

## **THE EFFECTIVENESS OF COGNITIVE STRUCTURE ANALYSIS IN ASSESSING STUDENTS' KNOWLEDGE OF THE SCIENTIFIC METHOD**

Mahnoor Ahmad and John Leddo

MyEdMaster, LLC, Leesburg, Virginia, USA

DOI: 10.46609/IJSSER.2023.v08i08.020 URL: <https://doi.org/10.46609/IJSSER.2023.v08i08.020>

Received: 17 August 2023 / Accepted: 27 August 2023 / Published: 31 August 2023

### **ABSTRACT**

Assessing students on their knowledge has been a key part of education, aiding in determining how much students have learned certain concepts. In the past, assessments have focused on whether students give the correct answer to problems, implying that the number of correctly-answered test items is a valid measure of how much students know. However, this emphasis on correct answers has resulted in negligence of assessments that could potentially provide diagnostic feedback to teachers and educators as to what concepts students have mastered, where the gaps in their knowledge are and how to remediate them. Having this framework could greatly benefit classrooms and day-to-day teaching practices. The present paper describes an assessment technique called Cognitive Structure Analysis that is derived from John Leddo's integrated knowledge structure framework (INKS-Leddo et al., 1990) that combines several prominent knowledge representation frameworks in cognitive psychology. While this framework has been used to determine its usefulness to mathematics, it has not been tested in other disciplines. The current paper is determined to test whether this framework can be utilized when it comes to testing students' knowledge in science by assessing them on a specific scientific topic: the scientific method. Using a Google Form, students were assessed on four types of knowledge considered the basis of mastery of scientific method concepts: factual, procedural, strategic, and rationale. Students gave responses to queries, and their results were measured where each type of knowledge was scored and a combined knowledge score was created. Students were then given real Advanced Placement style problems to solve, which generated a problem-solving score. Correlations between each knowledge component score and problem-solving performance were moderate and the correlation between overall CSA-assessed knowledge and problem-solving performance was .63. Factual, strategic and rationale knowledge also showed statistically significant correlations with problem solving performance, but procedural knowledge did not. Future research can investigate CSA's across other scientific topics and other subjects, and

whether incorporating CSA as part of day-to-day classroom instruction can lead to higher student achievement and create more efficient learning practices.

## **Introduction**

In the history of education, assessments have been used as a means of measuring the extent to which students have learned the content that they have been taught. In both classroom settings and in standardized testing, this content is operationally defined as the amount of correct answers a student gives on test questions. In the past, various frameworks have been utilized by teachers and educational organizations to test students' knowledge, typically categorized by whether students are required to select correct answers from a set of answer choices or required to construct their own answers to problems. While both categories of frameworks have their benefits, they do include drawbacks that affect their accuracy in assessing students' knowledge.

Multiple choice assessments require students to select and differentiate the correct answer choices from several distracting answer choices. They are widely used in standardized testing environments and classrooms due to their efficiency when it comes to grading (Chaoui, 2011). However, students often score higher on multiple choice tests than they do on constructive response tests as students can increase their scores through guessing (cf. Elbrink and Waits, 1970; O'Neil and Brown, 1997), which is often cited as a reason why multiple choice tests should not be used.

Constructive assessments require that students formulate their own answers to questions rather than choose from different answer choices as with multiple choice assessments. Researchers find, when investigating math problem solving, that students are more likely to use guessing strategies when given multiple choice tests but are more likely to reason mathematically when given constructive tests (Herman et al., 1994), thus increasing its validity in measuring students' actual learning (Frary, 1985).

These frameworks are based on Classical Test Theory, one of the major pillars of assessment methodology, which assumes that the total number of correctly-answered test items indicates the students' level of knowledge (cf., de Ayala, 2009). The challenge with the key assumption of Classical Test Theory, though, is that the assumption that correct answers indicate learning and vice versa may not be entirely true. A medical analogy works well here. Normally, if a person shows outward signs of illness, s/he is probably sick (although there could be non-medical reasons why a person can appear sick such as overexertion or lack of sleep). Similarly, a student who makes a lot of mistakes on a test probably has a lack of knowledge (unless, for example, they were distracted or sick during the test). However, a person can look healthy and still have an underlying illness. Similarly, a student may get correct answers on a test and have knowledge

deficiencies. They could be regurgitating information or formulas that they do not truly understand or guessing correctly on multiple choice exams.

More importantly, the lack of correct answers does not inform the teacher as to what concepts need to be remediated. A doctor does not stop his/her diagnosis after observing symptoms. The doctor conducts additional tests to discover the cause of the symptoms, so that an appropriate remedy can be applied. Indeed, we would consider it medical malpractice for a doctor to treat only the symptoms and not the underlying causes of diseases. Similarly, an incorrect answer to a test question is a symptom that may indicate an underlying knowledge deficiency. We consider it to be educational malpractice to stop the assessment at that point without diagnosing the underlying knowledge deficiency that is causing that incorrect answer. Unless that cause is identified, how can the appropriate remedial instruction be given?

Therefore, our research indicates that there needs to be a new way to assess students' knowledge, one that takes advantage of testing all aspects of their knowledge. The current paper expands on a presented assessment methodology called Cognitive Structure Analysis (CSA) which was designed to assess the underlying knowledge a student has on certain subjects, so that when a student does make a mistake, the source of that mistake can be identified and refined. This methodology is based on decades of cognitive psychology research that have illustrated that people possess a variety of knowledge types, each of which are organized and used differently in problem solving. Due to the fact that there are different types of knowledge that people possess, our framework is an integration of 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.

We note that the National Council of Teachers of Mathematics (2000) has developed a taxonomy of strands necessary for students to be considered mathematically proficient that uses similar terminology: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning. In many ways, the strands of conceptual, procedural and strategic do correspond to our own. The key difference is that the National Council of Teachers of Mathematics frames these strands in terms of desired skills/behavioral outcomes whereas the INKS framework conceptualizes these in terms of the specific knowledge needed to achieve those outcomes.

The INKS framework is based on research done by John Leddo (Leddo et al., 1990) which showed that true mastery in 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.

In order 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 because previous research describes a strong correlation between user knowledge — as assessed by CSA — and performance in practical problem solving. In one previous research project, we found that the using an automated multiple-choice CSA system to assess student learning produced measures of knowledge that correlated .88 with student problem solving performance and measures of change of knowledge as a result of instruction that correlated .78 with change in performance from pretest to post-test. Moreover, at-risk students taught based on the 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 using CSA in an open-ended response format. In this study, students' algebraic knowledge was assessed and this knowledge assessment was then correlated with problem-solving performance. Students were given open-ended questions to assess their factual (semantic), strategic (script-based), procedural and rationale (mental model) concept knowledge of Algebra 1. The total INKS knowledge and individual component knowledge scores were correlated with total number of correctly solved problems. Results showed correlations of .966 between problem solving and total INKS knowledge, .866 between problem solving and factual knowledge, .937 between problem solving and procedural knowledge, .819 between problem solving and strategic knowledge, and .788 between problem solving and rationale knowledge. In two other projects, assessments produced using the CSA methodology produced assessments of student 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).

In the current project, students were assessed on the scientific method for the facts (semantic knowledge), procedures (production rules), strategies (scripts), and rationales behind the concepts (mental models) they possessed. Each category built upon the last. Afterwards, open-ended questions were constructed and implemented into a post assessment to see if students possessed each of these knowledge components. Students' answers to the questions were analyzed for their correctness. Through the analysis, we were able to obtain scores for each of

the four types of knowledge that are considered necessary to master the scientific method topic, and an overall knowledge score. Additionally, a problem solving score is assessed based on students' correctness in the post assessment. These scores are then correlated together in order to determine the extent to which the INKS knowledge predicts students' knowledge and problem solving performance. The purpose of this project is to extend the research done in the Leddo et. al (2022) to another subject, the scientific method, as well as determine if the results determined in that study will endure when assessing a different subject.

## **Method**

### **Participants**

The participants in this experiment were 21 high school and college students from Northern Virginia. There were 20 female participants and 1 male participant. Participants had a varying level of scientific knowledge. They were not paid for participation, but were still given compensation in the form of 2 service hours (credit for performing 2 hours of volunteer work to help an outside organization)..

### **Materials and Equipment**

A Google Form was created that covered questions surrounding scientific experimental design. Topics and questions were created based on similar problems given in science classrooms.

The skills tested on the knowledge assessment were: identifying common aspects of an experiment, knowing the significance of different types of graphs, labeling graphs, writing different kinds of scientific statements like hypotheses and conclusions, calculating the percent error, designing an experiment, and analyzing the importance of various aspects in the experimental design.

A problem-solving assessment was also given after the knowledge assessment which tested the participant's ability to demonstrate their knowledge using AP Biology style questions, such as identifying variables, explaining terms in the context of a question, determining the function of a method in an experiment, and calculating percent error.

The knowledge assessment contained 4 sections with a different type of question in each section. The first section contained all the fact-based questions that would analyze the participant's ability to provide definitions of aspects of the experimental method.

### ***Fact-based Questions:***

“What is an independent variable?”

“What is a dependent variable?”

“What is a hypothesis?”

“What is the control group?”

“What is the experimental group?”

“What is a placebo?”

“What is a bar graph? What does it represent?”

“What is a line graph? What does it represent?”

“What is a scatter plot? What does it represent?”

“What is a confounding variable?”

Next were the procedure questions, which tested participants’ knowledge of the specific steps one executes in a larger process or strategy.

***Procedure-based Questions:***

“How do you graph error bars?”

“How do you label a graph?”

“How do you write a null and alternative hypothesis?”

“How do you write an experimental hypothesis?”

“In an experiment about testing the effects of colored light on plant growth, what could be a possible hypothesis?”

“In an experiment about the effects of caffeine on plants, what would you need to control for?”

Next, the participants would focus more on the functions of strategy-related processes, with questions that tested their ability to provide examples and write their own statements.

***Strategy-based Questions:***

“How do you graph different types of data? Give an example using a bar graph, line graph, and scatter plot.”

“How do you write a conclusion for your results from an experiment?”

“How do you calculate the percent error? Write out the steps.”

“Design an experiment to test the effects of different types of fertilizer on bean production in bean plants.”

“Describe the process of conducting an experiment.”

Finally, participants were tested on their rationale/critical thinking capabilities. They would determine the importance of a certain element in a larger process.

***Rationale-based Questions:***

“Why is an experiment necessary?”

“Why do we need the experimental method?”

“Why would a placebo be used in an experiment?”

“Why is it important to have large sample sizes?”

“Why is it important to have a control group?”

“Why do we randomly assign people to different groups in an experiment?”

“Why do scientists repeat the experiments that other scientists have already done?”

The second set of materials was the problem-solving assessment, which contained 8 total questions with 2 sub questions in each (except for question 7), giving a total of 15 questions and a total score of 15 points. Participants were required to answer all questions in detail.

The Google Form used is linked below:

<https://forms.gle/MVTszfLhzsWL4wBR7>

The problem-solving assessment contained 2 questions per INKS category which tested students’ ability to apply their knowledge to real Advanced Placement style questions relating to the scientific method.

***The first and fifth question related to fact-based knowledge:***

1. a. Explain what a control group is and then identify the control for this experiment
- b. Identify one constant used in the experiment

- 5) a. Explain what an independent variable is and identify the independent variable in this experiment.
- b. Explain what a confounding variable is and identify a possible confounding variable in this experiment.

***The second and sixth question related to procedural knowledge:***

2. a. Write a hypothesis for this experiment
- b. If the student planned on making a bar graph for this experiment, how would she label it?
- 6) a. Write a null hypothesis about the relationship of fruit fly traits
- b. Write an alternate hypothesis about the relationship between fruit fly traits

***The third and seventh question related to strategy-based knowledge, however, the seventh question only contained one part:***

3. a. What is the percent error for the masses of both species?
- b. Explain how you would design a graph for this data.
- 7) Design an experiment which tests the effects of temperature on the activity of a certain enzyme. Determine which type of graph to use for your experiment and describe how you'd label it. The units used for enzyme activity is  $\mu\text{mol/ml}$ .

***The fourth and eighth question related to rationale-based knowledge:***

4. a. Why did the researchers randomly assign the participants?
- b. Why would the researchers use a control group?
- 8) a. Why did the group of scientists use a large sample size?
- b. What was the purpose of repeating the original experiment?

**Procedure**

The two assessments were administered using a Google Form, which was sent to each participant through email. They were instructed to fill out the form completely and with detail. They were allowed to take as much time as needed. Participants were instructed to not use any outside



resources during the knowledge assessment and problem-solving-assessment, and had to answer each question in the order they came in.

## **Results**

Students' results were analyzed using an answer key that was made based on real answer keys created by school science teachers, along with multiple Advanced Placement test scoring sheets. For each assessment, students received a point for responses that were similar to the answer key created. Incorrect answers received no points. For some of the questions, participants could earn multiple points, such as when having to design 3 different graphs (bar graph, line graph, and scatterplot) in one question. All students' results were analyzed using the same answer key and standard.

The knowledge assessment's four sections were scored using a point system. The fact-based section had 10 questions with each question worth 1 point, the procedure-based section had 6 questions with each question worth 1 point, the strategy-based section had 5 questions with each question worth 1 point except for one question that contained 3 parts, giving a total point value of 7 points for this section, and finally, the rationale-based section had 7 questions with each question being worth 1 point. The entire knowledge assessment section was worth 30 points in total.

The problem-solving assessment followed the same system with 15 points in total. Students' answers that demonstrated a response similar to the answer key were given full credit (1 point). Students' answers that were completely different from official answers or students who did not respond or gave a non-scorable answer such as "I'm not sure" received 0 points.

In order to determine how well the INKS model could be used to represent students' scientific method knowledge as well as how the CSA framework could be used to assess how much of that knowledge would predict their problem solving performance, the knowledge component scores were correlated with problem solving scores. The results of this analysis showed a correlation between total INKS knowledge as assessed by CSA and problem-solving performance of .63,  $df = 19$ ,  $p < .005$

Next, the individual components of the CSA framework were analyzed to see how well they correlated with problem solving performance, as determined by the problem solving assessment. Factual knowledge correlated .64 with problem-solving performance,  $df = 19$ ,  $p < .005$ . Procedural knowledge correlated .04 with problem-solving performance,  $df = 19$ , ns, strategic knowledge correlated .62 with problem-solving performance,  $df = 19$ ,  $p < .005$ , and rationale knowledge correlated .43 with problems-solving performance,  $df = 19$ ,  $p = .05$ . Accordingly, all

components of the INKS framework, except procedural knowledge, correlated significantly with problem solving performance.

## **Discussion**

The results of this project demonstrate the efficiency of using the CSA as a method to determine whether assessing concept knowledge using the INKS framework can help predict problem solving capabilities. The correlations between the assessed student knowledge and problem solving performance were moderate and the overall correlation was 0.63.

When viewing each knowledge category tested, the correlations of factual knowledge, strategic knowledge and rationale knowledge on predicting problem solving performance are .64, .62, and .43, respectively. These knowledge categories were the only types of knowledge that had the ability to predict problem solving performance, with procedural knowledge having no relationship to the actual ability to solve problems. Interestingly, assessing all four types of knowledge did not prove more predictive of problem-solving knowledge than assessing either factual or strategic knowledge alone. Moreover, the correlations obtained in the present study were lower than those obtained in the original Leddo et al. (2022) study.

The difference in predictiveness of CSA of algebraic vs. scientific method problem solving may lie in the nature of the subject matter. Algebra 1 is a highly procedural discipline with very few variations in how to go about solving problems. Indeed, in the Leddo et al. (2022) study, procedural knowledge turned out to be the most predictive of overall problem solving performance. The scientific method, on the other hand, is as much art as it is method as there are many different ways to conduct an experiment. Accordingly, much of the scientific method is a set of core concepts (e.g., independent and dependent variables) and strategies that guide how to conduct experiments (e.g., create control conditions), which may explain why those two types of knowledge were the most predictive of problem solving performance. The very creative nature of scientific research may make it more difficult to articulate a core set of facts, procedures, strategies and rationales that can predict problem solving performance with great accuracy across a wide range of problems.

The CSA framework was easy to implement, indicating that in other educational settings it would be efficient and have minimal disruption to existing teaching methods and involve little training for administration by instructors. Another potential benefit of the CSA framework is that it may be helpful in diagnosing whether a student's mistake was due to a knowledge deficiency or a careless mistake. This was explored in a technology-based implementation of CSA (Liang and Leddo, 2020). In that study, when students made a mistake in solving Algebra 1 problems, the software used CSA to assess the knowledge components necessary to solve the problem. If

one or more knowledge components were found missing, the software concluded that the missing knowledge was the source of the mistake. If the student demonstrated mastery of the relevant knowledge, the software concluded that the student made a careless mistake. The results of the software's assessments were compared to those of experienced math teachers who watched recordings of the problem solving sessions. The percentage agreement between the assessments produced by the software and those of the teachers was in the low to mid 90's, which was statistically equal to the percentage agreement between the assessments produced by the teachers.

Another important research direction is to replicate the above experiment across other disciplines and age groups as the Leddo et al. (2022) and present studies were done with high schoolers and STEM-related subjects. Consideration of not only if the INKS framework has a predictive power that persists but also whether different types of knowledge will have different correlational strengths is necessary to further knowledge about the INKS framework. The scientific method is very abstract in nature, so it is unsurprising that of the four categories of INKS knowledge, factual knowledge and strategic knowledge were the most individually predictive of overall problem solving.

Mechanics in physics is much more procedural in nature, with calculating equations and evaluating trajectories of motion, and so it may be the case that procedural knowledge will prove even more important for assessing knowledge in physics than assessing knowledge of the scientific method. Similarly, while the scientific method is highly experimental in nature, physics involves visual/spatial reasoning - such as force diagrams - which is not accounted for in the present CSA framework. The CSA framework and its theoretical basis, INKS, need to be expanded to incorporate these types of knowledge.

The same type of reasoning may apply to non-science related subjects as well. Writing may incorporate extensive strategic knowledge for organizing material to be written, semantic knowledge requires vocabulary and a mental model/rationale knowledge for literary devices. The relative strengths of each type of knowledge may depend on the type of writing. Reading may involve little procedural knowledge and focus more on strategic/script based knowledge for understanding the structure of text, semantic knowledge for vocabulary, and mental model/rationale knowledge to understand author's purpose. In science, semantic knowledge is used for concepts, procedural knowledge for formulas, strategic knowledge for designing experiments, and mental model/rationale knowledge for scientific principles.

A fascinating question that arises from the present work is whether students can learn to self-assess using CSA. If this were possible, students could assess their own gaps in knowledge and then undertake corrective instruction to fill them. One challenge that may need to be overcome is

that students often are not very reliable in assessing what they know and do not know. Leddo, Clark and Clark (2021) found that middle schoolers who indicated that they understood algebra content they were taught correctly answered only two-thirds of questions based on that content.

Moreover, when middle schoolers indicated that they did not understand a concept, they still correctly answered three-eighths of questions based on that content. However, in the Leddo, Clark and Clark (2021) study, students were not taught how to self-assess their knowledge; they simply relied on a subjective impression of whether or not they understood the content. CSA could serve as a basis for helping students self-assess their knowledge. It may not be the case that students would be able to tell if their self-assessed knowledge is accurate (although they could fact check it), but they may be able to use CSA to identify what gaps they have in their knowledge based on whether they can even answer the questions that comprise the CSA technique.

The most important research question that remains to be addressed is whether CSA, when integrated into daily classroom instruction, can boost student achievement. Here, teachers would use CSA, perhaps as part of the daily homework or in-class assignments, to assess how well students understand key concepts being taught. Any concepts that are typically challenging could easily be There is some preliminary data that suggests this may be the case. Leddo and Sak (1994) found that changes in knowledge as measured by CSA before and after instruction correlated .78 with changes in pre-test/post-test problem solving performance after instruction was given based on the initially assessed needs. However, in this study, the “assessment-instruction-assessment” cycle only happened once. Future research should implement the assessment and instruction cycle on a more continuous basis. As can be seen from the present study’s results and above discussion, CSA offers a great deal of promise, both as an assessment methodology to identify what students know and how these knowledge gaps may impact performance and as part of a strategy for classroom instruction that is designed to boost student performance. However, further research is necessary to demonstrate this.

## **References**

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

Chaoui, N (2011) "Finding Relationships Between Multiple-Choice Math Tests and Their Stem-Equivalent Constructed Responses". CGU Theses & Dissertations. Paper 21.

de Ayala, R. J. (2009). The theory and practice of item response theory. New York: The Guilford Press

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

Elbrink, L., & Waits, B. (Spring, 1970). A Statistical Analysis of MultipleChoice Examinations in Mathematics. *The Two-Year College Mathematics Journal*, 1(1), 25-29.

Herman, J. L., Klein, D. C., Heath, T. M., & Wakai, S. T. (1994). A first look: Are claims for alternative assessment holding up? (CSE Tech. Rep. No. 391). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing

Frary, R. (Spring, 1985). Multiple-Choice Versus Free-Response: A Simulation Study. *Journal of Educational Measurement*, 22, 21-31.

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., 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. 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.

National Council of Teachers of Mathematics (2000). Principles and Standards for School Mathematics. Reston, VA: NCTM.

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

O'Neil Jr., H., & Brown, R. (1997). Differential Effects Of Question Formats In Math Assessment On Metacognition And Affect. *Applied Measurement in Education*, 331-351.

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

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.