology and Teacher Education.

Leddo, J., Zhang, Z. and Pokorny, R. (1998).Automated Performance Assessment Tools. Proceedings of the Interservice/Industry Training Systems and Education Conference. Arlington, VA: National Training Systems Association.

Liang, I. and Leddo, J. (2020). An intelligent tutoring system-style assessment software that diagnoses the underlying causes of students' mathematical mistakes. International Journal of Advanced Educational Research, 5(5), 26-30.

Letourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*, *10*(1), 29.

Nehra, P., & Leddo, J. (2024). The effects of Cognitive Structure Analysis in self-assessing and remediating knowledge gaps in introductory Spanish. Journal of Educational Psychology, 45(3), 78-89.

Newell, A. and Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice

Prakash, P. & Leddo, J. (2025a). Using Self-Assessment and Remediation to Raise Student Achievement in Reading Comprehension. International Journal of Social Science and Economic Research, 10(1), 277-286.

Prakash, P. & Leddo, J. (2025b). Using Self-Assessment and Remediation to Raise Student Achievement in Mathematics. International Journal of Social Science and Economic Research, 10(1), 447-456.

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Prakash, P. & Leddo, J. (2025c). Using Self-Assessment and Remediation to Raise Student Achievement in History. International Journal of Social Science and Economic Research, 10(3), 650-659.

Prakash, P. & Leddo, J. (2025d). Using Self-Assessment and Remediation to Raise Middle School Student Achievement in Reading Comprehension. International Journal of Social Science and Economic Research, 10(3), 1130-1140.

Prakash, P. & Leddo, J. (2025e). Using Self-Assessment and Remediation to Raise Middle School Student Achievement in Science. International Journal of Social Science and Economic Research, 10(4), 1471-1482.

Quillian, M.R. (1966). Semantic memory. Cambridge, MA: Bolt, Beranek and Newman.

Sathiyamoorthy, S.S. and Leddo, J. (2025). Using Self-Assessment and Remediation to Raise College Student Achievement in Psychology. International Journal of Social Science and Economic Research, 10(5), 1859-1869.

Schank, R.C. (1982). Dynamic Memory: A theory of learning in computers and people. New York: Cambridge University Press.

Schank, R.C. and Abelson, R.P. (1977). Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Erlbaum.

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

Villegas-Ch, W., Buenano-Fernandez, D., Navarro, A. M., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments*, *12*(1), 1-31.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International journal of educational technology in higher education*, *16*(1), 1-27.

Zhou, L.N. & Leddo, J. (2023). Cognitive Structure Analysis: Assessing Students' Knowledge of Precalculus. International Journal of Social Science and Economic Research, 8(9), 2826-2836.

