Work-in-Progress Paper—Developing an Evidence-Centered Model for Computational Thinking in Virtual Worlds with Children with Autism

Alex Barrett
Department of Educational Psychology and Learning Systems
Florida State University
Tallahassee, USA
abarrett3@fsu.edu

Nuodi Zhang
Department of Educational Psychology and Learning Systems
Florida State University
Tallahassee, USA
nz21a@fsu.edu

Fengfeng Ke
Department of Educational Psychology and Learning Systems
Florida State University
Tallahassee, USA
fke@fsu.edu

Jewoong Moon
Department of Educational Leadership, Policy, & Technology Studies
The University of Alabama
Tuscaloosa, USA
jmoon19@ua.edu

Zlatko Sokolikj
Department of Scientific Computing
Florida State University
Tallahassee, USA
zs09f@my.fsu.edu

***Abstract*—This work-in-progress paper reports on the establishment of preliminary reliability for a domain-agnostic evidence-centered assessment model to measure computational thinking (CT) in an online virtual world. Preliminary reliability was established between two researchers through manually coding 800 minutes of recorded learning sessions and over 350 minutes of consultation. Participants were three adolescents diagnosed with autism spectrum disorder. Findings indicate an acceptable level of reliability between the two coders, opening the way to more extensive application of the model in future studies.**

Index terms—computational thinking, evidence-centered design, behavioral coding, virtual worlds, autism spectrum disorder

I. Introduction

Jeannette Wing's computational thinking (CT) applies computer science terminology and thought processes to general problem solving [1]. Despite Wing's vision of CT being applied generally [1], most research in CT has focused on computer science, computer programming and other Science, Technology, Engineering, and Mathematics (STEM) subjects [2], [3]. Although the principles of CT readily agree with STEM, some advocate for the application of CT in non-technical domains, as originally conceptualized [4], [5].

Methods for assessing CT competencies continue to be refined [6]—[8]. For example, Siu-Yung Jong et al. and Tsai et al. developed survey-based scales for measuring features of CT [9], [10], however, evidence-centered observational assessments of CT competencies and disposition are less prevalent. Dagiene et al. [11] were among the first to suggest observable identifiers which indicate CT in action using the core competencies of CT confirmed by Selby and Woollard [7], but a reliable model has yet to be established. Furthermore, research in CT and STEM has largely ignored the marginalized population of children with autism spectrum disorder (ASD) [12].

The purpose of this study, therefore, is to measure the preliminary reliability of a model framework for assessing CT competencies in children with ASD. The following sections provide a background of the model's development and a description of how researchers established a preliminary reliability rating, justifying more extensive use of the model.

II. Background

Wing's conceptualization of CT was abstract and introductory [1], but has since been interpreted more concretely by various researchers. Palts and Pedaste provide an overview of the theoretical relationships among CT studies since Wing's original publication [13]. The present study aligns with the branch of CT research originated by Selby and Woollard whose review extracted the most common and well-defined CT competencies from the literature [7]. In addition to these, disposition has also been found to be relevant to successful CT [14]—[16]. Table 1 contains the eight CT dimensions included in the model developed for this study.

Assessment of CT requires a systematic framework for developing valid and reliable measurements of student learning. Evidence-centered design (ECD) is a principled assessment design framework that helps to specify and connect the target skills, work products, and tasks that students will engage in [17], [18]. ECD involves a conceptual assessment framework, where components and procedures needed for the assessment are defined. In this layer, assessment designers will focus on several models. 1) a student model for performance to be evaluated), 2) an evidence model for specifying student behaviors, and 3) a task model for determination of learning tasks that elicit students' target behaviors. ECD has the advantage of evaluating multidimensional and complex performances in dynamic, interactive environments [17], such as virtual worlds.

Research supported by National Science Foundation grant 1837917.

TABLE I. CT DIMENSIONS AND DEFINITIONS

	Dimension	Definition
Competency	Decomposition	The ability to decompose large or complex problems into smaller and manageable ones by functionality, structure, or relationship [7].
	Abstraction	The ability to reduce complexity by leaving out irrelevant details and focusing on essentials [7].
	Algorithmic Thinking	The ability to develop solutions to problems in a step-by-step manner [7].
	Evaluation	The ability to analyze and evaluate the problem-solving strategies and solutions in terms of efficiency and effectiveness [7].
	Generalization	The ability to generalize specific problem-solving strategies or solutions to different contexts [7].
Disposition	Tolerance for Ambiguity	Confidence in dealing with complexity and ambiguity [15].
	Persistence	Patience and endurance in working with difficult problems [15].
	Collaboration	Communicating and working with others to accomplish common goals [14].

III. EVIDENCE-CENTERED COMPUTATIONAL THINKING MODEL

The evidence-centered CT model being developed in this study is shown in Fig. 1. On the left of the model are the core dimensions of CT and the right of the model depicts theory-derived observable behaviors which are indicative of the CT dimension they are connected to. The evidence-centered design of this model affords demonstrative reasoning based on learners' observable behaviors within the CT model [19]. The first step of using an ECD approach is to conduct a domain analysis, identifying major conceptions included in the model for assessment [17]. A total of eight dimensions were identified as the focus of the model (see Table I).

IV. METHOD & RESULTS

Data for preliminary reliability were collected from three purposively-sampled adolescents diagnosed with ASD who participated in a virtual 3D modeling and programming-based learning program to practice CT. Every two weeks the participants engaged in a learning module on the basics of programming non-player characters (NPCs) conducted entirely in OpenSimulator, an online virtual world. Each session was facilitated by a researcher who is a subject matter expert in computer science. All the sessions were recorded and subsequently coded using BORIS© (Behavior Observation Research Interactive Software) [20].

A. Inter-rater Reliability

Reliability of the raters was calculated by a modified percentage agreement method used by Galbraith et al. [21], which is determined by adding up the number of codes with total agreement between researchers within a margin of time (15 seconds in the present study) and then dividing the sum by the total number of codes entered in that session. According to Hartmann et al. [22], an inter-rater agreement percentage of 70% would be considered excellent for a coding system like the one deployed here.

To establish preliminary reliability, two researchers independently observed and coded 11 randomly selected learning sessions of two participants, which comprised 634 minutes of recorded video, or 16% of the total data for these two participants in this learning module. After coding each video researchers compared and discussed their behavioral coding, resulting in over 350 minutes of consultation. During these consultations, any discrepancy between the data was scrutinized until 100% agreement was achieved. Lastly, three additional videos were coded by the researchers, one from each of the three participants who completed the same learning module. These additional sessions totaled 166 minutes of recorded video. The codes derived from these sessions were used to calculate the reported percent agreement of preliminary inter-rater reliability.

B. Coding Procedure

To code observable behaviors of participants, the ethograms of CT dimensions were compiled before coding and revised as needed during the consultation sessions. Behaviors were coded through a series of modifiers, starting with the CT dimension and then the observable behavior, and lastly the state of the behavior (independent or assisted by the facilitator for CT competencies, and desirable (strong or weak) or undesirable for CT disposition). With our model, each behavior was treated as a state event with a starting and a stopping time. Also, each CT competency was coded without imbricating another competency, as CT competencies are distinct from each other, however, overlap between CT competencies and disposition was permitted. In this exploratory phase of the model's use, a fixed time interval for CT competency behaviors was not imposed, but a sample (one 60-minute session from each coder) showed that CT competency behaviors endured for an average of 43.3 seconds (SD 23.3). Disposition behaviors, on the other hand, were confined with a 15-second minimum time interval as terminal points were ambiguous for these behaviors.

Results of the final reliability calculations returned an inter-rater agreement percentage of 86.2%, which is considered strong [22].

V. DISCUSSION AND CONCLUSION

This study reported preliminary evidence of reliability in using a domain-agnostic, evidence-centered model for CT. Reliability was achieved after about 20 hours of coding and consultation. This study corroborates the practicality of using evidence-centered assessments of CT in addition to demonstrating the applicability of CT behavioral coding in online virtual worlds with individuals diagnosed with ASD.

Some unique challenges arose from coding behaviors in online virtual worlds. For instance, when using the evidence-centered CT model for coding, the behaviors needed to be germane to the competency in question, therefore it was important to distinguish between actions that took place in the

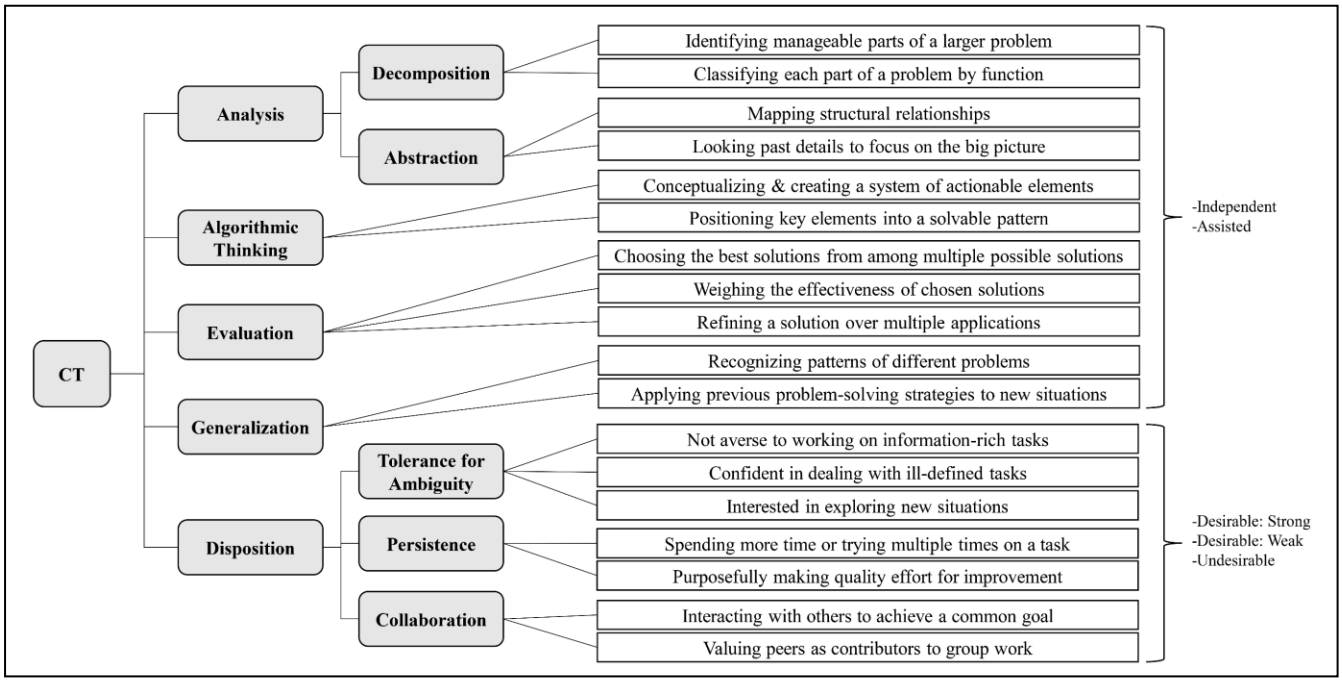

Fig. 1. Evidence-centered model depicting CT dimensions and behaviors.

virtual world and actions that took place in the participants' physical environment. Some participants were distracted by something in their physical environment, or had technical issues with their computers, which could appear as an unresponsive avatar in the virtual world. These task-irrelevant behaviors needed to be recognized so as not to code them as, for example, undesirable persistence.

With preliminary reliability established, the authors are now further validating this evidence-centered CT model in a longitudinal virtual world-based active learning project for adolescents diagnosed with ASD.

REFERENCES

[1] J. M. Wing, "Computational thinking," *Communications of the ACM*, vol. 49, no. 3, pp. 33-35, 2006.

[2] S. Sukirman, L. F. Md Ibharim, C. S. Said, and B. Murtiyasa, "A strategy of learning computational thinking through game based in virtual reality: Systematic review and conceptual framework," *Informatics in Education*, pp. 1-23, 2021.

[3] K.-Y. Tang, T.-L. Chou, and C.-C. Tsai, "A content analysis of computational thinking in research: An international publication trends and research typology," *Asia-Pacific Education Researcher*, vol. 29, no. 1, pp. 9-19, 2020.

[4] B. C. Czerkawski, and E. W. Lyman III, "Exploring issues about computational thinking in higher education," *Tech Trends*, vol. 59, no. 2, pp. 57-65, 2015.

[5] R. P. Lai, "What underlies computational thinking: Exploring its cognitive mechanism and educational implications," In S.C. Kong, et al. Eds. *Proceedings of International Conference on Computational Thinking Education*, 2019.

[6] M. Roman-González, J.-C. Pérez-González, and C. Jiménez-Fernández, "Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test," *Computers in Human Behavior*, pp. 1-14, 2016.

[7] C. C. Selby, and J. Woollard, "Computational thinking: The developing definition," http://eprints.soton.ac.uk/id/eprint/356481, 2013.

[8] J. M. Wing, "Computational thinking and thinking about computing," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1881, pp. 3717–3725, 2008.

[9] M.-J. Tsai, J.-C. Liang, and C.-Y. Hsu, "The computational thinking scale for computer literacy education" *Journal of Educational Computing*, vol. 59, no. 4, pp. 579-602, 2021.

[10] M. Siu-Yung Jong, J. Geng, C. S. Chai, and P.-Y. Lin, "Development and predictive validity of the computational thinking disposition questionnaire," *Sustainability*, vol. 12, no. 4459, pp. 1-17, 2020.

[11] V. Dagiene, S. Sentance, and G. Stupuriene, "Developing a two-dimensional categorization system for educational tasks in informatics," *Informatica (Netherlands)*, vol. 28, no. 1, pp. 23-44, 2017.

[12] C. E. Hawley, E. Cardoso and B. T. McMahon, "Adolescents to adulthood in STEM education and career development" The experience of students at the intersection of underrepresented minortity status and disability," *Journal of Vocational Rehabilitation*, vol. 39, no. 3, pp. 193-204, 2013.

[13] T. Palts, and M. Pedaste, "A model for developing computational thinking skills," *Informatics in Education*, vol. 19, no. 1, pp. 113-128, 2020.

[14] D. Barr, J. Harrison, and L. Conery, "Computational thinking: A digital age skill for everyone," *Learning & Leading with Technology*, vol. 38, no. 6, pp. 20-23, 2011.

[15] V. Barr, and C. Stephenson, "Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community?" *AMC Inroads*, vol. 2, no. 1, pp. 48-54, 2011.

[16] J. Bilbao, E. Bravo, O. Garcia, C. Varela, and C. Rebollar, "Assessment of computational thinking notions in secondary school," *Baltic Journal of Modern Computing*, vol. 5, no. 4, pp.391-397, 2017.

[17] R. J. Mislevy, and G. D. Haertel, "Implications of evidence-centered design for educational testing," *Educational measurement: issues and practice*, vol. 25, no. 4, pp. 6-20, 2006.

[18] R. J. Mislevy, and M. M. Riconscente, "Evidence-centered assessment design," In *The Handbook of Test Development*, New York: Routledge, 2011, pp. 75-104.

[19] R. J. Mislevy, R. G. Almond, and J. F. Lukas, "A brief introduction to evidence-centered design," *ETS Research Report Series*, vol. 1, pp. i-29. 2003.

[20] O. Friard, and M. Gamba, "BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations," *Methods in Ecology and Evolution*, vol. 7, no. 11, pp. 1325-1330, 2016.

[21] L. Galbraith, K. Bull, and C. M. Hill, "Video analysis of parent-child interactions in behavioral sleep disorders: Development of a scoring algorithm," *Frontiers in Psychiatry*, vol. 10, no. 861, pp. 1-8, 2019.

[22] D. P. Hartmann, B. A. Barrios, and D. D. Wood, "Principles of behavioral observation," *Comprehensive handbook of psychological assessment*: *Volume 3, behavioral assessment*, John Wiley & Sons, Inc., pp. 108-127, 2003.