Developing Computational Thinking Assessments for Elementary Students: Connecting Cognition, Observation, and Interpretation

Brian Gane, bgane@uic.edu, University of Illinois at Chicago
Noor Elagha, nelagh2@uic.edu, University of Illinois at Chicago
Feiya Luo, sophia.l1998@ufl.edu, University of Florida
Ruohan Liu, liuruohan1990@ufl.edu, University of Florida
Wei Yan, wei1993yan@ufl.edu, University of Florida
Carla Strickland, castrickland@uchicago.edu, University of Chicago
Diana Franklin, dmfranklin@uchicago.edu, University of Chicago
Kathryn M. Rich, richkat3@msu.edu, Michigan State University
James W. Pellegrino, pellegjw@uic.edu, University of Illinois at Chicago
Maya Israel, misrael@coe.ufl.edu, University of Florida

Abstract: We developed Computational Thinking (CT) assessments to pair with a curriculum that integrates CT into mathematics, based on CT learning trajectories. We used the assessment triangle to build a validity argument, supported by our evidence-centered design process and preliminary analysis of pilot data. Elementary students completed assessments after receiving instruction that integrated CT into fraction lessons. Data analysis using CTT and IRT procedures suggest items are suited for a range of CT learners.

Major issues

Computational thinking (CT) is a key aspect of recent K-12 standards documents including those for computer science (K-12 Computer Science Framework, 2016) and science (NGSS Lead States, 2013). CT definitions center on characterizing the thought processes engaged in by humans as they express solutions to problems as computational steps that could be executed by an information-processing agent (e.g., a computer program) (Grover & Pea, 2013; Peel & Sadler, 2018). Introducing CT to young students is important: CT practices and skills help students connect concepts, improve the way they express their ideas, and can prepare students for future learning about programming and design. Further, CT practices and skills prepare students to think in abstraction and use systematic strategies in problem solving (Brennan & Resnick, 2012). Our project developed assessments for an integrated math + CT curriculum that we are developing using CT learning trajectories (e.g., Rich, Strickland, Binkowski, Moran, & Franklin, 2017). We use the assessments for dual purposes: to evaluate student learning during instruction and to gather empirical evidence to support and refine the learning trajectories. After instruction with the curriculum, students should be able to demonstrate CT across a variety of trajectories (topics) as they solve items that require CT *knowledge, skills, and abilities* (KSAs).

Theoretical frameworks

The *assessment triangle* connects three related elements of assessment design and use: cognition, observation, and interpretation (NRC, 2001). We use the assessment triangle framework to design assessments and interpret the assessment results (e.g., Streveler et al., 2011). *Evidence-centered design* (ECD) is a framework and design process that considers assessment design across multiple layers of implementation (Mislevy, Steinberg, & Almond, 2003). ECD through its steps and processes ensures that assessment developers focus on all three vertices of the assessment triangle. In an *argument-based approach to validity* (Kane, 1992) one makes claims about the intended interpretive use of the assessment and provides evidence to support those claims. This evidence should include information about the design and empirical evidence from students (Pellegrino et al., 2014).

Design methodology, data sources, and evidence

Our *cognition vertex* of the assessment triangle is built from different CT learning trajectories: sequence, repetition, decomposition, variables, and conditionals. Our *observation vertex* is directly connected to the cognition vertex: We used ECD to create *design patterns* that describe families of assessment items (Mislevy & Haertel, 2006). These design patterns are critical because they define the KSAs that underlie each of the different CT trajectories. A design pattern also includes (a) potential work products and observations of student work, and (b) characteristic and variable item features.

Data were collected from students in Grades 3 (Teacher N = 4; Student N = 90) and 4 (Teacher N = 3; Student N = 54) that participated in the curricular intervention and consented to take part in data collection. The assessment instruments were administered after the curricular lessons. We interpreted students' responses using objective criteria: we coded student responses for accuracy and characteristics of their responses (*interpretation vertex*). To ensure reliability of coding of students' responses we used a multi-rater method. We had high agreement among scorers (Grade 3 $M = .99$, $SD = .01$; Grade 4 $M = .95$, $SD = .07$). Disagreements were resolved through discussion and the scoring criteria and/or rubrics were updated.

Results and findings

In Grade 3, we found a high degree of internal consistency among items ($\alpha = .77$). Further, alpha-if-deleted values indicated that no items had values greater than .77, indicating that dropping any item would lower reliability. CTT analysis of difficulty indicated all items fell within the .2–.8 range of acceptability. Discrimination values were all greater than .2. Although a small sample, we fit a Rasch model to the data (Anderson LR-test, $X^2_{(6)} = 11.14$, $p = .08$). In Grade 4, internal consistency was lower ($\alpha = .40$). Three items had alpha-if-deleted values greater than .4, indicating we could increase internal consistency by dropping these items. These items also had discrimination values < .2, indicating they should be carefully examined for possible revision. We dropped these potentially problematic items before fitting the Rasch model (Anderson LR-test $X^2_{(4)} = 5.50$, $p = .24$).

Significance

Preliminary analyses of item and test performance suggest two items might need to be revised. The difficulty indices suggest that the items are appropriate for students with a range of CT abilities. Although the sample size was small, we fit Rasch models to data from assessment instruments from both grades. These assessment instruments seem able to form a composite score that indicates students' ability along a latent trait (CT) continuum. In developing these CT assessments, all three vertices of the assessment triangle were defined and coordinated, using ECD, to yield specific claims for a validity argument. Further examinations of structural aspects of the instrument are warranted as they might provide evidence for the reliability and validity of sets of items that are related to different CT trajectories (topics).

References

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. Paper presented at the *2012 Annual Meeting of the American Educational Research Association*, Vancouver, Canada.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

K-12 Computer Science Framework (2016). *K-12 computer science framework*. https://www.k12cs.org/

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.) *Educational Measurement* (4th Ed., pp. 17-64). Westport, CT: Praeger Publishers.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspective*, *1*(1), 3–62.

National Research Council [NRC]. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

Peel, A., & Sadler, T. (2018). Computational thinking: Unplugged design of algorithmic explanations. Paper presented at the *2018 NARST Annual International Conference*, Atlanta, GA.

Pellegrino, J. W., DiBello, L. V., & Brophy, S. P. (2014). The science and design of assessment in engineering education. In A. Johri & B. M. Olds (Eds.), *Cambridge Handbook of Engineering Education Research* (pp. 571–598). Cambridge: Cambridge University Press.

Rich, K. M., Strickland, C., Binkowski, T. A., Moran, C., & Franklin, D., (2017). K-8 learning trajectories derived from research literature: Sequence, repetition, conditionals. In *Proceedings of the 2017 ACM Conference on International Computing Education Research (ICER '17)*. New York, NY: ACM.

Streveler, R. A., Miller, R. L., Santiago-Roman, A. I., Nelson, M. A., Geist, M. R., & Olds, B. M. (2011). Rigorous methodology for concept inventory development: Using the 'assessment triangle' to develop and test the Thermal and Transport Science Concept Inventory (TTCI). *International Journal of Engineering Education, 27*(5), 17.

Acknowledgments

This work is supported by a grant from the National Science Foundation #1932920.