International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

DO YOU WANT TO RAISE STUDENT ACHIEVEMENT? THEN, ASSESS AND REMEDIATE KNOWLEDGE, NOT PROBLEM-SOLVING PERFORMANCE

John Leddo and Mahnoor Ahmad

MyEdMaster, LLC

DOI: 10.46609/IJSSER.2024.v09i01.017 URL: https://doi.org/10.46609/IJSSER.2024.v09i01.017

Received: 16 Jan. 2024 / Accepted: 27 Jan. 2024 / Published: 31 Jan. 2024

ABSTRACT

A widespread practice in education is to give students problems to solve, ask them to show their work and, when they make mistakes, show them the correct way to solve the problem to arrive at the right answer. This approach is consistent with the precepts of Classical Test Theory, which operationalizes learning in terms of the number of correct answers given to problems. The present paper challenges this approach by comparing it to an alternative that assesses what subject matter knowledge students have using a methodology called Cognitive Structure Analysis (CSA) and remediating any knowledge gaps. A study was conducted with Algebra 2 students in which students were initially either given problems to solve and feedback in the form of being shown correct step-by-step solutions to the problems or had their concept knowledge assessed using CSA and then provided feedback to remediate any knowledge deficiencies they had. All students were then given a 20-question problem-solving post-test. Results showed that while students given step-by-step feedback did raise their performance in the post-test, students given knowledge concept feedback scored, on average, 10 points higher on the post-test than did students given step-by-step feedback. Results suggest that assessing and remediating concept knowledge may provide a quick, cheap, and easy way to improve academic performance compared to the traditional assessment and remediation approach of emphasizing correct solutions to problems.

Introduction

As far back as we can remember, math classes across America (possibly around the world) and across time shared one feature in common. Whether for homework assignments or tests, teachers would tell their students, “Show all work.” Teachers would then review student problem-solving solutions and if the students made mistakes, the teacher would highlight the incorrect steps and

ISSN: 2455-8834
Volume:09, Issue:01 "January 2024"

show students the corrected versions. The implication is that by seeing what they did wrong and what they should have done, students would improve their problem-solving performance. Indeed, this method is so pervasive that it made its way into educational software. Many programs ask students to solve problems and if students enter the wrong answers, they are shown the correct step-by-step solutions. The purpose of the present paper is to question whether this time-honored tradition of correcting problem-solving mistakes is the best way to improve student performance and to explore whether an alternative approach, i.e., correcting students' knowledge, may be more effective.

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 standardized testing, this content is operationally defined as the number 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 to construct their own answers to problems. While both categories of frameworks have their benefits, they 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

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

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, s/he was 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 they do not truly understand or guessing correctly on multiple-choice exams.

More importantly, the lack of correct answers does not inform the teacher what concepts need to be remediated. A doctor does not stop his/her diagnosis after observing symptoms. Instead, they conduct 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 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?

In previous papers (Leddo et al, 2022; Ahmad and Leddo, 2023; Bekkari and Leddo, 2023; Zhou and Leddo, 2023), we have proposed an alternative method of assessment, one that measures what underlying concepts a student has about the subject matter rather than how well the student performs when solving problems. This method is called Cognitive Structure Analysis (Leddo et al., 1990). 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.

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, and 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.

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

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 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 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 the 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. These findings were extended to precalculus (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 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).

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

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

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 that directly measures that performance with one that measures underlying knowledge would be less appropriate. The purpose of the present study is to address that potential criticism directly, namely to compare the effectiveness of assessing and remediating problem-solving performance (as defined by measuring number of correct answers to problems and by showing the correct solutions, respectively) and assessing and remediating subject matter knowledge as measured using CSA.

Method

Participants

The participants in this experiment were 24 high school and college students from Northern Virginia. There were 18 female participants and 6 male participants. Participants had varying levels of mathematical knowledge. They were not paid for participation but were given compensation in the form of service hours (credit for performing service work to help an external organization).

Materials

Two Google Forms were created that covered questions surrounding logarithms, a specific mathematical topic. Questions were created using the Holt Algebra 2 textbook (Schultz et al., 2004).

The first Google form, the 'Knowledge Assessment,' tested participants' knowledge of logarithms. Questions, in the form of short answer responses, were created that tested factual, strategic, procedural, and rationale knowledge, i.e., knowledge based on the components of the INKS framework. Participants received feedback in the form of the correct concept knowledge associated with each question. The participants who received this Google form were called the 'Knowledge Feedback Group.'

The second Google form, called the 'Problem-Solving Assessment,' tested participants' ability to solve questions based on their knowledge of logarithms. They received correct step-by-step solutions after each problem. The participants who received this Google form were called the Performance Feedback Group.'

The knowledge assessment contained 27 questions with 4 types of questions in each section: fact, procedure, rationale, and strategy.

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

The first section contained all the fact-based questions that would analyze the participant's ability to provide definitions of various aspects of logarithms:

Fact-based Questions:

"What is a base?"

"What is an exponent?"

"What is a coefficient?"

"What is an exponential equation?"

"What is a logarithmic equation?"

"What is the one-to-one property of logarithms?"

"What is the zero property?"

"What is an inverse?"

"What is the product property of logarithms?"

"What is the quotient property of logarithms?"

Next were the procedure questions, which tested participants' knowledge of the specific steps executed in a larger process or strategy:

Procedure-based Questions:

"How does a root become an exponent?"

"How do you write a Logarithmic equation?"

"How do you write $a^b=c$ into logarithmic form?"

"How do you write $a^{(1/b)}=c$ into logarithmic form? "

"How do you write $a^{-b}=c$ into Logarithmic form?"

"What happens to the exponent when you have $\log(a^b)$?"

"What would you need to do when you have $\log(a)/\log(b)$?"

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

Then, 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 do you take the log of both sides when solving an exponential equation?"

"Why would you convert an exponent into a log when solving an exponential equation?"

" What is the function of the base of a logarithmic equation?"

"What is the purpose of an equation?"

"Why are there different properties of logarithms?"

Finally, the participants would focus more on the functions of strategy-related processes, with questions that tested their ability to provide solutions to each equation given:

Strategy-based Questions:

"Describe the step-by-step method of how you would solve or simplify each equation in terms of C."

"$h^c=b$"

"$a=\log_h(C)$"

"$a^c=\sqrt{b}$"

"$a^c=1/\sqrt{b}$"

"$a^c=(b/d)$"

"$a^c=b^d$"

The Problem-Solving Assessment contained 20 questions that mimicked logarithm-related problems seen in a classroom. The problems were separated into groups based on the instructions being asked. After each group of questions, the participant would receive the strategies and answers for that group. All problems were done by hand, except for problems 18, 19, and 20, which allowed for a calculator to be used.

"Find the value of d in each equation:"

1. "$\log_a(p)=b$"

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

2. “log(p)=-10”

3. “$\log_p(36)=2$”

“Expand each expression into one logarithm using the Product and Quotient Property of Logarithms:”

4. “$\log_a(x/b)$”

5. “$\log_a(b*c)$”

“Condense each expression into one logarithm using the Product and Quotient Property of Logarithms:”

6. “clog(a)-dlog(b)+log(p)”

7. “$4\log_a+\log_d$”

8. “$8\log_a-\log_b$”

“Solve using the One-to-One Property of Logarithms:”

9. “$\log_4 10=\log_4 5x$”

10. “ln(x+5)=ln(12+10x)”

“Solve for X:”

11. “$\log_3(ax)-\log_3(b)=c$”

12. “$\log_7(x-5)+\log_7(3x+5)=1$”

13. “$\log_4(x^2)=2$”

14. “$\log_4(2x^2-4x)=2$”

15. “$\log_x(6x-8)=2$”

16. “$\log_2(7x-6)=3$”

17. “$\log_2(x+3)$-log2(2x+1)=3”

“A calculator is allowed for these questions. Round to the thousandths place.”

18. “$6\ln(7x-2)-4=20$”

19. "$4\ln(6x+2)-16=16$"

20. "$5\ln(9x+3)=15$"

The answer key for all questions was created based on the Holt Algebra 2 textbook (Schultz et al., 2004).

Each participant was given 1 point for a correct answer, which meant if their response aligned closely with the answer written on the answer key, they would receive a point. If their response were inaccurate or did not answer the question in the correct way, they would receive no points for that question. No half-credit points were given for any question.

Finally, a 20-question post-test was constructed that contained problems for participants in both conditions to solve. The post-test questions were based on the subject matter covered in the two initial assessments.

The Google Forms used are linked below:

Logarithmic Functions Assessment - Google Forms

Logarithms Problem Solving Assessment - Google Forms

Procedure

Participants were randomly assigned to one of the two groups (the Performance Feedback Group and the Knowledge Feedback Group) with 12 participants assigned to each. Each participant received one assessment, The two assessments were administered through Google Forms, which were sent to each participant through email. The Google Forms contained both the initial assessment and the 20-question post-test. Participants were instructed to complete the form and answer in detail. They were given as much time as needed to complete the task. However, the participants were given no help throughout the session, nor were they allowed to use external resources.

Results

The pre-feedback and post-feedback scores for each group were tabulated. For the Performance Feedback Group, the number of correct answers (out of 20 questions) in both the pre-feedback and post-feedback tests was calculated. The scores of one participant in each condition were eliminated. Specifically, one participant in the Performance Feedback Group scored 17 on the pre-feedback test but only 1 on the post-test. Something similar happened with a participant in the Knowledge Feedback Group. Mean scores for the remaining 11 participants in the

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

Performance Feedback Group were 13.64 (68% correct) for the pre-feedback test and 15 (75%) for the post-test. This difference was statistically significant using a paired t-test, t= 2.43, df = 10, p = .035, suggesting that performance feedback can, indeed, boost student achievement.

For the Knowledge Feedback Group, the number of correct responses (out of 27) to the knowledge queries and the number of correctly answered questions on the post-test were calculated. In this case, a direct comparison between the two scores is not particularly meaningful, nor is there a way to equate answers to knowledge questions on the knowledge assessment with answers to problems on the pre-feedback assessment in the Performance Feedback Group. However, for the sake of disclosure, the mean number of correct answers to the knowledge queries was 21.3 or 79%. The mean number of correct answers to the post-test was 17 or 85%.

A t-test was performed, comparing the mean number of correct answers on the post-test problems. Results showed that students in the Knowledge Feedback Group scored significantly higher than those in the Performance Feedback Group, t = 2.41, df = 20, p = .026.

Discussion

The goal of the present research was to logically extend our previous research on Cognitive Structure Analysis as an alternative assessment method to measuring the number of correct answers as prescribed by Classical Test Theory. Here, we wanted to see if assessing and remediating student knowledge would produce greater student achievement than assessing and remediating student problem-solving performance. Results showed that, while assessing and remediating student problem-solving performance did produce a statistically significant increase in post-test scores, assessing and remediating student knowledge produced even greater post-test scores. The mean difference was 10 points (75% vs 85%) or about one letter grade (C grade vs. B grade).

As such assessing and remediating student knowledge using CSA offers a potential solution to raising student achievement. According to the US Department of Education's National Assessment of Educational Progress (NAEP, 2022), only 36% of US fourth graders scored at or above the proficient level (grade level) in math, while 26% of eighth graders did, and 24% of twelfth graders. Not only do most students perform below grade level in math, based on national standards, but the percentage of students who do perform below grade level increases as they progress through their schooling. In other words, these students fall further and further behind.

If the present research results hold up, CSA offers a quick, cheap and easy way to immediately raise student achievement. This approach can be easily taught to teachers and can be

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

administered through simple media like Google Forms without resorting to rewriting textbooks (and thereby incurring enormous expense).

Admittedly, more research ought to be done. We would want to generalize our approach to other topics in math, other subjects, and other grade levels. However, our previous research in CSA does point to a consistency that how much knowledge a student has as measured by CSA reliably predicts how well they will solve problems in various subjects. We have now extended this research to show that remediating INKS-based knowledge deficiencies assessed by CSA will lead to higher problem-solving performance than simply correcting their problem-solving performance mistakes.

There is another way CSA-based assessment and remediation may boost student achievement. After years of watching students say things like “I don’t get it” or “I’m stuck” when learning new material, we began conducting studies to see if students can be taught to assess their own knowledge using CSA. It turns out that students can be taught to reliably assess what they know and what they do not know (Cynkin and Leddo, 2023; Dandemraju, Dandemraju and Leddo, 2024). The next logical step is to see whether they can use that self-assessment to seek additional instruction to correct those deficiencies. If that turns out to be the case, that would greatly improve students’ self-efficacy in their learning process.

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.

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)

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

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. In preparation.

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

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 Multiple Choice 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., 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.

NAEP Mathematics: National Average Scores. (n.d.). Www.nationsreportcard.gov. https://www.nationsreportcard.gov/mathematics/nation/scores/

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

ISSN: 2455-8834

Volume:09, Issue:01 "January 2024"

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.

Schultz, J., Ellis, W., Hollowell, K., & Kennedy, P. (2004). *Algebra 2* (W. Ellis, Ed.; 1st ed.) [Review of *Algebra 2*]. Holt, Rinehart, and Winston.

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.

