

Assessing Knowledge Rather Than Correct Answers: A Cognitive Structure Analytics Framework for Personalized Learning

Sahasra Ravirala¹ Matteen Safi¹ and John Leddo^{1,2,3}

¹MyEdMaster, LLC, Virginia, USA

²METY Technology, Inc., Virginia, USA

³METY Foundation, Virginia, USA

DOI: 10.46609/IJSSER.2026.v11i09.056 URL: <https://doi.org/10.46609/IJSSER.2026.v11i09.056>

Received: 30 August 2026 / Accepted: 18 September 2026 / Published: 30 September 2026

ABSTRACT

Most educational assessment systems summarize learning as the proportion of questions answered correctly. Although useful for measuring performance, correct-answer scores often provide limited diagnostic information about why a student succeeded or failed and what should be taught next. This paper presents a knowledge-centered analytics framework that shifts the unit of analysis from answers to underlying knowledge. The framework is grounded in the Integrated Knowledge Structure (INKS) model and Cognitive Structure Analysis (CSA), which represent mastery in terms of facts, strategies, procedures, and rationales. Prior CSA studies have found strong relationships between assessed knowledge and problem-solving performance, and experimental studies in high- and middle-school mathematics found larger posttest gains when students' knowledge gaps were assessed and remediated than when incorrect problem-solving steps were assessed and corrected. We describe a software analytics architecture for converting student self-assessment responses into item-, student-, class-, and longitudinal knowledge profiles, and for using those profiles to guide personalized AI remediation. The current software prototype records the four mastery dimensions before and after sessions and contains conventional predictive models, but real-student validation and fuller knowledge-profile reporting remain future work. The framework offers a testable alternative to answer-centered analytics: assessment as diagnosis of the knowledge structures that generate performance.

Introduction

Educational assessment has traditionally emphasized observable performance: whether a learner selects or produces the correct answer, how many items are correct, and whether a score exceeds a criterion. Such measures are indispensable for many summative purposes, but they compress

potentially different causes of performance into a single outcome. A correct answer can arise from robust conceptual understanding, memorized procedures, partial knowledge, or guessing; an incorrect answer can arise from a factual gap, an inappropriate strategy, a procedural error, a missing rationale, or an execution slip. When the instructional question is "What should this learner study next?" the distinction matters.

Formative-assessment research has long emphasized that assessment becomes instructionally powerful when evidence is used to adapt teaching to learners' needs (Black & Wiliam, 1998). Intelligent tutoring research similarly demonstrates the value of maintaining models of learners' evolving knowledge rather than relying only on aggregate test scores (Corbett & Anderson, 1995). The present framework extends this diagnostic logic by explicitly representing several types of knowledge and by making those representations accessible to students themselves through structured self-assessment.

The central proposition is that educational analytics should not stop at determining whether an answer is correct. It should estimate what knowledge a learner possesses, what knowledge is faulty or absent, which gaps the learner can recognize, and how those states change after remediation. This paper describes the theoretical basis, prior empirical evidence, and software architecture for such a system.

From Correct Answers to Knowledge Diagnosis

An answer-centered assessment treats item performance as the principal observable. Item analysis may identify which questions are difficult, discriminate among students, or reveal common distractors, but the instructional interpretation still depends on inferring the knowledge that produced the response. A knowledge-centered system reverses the emphasis. It begins with a model of the knowledge required for competent performance and asks questions designed to sample that knowledge directly.

This distinction is especially important for remediation. Correcting the steps of a missed problem teaches the learner how that particular problem should have been solved. Diagnosing a missing concept can support transfer because the remediation targets a knowledge component that may participate in many problems. The hypothesis is therefore not that correct-answer data are useless, but that knowledge-level diagnosis can provide a more causally informative target for instruction.

The INKS and Cognitive Structure Analysis Framework

The framework derives from the Integrated Knowledge Structure (INKS) approach developed from research on expert knowledge elicitation and representation (Leddo et al., 1990). INKS

integrates multiple forms of knowledge that cognitive research has often treated separately. In its educational application, four categories are emphasized: facts, strategies, procedures, and rationales. Facts describe what entities, concepts, relations, and properties are; strategies represent general goal-directed approaches for solving a class of problems; procedures specify the operations or steps used to implement strategies; and rationales represent the causal or principled knowledge that explains why concepts, strategies, or procedures work.

Cognitive Structure Analysis (CSA) operationalizes INKS as an assessment methodology. Instead of asking only whether students can solve problems, CSA uses targeted queries to assess the component knowledge presumed to support problem solving. Leddo, Li, and Zhang (2022) reported a .966 correlation between overall CSA-assessed Algebra I knowledge and problem-solving performance, with substantial correlations for factual, procedural, strategic, and rationale knowledge separately. Replications and extensions reported correlations of .63 for scientific-method knowledge (Ahmad & Leddo, 2023) and .80 for precalculus knowledge (Zhou & Leddo, 2023). These findings support the premise that a structured knowledge profile can be strongly related to performance while also being more diagnostically specific than a total score.

Self-Assessment as a Scalable Source of Knowledge Data

CSA was initially conceived as an external diagnostic method, but its logic can also be taught to learners. Cynkin and Leddo (2023) showed that students could use a CSA-based structure to self-assess what they did and did not know. Subsequent work has used templates that prompt learners to inventory facts, strategies, procedures, and rationales and then remediate the deficiencies they identify (Ravi & Leddo, 2024). This approach changes the student's role from a passive recipient of a score to an active contributor to a model of his or her own knowledge.

This direction is consistent with the broader self-assessment literature. Panadero, Jönsson, and Botella's (2017) meta-analyses found positive effects of self-assessment on self-regulated learning and self-efficacy. The CSA approach is distinctive, however, because the self-assessment is organized around a cognitive model of subject-matter knowledge and is intended to produce actionable diagnostic data for remediation.

Evidence for Remediating Knowledge Rather Than Problem-Solving Steps

The strongest empirical motivation for knowledge-centered analytics comes from direct comparisons of remediation targets. Leddo and Ahmad (2024) compared high-school Algebra II students who received feedback on step-by-step problem solving with students whose underlying concept knowledge was assessed and remediated using CSA. On a subsequent 20-item problem-solving posttest, the knowledge-assessment condition scored approximately 10 percentage points higher on average than the step-remediation condition.

Challagulla and Leddo (2025) extended this comparison to middle-school mathematics. Students whose knowledge was assessed and remediated using CSA scored approximately 13 percentage points higher on the posttest than students who received remediation focused on problem-solving steps. A later synthesis presented at SITE 2026 summarized both studies and concluded that assessing and remediating concept knowledge produced significantly greater learning than emphasizing correction of step-by-step performance (Leddo, Challagulla, & Ahmad, 2026). These studies motivate an analytics system whose primary objects are knowledge components and knowledge gaps rather than merely right and wrong answers.

A Knowledge-Centered Learning Analytics Architecture

The software work described here was developed in partnership with the University of Virginia. The delivered pipeline records four mastery scores—facts, strategies, procedures, and rationales—before and after each learning session, along with information about how explanations were delivered. The current code uses five conventional prediction models (Random Forest, XGBoost, Ridge Regression, and two Support Vector models) to predict overall learning gain and whether a session would be characterized as too basic, too complex, or just right. The modeling methods themselves are standard; the distinctive feature is the knowledge representation supplied by the CSA framework.

The current models use the four mastery scores as inputs but do not yet generate a per-student knowledge profile or separately track improvement in each knowledge type. Thus, the architecture described below is both a representation of what the software already captures and a specification for the next analytic layer.

Proposed Analytic Classification of Self-Assessment Responses

At the item level, each self-assessment response can be mapped to a knowledge item and knowledge type and then classified according to its diagnostic status. The interface supplied with this project illustrates five outcome categories: Correct, Faulty, Known Gap, Missed, and Irrelevant. These categories distinguish qualitatively different learning states that would be collapsed by a conventional correct/incorrect score.

Correct: The learner expresses the target knowledge accurately.

Faulty: The learner expresses relevant knowledge, but the knowledge is incorrect or materially distorted.

Known Gap: The learner recognizes that the knowledge is missing or incomplete, providing evidence of metacognitive awareness.

Missed: The learner fails to supply a target knowledge component and does not explicitly identify it as a gap.

Irrelevant: The learner supplies content that does not address the target knowledge item.

These categories permit analytics at several levels. At the item level, the system can identify concepts that are widely correct, widely faulty, or frequently missed. At the knowledge-type level, it can determine whether a learner or class is relatively stronger in facts, strategies, procedures, or rationales. At the student level, the distribution of classifications can form a diagnostic profile. Longitudinally, transitions from Faulty, Known Gap, or Missed to Correct can provide direct evidence of conceptual change following remediation.

Illustrative Item-Analysis Output

Table 1 reproduces the visible rows of the supplied METY item-analysis interface as an editable table. The screenshot is an illustrative interface rather than a report of validated empirical results. It shows how the framework can summarize knowledge items by type and by diagnostic classification. The interface also displays 28 total items, 62% average correct, 36% for the most-missed knowledge item, and 18% for the most-faulty incorrect item.

Table 1
Illustrative Knowledge-Centered Item Analysis for Photosynthesis

#	Knowledge item	Type	Correct %	Faulty %	Known Gap %	Missed %	Irrelevant %
1	Plants convert light energy to chemical energy	Facts	86	4	6	2	2
2	Chlorophyll is the pigment that absorbs light	Facts	78	8	10	2	2
3	Write the balanced equation for photosynthesis	Procedures	56	18	18	8	4
4	Identify when photosynthesis occurs	Strategies	48	12	26	10	4
5	Explain why light is required for photosynthesis	Rationales	32	28	26	10	4
6	Describe the role of carbon dioxide	Facts	68	12	14	4	2
7	Explain how glucose is produced	Rationales	30	24	32	10	4
8	Describe the process with a diagram	Procedures	46	18	24	8	4

Note. Values and labels reproduce the visible rows of the supplied interface mock-up. Because the display is illustrative, percentages should not be interpreted as findings from real students.

From Descriptive Analytics to Personalized Remediation

The instructional value of the framework lies in linking diagnosis to action. A conventional item report might tell a teacher that 32% of students answered an item correctly. A knowledge-centered report can indicate that the item concerns a rationale, that 28% of responses contain faulty rationale knowledge, 26% reveal a recognized gap, and 10% miss the rationale without recognizing the gap. Those states suggest different responses. Faulty knowledge may require conceptual correction; a known gap may require targeted explanation or retrieval support; a missed gap may require both metacognitive prompting and instruction.

For AI tutoring, the same profile can condition generation. Rather than asking a chatbot to provide a generic explanation of photosynthesis, the system can provide the learner's knowledge profile as context and request remediation specifically for deficient knowledge components. A learner with strong facts but weak rationales can receive causal explanations and "why" questions; a learner with weak strategies can receive guidance on when and how to approach a class of problems; a learner with procedural gaps can receive worked sequences and practice. The purpose is not merely to vary wording or difficulty, but to personalize instruction around the knowledge structure that appears to be missing.

Recently, we have been conducting experiments in which we incorporate self-assessment into LLM-based chatbots. Students perform an initial self-assessment on the topic they wish to learn about, and that information is fed to the chatbot with the instructions to tailor answers to students' questions to focus on filling self-assessed knowledge gaps. We have tested this approach on a variety of subject areas (Ganne and Leddo, 2026; Rapolu and Leddo, 2026a; Maviti and Leddo, 2025; Maviti, Leddo and Prakash, 2025; Wang and Leddo, 2025). In these tests, students learn a new subject by using either our self-assessment chatbot or a standard LLM like ChatGPT or Gemini. The instruction is followed by a post-test. We find that those using the self-assessment chatbot score, on average, the equivalent of one to four letter grades (roughly 10 to 40 percentage points) higher on the post-test than those who use a standard LLM. Rapolu and Leddo (2026b) found that while students who used self-assessment showed improvement in performance, those who used a self-assessment chatbot showed an additional 10 percentage points improvement. This suggests that AI can enhance the effectiveness of CSA-based self-assessment and remediation by assisting the remediation process. Further research could be done to see if AI can help the assessment process as well.

Class, School, and Longitudinal Analytics

The same data model can support aggregation without losing diagnostic meaning. For a class, teachers could view mean mastery and dispersion for each knowledge type, the most frequently

faulty concepts, the most common known and unrecognized gaps, and the proportion of learners moving from gap states to correct knowledge after instruction. School-level dashboards could aggregate these measures across classes, grades, subjects, or demographic groups while maintaining appropriate privacy protections.

Longitudinal analytics are particularly important because learning is a change process. The present code already stores pre- and post-session mastery measures, but the current analysis predicts only a blended overall gain. A knowledge-centered extension should estimate change separately for facts, strategies, procedures, and rationales and should display the trajectory of each dimension over time. Such trajectories would allow researchers to test whether particular forms of remediation selectively improve the intended knowledge type and whether changes in knowledge precede changes in problem-solving performance.

Relationship to Existing Learning Analytics

The proposed framework complements rather than replaces established student-modeling approaches. Knowledge tracing, for example, estimates the probability that a learner has acquired procedural knowledge components from sequences of observed performance (Corbett & Anderson, 1995). The present approach differs in two ways. First, it explicitly distinguishes factual, strategic, procedural, and rationale knowledge. Second, it can incorporate learners' direct self-reports and open-ended demonstrations of what they know, rather than inferring all knowledge states from correctness histories. A productive future direction is to combine these sources: self-assessed knowledge, externally scored CSA responses, problem-solving behavior, and longitudinal performance.

Research Questions and Validation Plan

Moving from a conceptual analytics framework to a validated system requires real-student data. The current dataset is synthetic; therefore, model accuracy on those data cannot establish educational validity. The next phase should address at least four questions: (1) How reliably can human raters and automated language models classify open-ended self-assessment responses as Correct, Faulty, Known Gap, Missed, or Irrelevant? (2) Do item- and knowledge-type profiles predict subsequent problem-solving performance beyond conventional correct-answer scores? (3) Do changes in the four knowledge dimensions predict learning gains over time? and (4) Does AI remediation conditioned on the diagnosed knowledge profile outperform generic remediation or remediation based only on incorrect answers?

Validation should include inter-rater reliability for response classifications, convergent validity with independent CSA measures, predictive validity for later performance, and randomized comparisons of remediation strategies. Because the framework is intended to guide consequential

instructional decisions, calibration and subgroup analyses should also examine whether classification accuracy and remediation benefits are comparable across age, subject, language background, and demographic groups.

Discussion

The distinction between assessing answers and assessing knowledge changes the purpose of educational analytics. If the goal is ranking, certification, or accountability, a total score may be sufficient. If the goal is improving learning, however, the system must help answer a different question: what does the learner need to know next? CSA and INKS provide a theoretically-motivated vocabulary for answering that question, and self-assessment provides a potentially scalable way to collect learner-generated evidence.

The prior mathematics experiments are especially relevant because they compare two remediation philosophies directly. Students who were shown and corrected problem-solving steps improved less than students whose underlying knowledge was diagnosed and remediated (Leddo & Ahmad, 2024; Challagulla & Leddo, 2025). These findings do not imply that worked examples or procedural feedback are unimportant. Rather, they suggest that remediation may be more effective when procedural errors are interpreted as evidence about underlying knowledge and instruction is directed at that knowledge.

The analytics interface makes this distinction operational. Instead of reporting only a percent correct, it can expose what kind of knowledge is at issue and whether the learner's representation is correct, faulty, knowingly absent, unknowingly missed, or irrelevant. If validated, these distinctions could make dashboards more instructionally actionable and provide AI tutors with structured information for personalization.

Limitations

The present paper describes a framework and software design, not a completed effectiveness trial of the analytics platform. The supplied analysis explicitly notes that its sample students were generated synthetically and that the current predictive models do not yet produce the per-student and per-knowledge-type reports envisioned here. The item-analysis screenshot is therefore treated as an illustrative target output. In addition, the evidence favoring knowledge remediation over step remediation currently rests on a limited number of studies and subject areas. Replication by independent research teams, larger samples, preregistered trials, and comparisons across domains will be important.

Conclusion

Correct answers are outcomes of learning; they are not the learning itself. A student can fail for different cognitive reasons, and those reasons imply different instructional responses. The INKS/CSA framework offers a way to represent those differences through facts, strategies, procedures, and rationales. Prior research indicates that these knowledge measures can be strongly associated with problem-solving performance and that directly remediating knowledge gaps can produce larger gains than correcting problem-solving steps alone. The next step is to convert that research framework into validated learning analytics that continuously identify knowledge states, track their change, and use them to personalize remediation. Such a system would move educational analytics from reporting performance toward explaining—and improving—the knowledge that produces it.

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. <https://doi.org/10.46609/IJSSER.2023.v08i08.020>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- 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*, 10(8), 3574–3586. <https://doi.org/10.46609/IJSSER.2025.v10i08.032>
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278. <https://doi.org/10.1007/BF01099821>
- Cynkin, C., & 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. <https://doi.org/10.46609/IJSSER.2023.v08i09.040>
- Ganne, Y. and Leddo, J. (2026). Evaluating the Impact of Personalized Adaptive AI vs. Standardized ChatGPT on the Short-Term Academic Performance in Adolescents. *International Journal of Social Science and Economic Research*, 11(3), 1059-1072.
- Leddo, J., Challagulla, M., & Ahmad, M. (2026, March). Improving educational performance by assessing and remediating knowledge rather than step-by-step performance. Paper presented

at the Society for Information Technology & Teacher Education International Conference (SITE 2026), Philadelphia, PA.

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

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. <https://doi.org/10.46609/IJSSER.2022.v07i11.011>

Leddo, J., & Ahmad, M. (2024). Do you want to raise student achievement? Then, assess and remediate knowledge, not problem-solving performance. *International Journal of Social Science and Economic Research*, 9(1), 249–261. <https://doi.org/10.46609/IJSSER.2024.v09i01.017>

Panadero, E., Jönsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>

Maviti, A. and Leddo, J. (2025). A Self-assessment Chatbot Greatly Outperforms Chat GPT in Teaching High School Students Calculus. *International Journal of Social Science and Economic Research*, 10(10), 5519-5531.

Maviti, A., Leddo, J. and Prakash, P. (2025). A Self-assessment Chatbot Greatly Outperforms Gemini in Teaching High School Students Calculus. *International Journal of Social Science and Economic Research*, 10(11), 6073-6085

Rapolu, S. and Leddo, J. (2026a). The Impact of a Self-Evaluation Chatbot on High School Students' Learning of Cell Communication. *International Journal of Social Science and Economic Research*, 11(1), 320-333.

Rapolu, S. and Leddo, J. (2026b). A Comparison of the Relative Effectiveness of Student Self-Assessment/Remediation vs. Using a Self-Assessment Chatbot on Raising Student Performance. *International Journal of Social Science and Economic Research*, 11(7), 3613-3629.

Ravi, D., & Leddo, J. (2024). Improving student performance by having students assess and remediate their own knowledge deficiencies. *International Journal of Social Science and Economic Research*, 9(10), 4717–4730. <https://doi.org/10.46609/IJSSER.2024.v09i10.043>

- Wang, Y & Leddo, J. (2025). Improving the Effectiveness of Chatbots by Incorporating Self-Assessed User Knowledge into the Question-Answering Process. *International Journal of Social Science and Economic Research*, 10(8), 3618-3629.
- 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. <https://doi.org/10.46609/IJSSER.2023.v08i09.025>