ASEE'S VIRTUAL CONFERENCE

At Home with Engineering Education

JUNE 22 - 26, 2020 #ASEEVC

Paper ID #30104

Measuring student computational thinking in engineering and mathematics: Development and validation of a non-programming assessment

Mr. Timothy Ryan Duckett, The University of Toledo

T. Ryan Duckett is a research associate with Acumen Research and Evaluation, LLC., a program evaluation and grant writing company that specializes in STEM and early childhood education. He is a PhD student in the Research and Measurement department at the University of Toledo.

Dr. Gale A Mentzer, Acumen Research and Evaluation, LLC

Gale A. Mentzer, PhD, the owner and director of Acumen Research and Evaluation, LLC, has been a professional program evaluator since 1998. She holds a PhD in Educational Research and Measurement from The University of Toledo and a Master of Arts in English Literature and Language—a unique combination of specializations that melds quantitative and qualitative methodologies. She and has extensive experience in the evaluation of projects focused on STEM education including evaluations of several multi-million dollar federally funded projects. Previously she taught graduate level courses for the College of Education at The University of Toledo in Statistics, Testing and Grading, Research Design, and Program Evaluation.

Measuring student computational thinking in engineering and mathematics: A work in progress examining the development and validation of a non-programming assessment

This work in progress presentation chronicles the development and validation of an assessment that measures student computational thinking skills (CT). As evidence of the growing need to integrate CT into problem-solving, particularly for ambiguous, open-ended problems, the International Society for Technology in Education created CT Competencies that coincide with the K-12 Computer Science Framework. In its simplest form, CT is "procedural thinking" [1] but over the past 25 years its definition has grown and evolved matching that of computers [2]. Definitions vary among researchers ranging from coding skills to using computers to solve problems to applying computer information-processing methods to one's thinking to define and solve complex problems [3]. For the purposes of this presentation, we use the computational thinking framework that guided the theory of action for the NSF STEM + C funded project, "Understanding How Integrated Computational Thinking, Engineering Design, and Mathematics Can Help Students Solve Scientific and Technical Problems in Career Technical Education (INITIATE) (#1741784). INITIATE used the definition provided by Computational Thinkers (computationalthinkers.com) that divides the process into four steps: (1) students take a complex problem and break it down into a series of small, more manageable problems (decomposition); (2) each of these smaller problems can be looked at individually, considering how similar problems have been solved previously (pattern recognition) and focusing only on the important details; while (3) ignoring irrelevant information (abstraction); and finally, (4) simple steps or rules to solve each smaller problem can be designed (algorithms).

With the emphasis on the development of CT skills comes the challenge of accurately measuring CT. Because of its close association with computer science, CT is often measured using programming tools (such as Scratch, Zoombinis, gaming, or simulation-based situations) on a computer [3]. CT skills, however, go well beyond programming and should be measurable as a skill that one can implement in other problem-solving situations [4]. The majority of CT measures that do not use technology and programming as the medium for measurement are project-specific, examine attitudes towards CT, use a longitudinal approach by examining a project-based process [3], or do not examine the transfer of CT to situations other than computer programming [5].

INITIATE is a three-year project that aims to improve high school student engagement in mathematics as well as attainment of mathematical and CT skills through the integration of project-based learning into the high school mathematics classroom. To do so, Career and Technical Education teachers (typically computer science and manufacturing technology) joined with mathematics teachers in a two-week intensive summer institute that provided opportunities for the joint development of lessons that integrate mathematics with computer science knowledge to program self-driven model automobiles to perform a variety of tasks. The goal was for students to learn to apply mathematics knowledge to the problem of programming the cars thereby realizing the practical value of mathematics. To measure student outcomes, three variables were examined: (1) student engagement; (2) student mathematics ability; and (3) student development of CT skills. Research on CT assessments revealed that existing instruments either use a computer for assessment (such as testing student programming skills or using deductive reasoning using computer fantasy-type games) or a test that was so specific to the project that it could not validly be used in other situations. In addition, few provided opportunities for students to transfer CT knowledge to new, realistic situations and researchers have noted a lack of a comprehensive CT measurement approach [6]. As a result, we set out to develop a portable CT assessment (PCTA) that measured problem-solving skills without being specific to a particular context. The goal was also to develop an instrument that would be economical to administer.

The development of the PCTA began by researching the four steps of CT based upon the ComputationalThinkers.com framework (which in turn is based upon Wing's seminal article [7] which states: *CT requires students to take a complex problem and break it down into a series of small, more manageable problems (decomposition). The smaller problems can be looked at individually, considering how similar problems have been solved previously (pattern recognition) and focusing only on the important details, while ignoring irrelevant information (abstraction). Next, simple steps or rules to solve each smaller problem can be designed (algorithms).* Research on CT assessment indicated that focusing on the *process* students follow to solve problems was essential to uncovering CT skills rather than relying upon summative testing. Assessments should include items that examine how students process, scaffold, and reflect upon information as well as the steps they follow to solve problems including reviewing and correcting errors [8]. The resulting PCTA was developed using items that included process of solving problems based upon the four key steps to CT. In addition, it was essential that the items could be understood by a general audience because students in the treatment classrooms varied in age and educational experience.

The initial PCTA had a total of 15 items. The initial set of questions were developed by the INITIATE research and evaluation team based upon the CT framework. The first eight items were multiple choice asking students about preferred problem-solving process (i.e., "How likely are you to do the following when faced with a complex mathematics problem?"). Using a 4-point scale, students responded to items such as, "Solve what I can and move on" and "Look for similarities between this one and others I have solved in the past." The remaining seven items were open-ended and asked students to elaborate on the steps they would take to solve problems like finding the fastest route from a bus stop to the library (a road map is included) and finding the area of an irregular polygon (students are asked to list the steps or process, not solve the problems; see Figure 1 below). A scoring rubric was developed by the research and evaluation team to describe the components required for correct answers to the open-ended items. The PCTA was then reviewed by teachers in the INITIATE program to provide content validity and to ensure the assessment contained no ambiguous items or instructions that students could not understand.

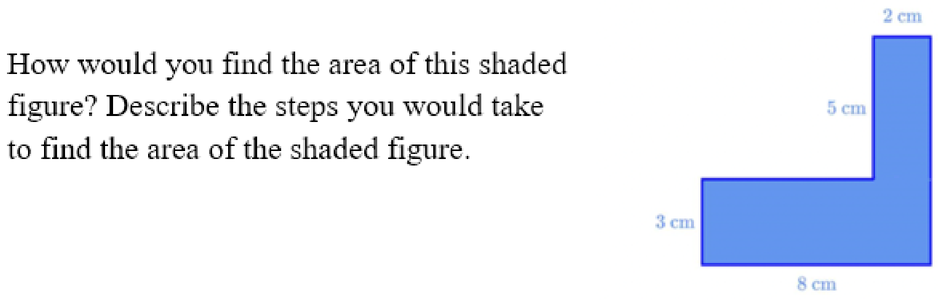
How would you find the area of this shaded figure? Describe the steps you would take to find the area of the shaded figure.

Figure 1. Example of an open-ended problem on the PCTA

The PCTA was administered electronically as a survey but could have been completed on paper in the spring academic semester by teachers participating in the professional development intervention, as well as by control teachers matched by district and course offerings. Demographics for the school district that participated in INITIATE show that 85.7% of the student population were classified as economically disadvantaged; 34.7% of the students received a rating of proficient; and the district overall had a 72% graduation rate. It should be noted that 35 students (approximately 15%) were flagged as not participating faithfully in the testing procedure (e.g., answers to open-ended questions left blank or filled with nonsense, students completing the instrument too quickly to engage with the items).

The responses to the 8 multiple choice items on the 230 surveys (n = 122 and 108 for treatment and control, respectively) were then analyzed using the Rasch-Andrich rating scale model. Generally, the Rasch measurement model converts raw scores into logarithmic estimates of the ability of students taking the test and the difficulty of the items on the tests. Winsteps 3.71 uses an iterative version of the PROX method to provide starting values for the joint maximal likelihood estimation of the free parameters (person ability, item difficulty, and k-1 threshold calibrations) in the data [9, 10]. This procedure builds off a stochastic Guttman pattern that posits as items increase in difficulty, they require higher CT ability on the part of the student in order to pass the item. In other words, students with higher CT skill are more likely to get the more difficult items correct. The ability of the parameters estimated in the Rasch analysis (ability of students and difficulty of items) to explain variance in the observed scores provides evidence for construct validity, i.e. the extent to which we are measuring CT and not a different construct.

Additional outputs were consulted to examine the functionality of the instrument. Mean-square infit and outfit statistics describe the extent to which students and items performed as predicted by the model: students with higher CT ability were expected to get easier items correct; more difficult items would not be answered correctly by lowest performing students. Any residual variance that could not be explained by the estimated measures of student ability and item difficulty was examined to determine if there were unwanted secondary (confounding) measures or noise affecting the measurement of CT ability.

The initial analysis provided a basic summary of the functionality of the PCTA instrument. Two of the multiple choice problems were worded in such a manner that expected students to negatively endorse them (e.g. if the question asked how likely they are to look at a problem as a whole and solve it, then one would expect them to answer "highly unlikely" because approaching a problem as a whole opposes the CT skill of decomposition). The wording of these two questions caused unnecessary confusion and as a result students did not perform as expected (e.g. students with higher abilities did not negatively endorse these items).

Two additional multiple-choice items exhibited slight evidence of misfit pertaining to the measurement of CT ability in high school students. The two items that asked about students' willingness to review textbook or other external sources to find similar examples to help solve a problem (i.e. pattern recognition) were not always endorsed favorably by students with higher levels of CT ability. One explanation could be that these students enjoyed the process of figuring solutions on their own as a sort of intellectual exercise. Additionally, the inclusion of "sources such as the internet" as one such aid may be interpreted as "looking up the answer online" and actually be directly opposed to the CT skills the item attempts to measure.

Overall, the refined 11 item PCTA instrument displayed adequate to good psychometric properties. The instrument was able to distinguish between four levels of student CT ability: very low (corresponding to PCTA scores of 0% to 26%), low (PCTA scores from 32% to 52%), moderate (PCTA scores from 58% to 79%), and high (PCTA scores from 84% to 100%). There was a Cronbach's alpha reliability coefficient of .73, indicating a moderately acceptable probability that another group of similar students completing the PCTA would produce the same estimate and range of CT abilities.

The items formed a meaningful hierarchy that matched the qualitative expectations established prior to testing; those items expected to be easier and require lower levels of CT ability were successfully completed by most students, while those items expected to be more difficult were only successfully passed by students with higher levels of CT ability. Four of the five easiest problems were the multiple choice questions, while the hardest items were the open-ended problems that required students to provide examples of CT skills in action. These findings suggest that it was easier for students to display attitudes that aligned with the spirit of CT, but struggled in applying the CT skills in real-world problems. The identification of this gap between students' proclaimed approaches to problem-solving and difficulty in application can help teachers develop pedagogy that focuses on developing targeted CT skills.

The answers provided by students with three broad levels of abilities to items with a robust range of difficulty accounted for 48.3% of the variance. A principal components analysis revealed the minor presence of a secondary cluster: the multiple-choice items cluster together, suggesting that students answer those items in a manner that is distinct from their responses to the open-ended items. This was expected since the multiple-choice items asked for students to provide their problem-solving strategies and the open-ended items asked them to actually solve problems.

This work in progress paper discussed the development and initial validation of a student CT instrument that can be completed online or as a paper and pencil survey. *With the growing demand to integrate CT into education, this instrument can provide STEM educators with an easy and economical way to measure CT.* While developed as part of a focused project, the PCTA was created as a generic test of CT base [11]d upon mathematics because it was administered to both the intervention and control students in high school mathematics courses.

Pilot study data was collected and analyzed using a Rasch measurement model. Construct validity was assessed through examining how well the data met the requirements of fundamental measurement; e.g. student CT ability can be measured when a student answers items of similarly difficulty; the measurement of CT proceeds in a monotonic and intervallic manner. The instrument displayed adequate reliability and excellent ordering of items from least to most difficult, but several steps could improve the precision of the instrument such as making the tasks in the open-ended items even more explicit. Including additional items, especially at the highest and lowest end of the scale, can lead to a more robust measure of student CT ability.

Given the promising initial findings regarding the PCTA's reliability in measuring CT ability, future data collected from the same population will extend the potential use of the instrument to track growth over time. Revisions to the PCTA are being incorporated to improve its reliability and the scoring rubric for the open-ended items is being reviewed in light of student responses so that it provides a closer match to the types of responses expected. Differential item function analyses will explore any potential biases in the instrument according to age, gender, race, and education level. Eventually the results will be compared between treatment and control groups to provide evidence toward the efficacy of programs that focus on developing teachers' CT instructional competencies.

References

[1] S. Papert, & I. Harel (1991). "Situating constructionism," *Constructionism*, 36(2), 1-11.

[2] S. Cansu, F Cansu (2019). "An overview of computational thinking," *International Journal of Computer Science Education in Schools, 3* (1) 1-11.

[3] V. Shute, C. Sun, & J. Asbell-Clarke (2017). "Demystifying computational thinking'" *Educational Research Review*, *22*, 142-158.

[4] M. Berland & U. Wilensky (2015). “Comparing virtual and physical robotics environments for supporting complex systems and computational thinking,” *Journal of Science Education and Technology 24*(5), 628-647.

[5] M. Bers, L. Flannery, E. Kazakoff, & A. Sullivan, (2014). “Computational thinking and tinkering: Exploration of an early childhood robotics curriculum,” *Computers & Education 72*, 145–157.

[6] B. Zhone, Q. Wang, J. Chen, & Y. Li (2016). “An exploration of three-dimensional integrated assessment for computational thinking,” Journal of Educational Computing Research, 53(4), 562-590.

[7] J. Wing (2006). “Computational thinking,” *Communications of the ACM 49*(3), 33-35.

[8] S. Lye & J. Koh (2014). “Review on teaching and learning of computational thinking through programming: What is next for K-12?” *Computers in Human Behavior*, *41*, 51-61.

[9] J. M. Linacre, "WINSTEPS Rasch measurement computer program," 2006.

[10] B. D. Wright and G. N. Masters, *Rating scale analysis*. MESA press, 1982.

[11] D. Andrich, "A rating formulation for ordered response categories," *Psychometrika*, vol. 43, no. 4, pp. 561-573, 1978.