International Society of the Learning Sciences

Validating a Code-Free Computational Thinking Assessment for Elementary School Students: Preliminary Results

Arif Rachmatullah, Carol Tate, Andrea Beesley, and Elise Levin-Guracar
arif.rachmatullah, carol.tate, andrea.beesley, elise.levin-guracar@sri.com
SRI International

Abstract: Computational thinking (CT) assessments are increasingly available, but it is still challenging to measure the impact of CT interventions. Assessments employing the same coding context as the intervention confer an advantage to treatment students, and possibly also to comparison group students with prior coding experience. Here we report initial findings of our effort to validate a CT assessment without a coding context and to account for prior coding experience in the validation process.

Introduction

The increasing number of computing-related careers has expanded the essential skills students need to develop throughout their school years. Such skills include computational thinking (CT), which Shute et al. (2017) defined as "the conceptual foundation required to solve problems effectively and efficiently (i.e., algorithmically, with or without the assistance of computers) with solutions that are reusable in different contexts" (p. 151). Due to its roots in computer science (CS), CT has been nurtured through engaging students in computer programming or coding-infused activities, such as scientific computational modeling where students build simulations of scientific phenomena in a coding environment. Therefore, many available CT assessments use coding languages as the context of questions (e.g., Román-González et al., 2017; Wiebe et al., 2019). This practice does not align completely with Shute et al.'s CT definition, which includes contexts beyond programming. In addition, using coding contexts in a CT assessment gives students with prior coding experience an advantage because they are more familiar with such representations compared to students with no experience. Hence, such an assessment might lead to biased interpretations and difficulty measuring the impact of CT activities. Relatedly, code-centric CT assessments might not be usable when conducting a study comparing computational modeling activity to a 'business as usual' condition in which students have not been exposed to coding, as the assessment would be overly aligned with the intervention. We encountered such a situation in the project from which this validation study stems. In response, we assembled a coding-free CT assessment for elementary school students to minimize pre-existing differences in students' prior coding experiences. The questions in this CT assessment measure five out of six CT facets proposed by Shute et al. (2017): decomposition, abstraction, algorithms, debugging, and generalization. This paper attempted to answer one research question: is this CT assessment valid and reliable?

Methods

Context and Participants

This study is part of a larger project focused on exploring the impact of learning mathematics through computer programming on fifth graders' CT skills. The study is still ongoing; the data we used in this paper were from the pilot study. A total of 127 fifth-grade students (age $M = 10.52$, $SD = 0.66$) participated in this pilot study and took the CT assessment. These students were from rural areas in a central US state and predominantly White (72%), 2% Black/African American, 1% Latinx, 6% mixed-race, and the remaining 19% did not provide their racial group information. Regarding gender, 46% of students identified as female, 48% male, and 6% opted not to answer the question. The number of students who reported never having any experience with coding (31%) was almost equal to the number of students who had some coding experience (at least a one-time experience learning how to code; 64%), and 5% of students did not report their prior coding experience.

The Code-Free CT Assessment

The pilot assessment was comprised of 16 questions. The questions were drawn mainly from three CT instruments for elementary school students: Bebras CT Challenges, the National Assessment of Educational Progress's (NAEP) mathematical problem-solving questions, and Math and Computational Thinking Through 3D Making (MPACT) assessment. Among the 16 questions, nine were from Bebras Challenges, five were from NAEP, and two were from MPACT. These questions were selected based on their concordance with Shute et al.'s (2017) CT facets. Each question measures multiple CT facets. Hence, we conceptualized that this assessment is unidimensional – measuring one construct, which is CT.

 © ISLS

Analysis and Validation Procedure

Rasch modeling was performed to gather evidence of the assessment's construct validity. Infit and outfit mean-square (MNSQ) values generated through Rasch modeling were used to evaluate whether each question could differentiate students with low CT scores from ones with higher CT scores. The MNSQ cutoff of 0.7 – 1.3 proposed by Wright and Linacre (1994) was used in this study. In addition, differential item functioning (DIF) analysis was run to identify any questions exhibiting bias based on students' prior coding experience. We used the cutoff of $p < .01$ for X^2_{13} to identify significant DIF as suggested by Choi et al. (2011). The questions either flagged as biased in the DIF analysis and/or having infit and outfit MNSQ values outside the acceptable range were removed. Reliabilities were computed using Rasch item and person reliabilities and Cronbach's alpha.

Initial Findings and Future Directions

We found that almost all questions had MNSQ values within the acceptable range and were not flagged as DIF prior experience, except Questions 13 and 16. Question 13 particularly had both MNSQ values outside the acceptable range and were flagged as DIF item, and Question 16 was flagged as a DIF item. We then removed these two questions and re-ran the analysis. The fourteen-question assessment (Table 1) had acceptable person and separation reliabilities as well as Cronbach's alpha values which were .64, .61, and .66, respectively. Furthermore, even though the findings do not yet have high interpretative power due to the limited sample size, we believe that the results are promising, especially toward establishing a more generalizable CT assessment. Hence, we plan to collect more data to strengthen the current findings and to have a more valid and reliable CT assessment for elementary school students.

Table 1

Psychometric properties of the final fourteen-questions CT assessment based on preliminary analyses

Task ID	Item Location	Outfit MNSQ	Infit MNSQ	DIF Prior Experience (p for X^2_{13})
Q_01	0.07	1.01	1.00	0.405
Q_02	1.11	1.12	1.10	0.808
Q_03	-1.21	0.90	0.98	0.893
Q_04	0.41	1.13	1.08	0.185
Q_05	-0.22	1.05	1.03	0.145
Q_06	0.88	0.88	0.90	0.993
Q_07	0.46	1.07	1.03	0.838
Q_08	-0.75	1.02	1.01	0.606
Q_09	0.49	0.95	0.94	0.150
Q_10	0.10	0.98	0.99	0.044
Q_11	0.50	0.99	1.00	0.655
Q_12	0.29	0.92	0.93	0.761
Q_14	-1.12	0.92	0.95	0.095
Q_15	-0.14	1.06	1.04	0.836

References

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8), 1-30.

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, *72*, 678-691.

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

Wiebe, E., London, J., Aksit, O., Mott, B. W., Boyer, K. E., & Lester, J. C. (2019, February). Development of a lean computational thinking abilities assessment for middle grades students. In *Proceedings of the 50th ACM technical symposium on computer science education* (pp. 456-461).

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8* (3), 370.

Acknowledgments

This research was supported by an Education Innovation and Research grant from the U.S. Department of Education to the eMINTS National Center at the University of Missouri College of Education.