An Engineering Computational Thinking Diagnostic: A Psychometric Analysis

Noemi V. Mendoza Diaz
Dept. of Educational Administration and Human Resource Development
Texas A&M University
College Station, TX, USA
nmendoza@tamu.edu

Deborah A. Trytten
Dept. of Computer Science
University of Oklahoma
Norman, OK, USA
dtrytten@ou.edu

Russ Meier
Dept. of Electrical Engineering and Computer Science
Milwaukee School of Engineering
Milwaukee, WI, USA
meier@msoe.edu

So Yoon Yoon
Dept. of Engineering Education
University of Cincinnati
Cincinnati, OH, USA
yoons5@ucmail.uc.edu

***Abstract*— This research-track work-in-progress paper contributes to engineering education by documenting progress in developing a new standard Engineering Computational Thinking Diagnostic to measure engineering student success in five factors of computational thinking. Over the past year, results from an initial validation attempt were used to refine diagnostic questions. A second statistical validation attempt was then completed in Spring 2021 with 191 student participants at three universities. Statistics show that all diagnostic questions had statistically significant factor loadings onto one general computational thinking factor that incorporates the five original factors of (a) Abstraction, (b) Algorithmic Thinking, (c) Decomposition, (d) Data Representation and Organization, and (e) Impact of Computing. This result was unexpected as our goal was a diagnostic that could discriminate among the five factors. A small population size caused by the virtual delivery of courses during the COVID-19 pandemic may be the explanation and a third round of validation in Fall 2021 is expected to result in a larger population given the return to face-to-face instruction. When statistical validation is completed, the diagnostic will help institutions identify students with strong entry level skills in computational thinking as well as students that require academic support. The diagnostic will inform curriculum design by demonstrating which factors are more accessible to engineering students and which factors need more time and focus in the classroom. The long-term impact of a successfully validated computational thinking diagnostic will be introductory engineering courses that better serve engineering students coming from many backgrounds. This can increase student self-efficacy, improve student retention, and improve student enculturation into the engineering profession. Currently, the diagnostic identifies general computational thinking skill**

Keywords—Diversity, engineering curriculum, computing knowledge, persistence, psychometric analysis

I. INTRODUCTION

This research-track work-in-progress paper is a second-year report on a National Science Foundation funded project to develop a new standard Engineering Computational Thinking diagnostic (ECTD). In a systematic literature review, existing computational thinking frameworks were found to be broad instruments focused primarily on pre-university students [1] [2] [3] [4]. These frameworks did not target engineering students specifically and some targeted only a limited set of computational thinking factors.

Frameworks define computational thinking in various ways. Generally, however, computational thinking is understood to be deeper than computer programming skill [5] [6]. Computational thinkers can ***abstract*** problems into experiments, models, and data that hide details irrelevant to the question at hand. Computational thinkers can ***write algorithms*** that apply mathematics and logic to solve problems. Computational thinkers can ***transform raw data into appropriate data representations*** for analysis. Computational thinkers can ***pattern match similarities in data*** to choose appropriate solution techniques. Computational thinkers can ***decompose problems*** into subsets and ***automate*** solutions. And, computational thinkers ***understand the impact*** of computing solutions as benefits and risks to multiple groups of users. Many of these computational thinking skills are implied within the ABET student outcomes as well as the Taxonomy of Engineering Education [7] [8] [9]. Given the centrality of computing to modern engineering, engineering students must develop computational thinking in engineering problem domains and integrate it in the product design, analysis, and maintenance workflows.

In 2019, our multi-institutional team of researchers began working toward the goal of a targeted diagnostic for engineering students through revision of an initial version of a computational thinking diagnostic developed by one of its investigators in 2017 [10]. The new diagnostic was organized as five computational thinking factors: (1) Abstraction, (2) Algorithmic Thinking and Programming, (3) Data Representation, Organization, and Analysis, (4) Decomposition, and (5) Impact of Computing. For each of the five factors, different diagnostic items were created at each of three levels of developer-perceived difficulty: low, medium, and high. The 2019 diagnostic went through an unsuccessful round of validation that showed excessive correlation between factors [11].

First-year engineering courses often teach computational thinking while introducing computer programming as a computational thinking tool. The goal of a validated diagnostic is a standard assessment to identify students that have strong entry level skills in computational thinking, to identify students that need extra help, and to identify areas for targeted intervention. A broader impact of a successful computational thinking diagnostic is better instruction of engineering students coming from many backgrounds – increasing student self-efficacy, persistence, and student enculturation into the engineering profession.

II. Development and Revisions of the ECTD

The ECTD has been under development for four years. Table I documents the number of engineering students testing each version, as well as the total number of diagnostic questions.

Table I: Development of the ECTD

Time	Version	Questions	Participants
Fall 2017	Pilot	15	1951
Fall 2019	Alpha	Version A: 15 Version B: 15	Version A: 373 Version B: 153
Spring and Fall 2020	Beta	Version A: 15 Version B: 15	Version A: 480 Version B: 436
Spring 2021	Gamma	20	191

The first version of the ECTD was created in 2017 [10]. This pilot version was given to engineering students in their first year of study at a large university in the southwestern United States. Each question had five multiple choice responses. Survey responses were analyzed to assess the quality of the items, and especially that the level of difficulty of each question was consistent with the design.

NSF funded the refinement and expansion of the ECTD in 2019 to two additional institutions: a large public institution in the midwestern United States, and a small private institution in the Great Lakes region of the United States. This research team then created the second version of the ECTD, called ECTD Alpha, in the summer of 2019. The existing diagnostic questions were categorized into the five computational thinking factors noted above and additional questions were added. Two ECTD versions were created so the diagnostic could be given as a pre/post assessment in a class without repeating survey items. Each version was designed to contain three questions of varying difficulty for each factor category: one each of high, medium, and low difficulty. Five hundred and twenty-six first-year engineering students at the large Southwestern university took ECTD alpha versions A and B in the fall of 2019 [11].

The analysis of the ECTD Alpha surveys did not produce the desired psychometric properties. Several pairs of survey items were found to have negative correlation coefficients instead of the positive correlations that were expected. When the eigenvalues were analyzed using an inflection point in the scree plot, five eigenvalues were greater than 1.0 [12] [13]. This result could have indicated that the five factors used to design the diagnostic were present. Unfortunately, the experimentally determined factor loadings from both Version A and B did not match the diagnostic design goal. When evaluating the five-factor model, Version A had one factor supported by seven questions and the other four factors were supported by only one question. The five-factor model for Version B similarly had one factor indicated by eight questions with only one or two questions supporting three other factors. These results caused the team to revise the ECTD again.

The research team created ECTD Beta (versions A and B) by modifying questions from ECTD Alpha that had undesirable psychometric properties. This included the reconsideration of question content and phrasing, as well as the choice of distractors. As one might expect, data collection during the 2020 was challenging due to the COVID-19 pandemic. Students were more reluctant to take the survey, probably due to fatigue. Through multiple recruitment attempts, the large university in the southwestern United States was able to recruit 916 first-year engineering students to take ECTD Beta. This survey had some better psychometric properties, including having positive correlation coefficients between all pairs of items. Unfortunately, only four eigenvalues were found to be greater than 1.0, indicating that the survey was measuring four factors instead of the five it was designed to measure. As with ECTD Alpha, only one or two questions were loaded onto three factors, with a single factor having most of the questions loaded. These problems were found in both the A and B versions of ECTD Beta. Since the psychometric properties of the diagnostic did not match the desired design goal, more modifications were necessary.

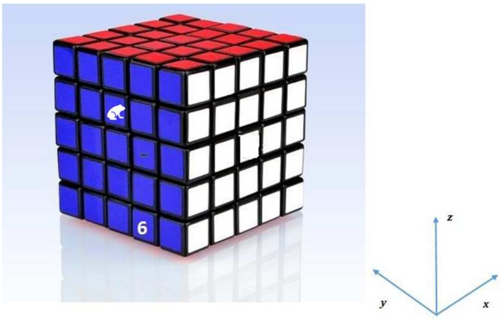
To improve the psychometric properties of the diagnostic, a subset of the items in the A and B versions of ECTD Beta were combined into ECTD Gamma. Four questions were selected for each factor based on measured psychometric properties. Hence, ECTD Gamma has twenty questions. Questions were shortened and simplified to improve clarity and reduce wordiness. In addition, questions were re-examined to remove possible cultural bias, for example, by providing contextual information that some students such as international students or students of low socio-economic status might not have. The revision process for a question is shown in Table II. The exponent in the original question was changed to the whole number 125 because exponentiation is not critical to understanding or solving the problem. The concept of traversal was replaced with counting, which is simpler and more direct. In addition, we added another label for block 26 to help students see that decomposition to layers might be helpful.

III. Method

The purpose of this research phase was to explore the psychometric characteristics of ECTD Gamma for engineering students, such as evidence of construct validity and reliability.

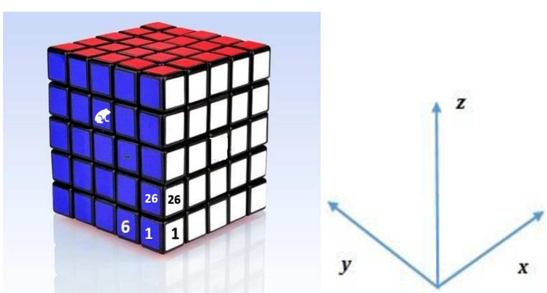
TABLE II: DECOMPOSITION PROBLEM FROM ECTD BETA

The Rubik's cube in the figure is composed of 5^3 (5 to the third power) blocks. A program counts the number of blocks traversed from the origin to the desired block by first traversing along the x-axis, then the y-axis, and finally the z-axis. For example, the block labeled 6 is the 6th block accessed. How many blocks are traversed to get to the block with the frog icon?



a) 6 b) 11 c) 18 d) 78 e) **86**

The Rubik's cube in the figure is composed of 125 blocks. A program counts the number of blocks from the origin of the coordinate system to a given block by counting the blocks along the x-axis, then along the y-axis, and finally along the z-axis. For example, the block labeled 6 is the sixth block counted and the block labeled 26 is the twenty-sixth block counted. How many blocks have been counted by the program when it arrives at the frog icon?



a) 6 b) 11 c) 18 d) 78 e) **86**

These research questions guided the work.

- Does construct validity of the ECTD Gamma hold for engineering students?
- Does internal consistency and reliability of the ECTD Gamma version exist for engineering students?

A. Participants

Study participants were recruited through email at three institutions previously described: a small private Great Lakes university (Institution A), a large midwestern public institution (Institution B), and a large public southwestern university (Institution C). Recruitment response rates were lower than expected, likely due to COVID-19 fatigue. Table III shows the demographic characteristics of the participants. The average age of the 191 students was 19.09 years old with SD = 1.49.

TABLE III: PARTICIPANT DEMOGRAPHICS

Sub categories	A *n*	A %	B *n*	B %	C *n*	C %	Total *n*	Total %
Sex								
Female	0	0.0	9	50.0	63	37.1	72	37.7
Male	3	100	8	44.4	105	61.8	117	60.7
Other/No Answer	0	0.0	1	5.6	2	1.2	3	1.6
Race/Ethnicity								
Hispanic	*0*	*0.0*	*2*	*11.1*	*32*	*18.8*	*34*	*17.8*
Non-Hispanic								
Asian	1	33.3	4	22.2	42	24.7	47	24.6
Black	0	0.0	0	0.0	1	0.6	1	0.5
Multi-racial	0	0.0	2	11.1	10	5.9	12	6.3
White	2	66.7	8	44.4	78	45.9	88	46.1
First Generation								
Yes	1	33.3	2	11.1	20	11.8	23	12.0
No	1	33.3	16	88.9	146	85.9	163	85.3
First Time in College								
Yes	3	100	16	88.9	155	91.2	174	91.9
No	0	0.0	2	11.1	15	8.8	17	8.9
Residency								
Intl.	0	0.0	2	11.1	7	4.1	9	4.7
USA	3	100	16	88.9	163	95.9	182	95.3
Student Level								
Year 1	3	100	10	55.6	120	70.6	133	69.6
Year 2	0	0.0	6	33.3	20	11.8	26	13.6
Year 3	0	0.0	0	0	28	16.5	28	14.7
Year 4	0	0.0	2	11.1	2	1.2	4	2.1
Totals	3	100	18	100	170	100	191	100

Note. Due to unspecified responses, category totals do not add up to 100%.

B. Data Analysis Approach

Student responses on the questions were coded binary, 0 for incorrect and 1 for correct answers. This binary coding is naturally categorical and the distribution of responses for each item was skewed and did not follow a normal distribution. Therefore, robust weighted least squares (WLSMV) employed in Mplus 7.11 was utilized as an estimator to obtain parameter estimates for factor analyses with categorical data [14].

First, an exploratory factor analysis (EFA) was conducted to identify underlying factor structure and irrelevant items that did not fit into any factors that exist in the scale. Eigenvalues and factor loadings after oblique rotation of GEOMIN, which is the default rotation of Mplus, were calculated to judge the number of factors and items for each factor. Second, as we identified a factor structure and items for ECTD Gamma from the EFA, we calculated the reliability coefficient of internal consistency, Cronbach's α, using SPSS Statistics Version 25 to investigate how items are inter-related within each factor, subfactor, and the overall instrument [15]. Finally, we explored item level characteristics, such as item difficulty and discrimination.

IV. RESULTS

A. Exploratory Factory Analysis

Tetrachoric correlation coefficients among the 20 items, which are binary categorical variables, revealed that the coefficients were positively correlated, and ranged from 0.082 to 0.845. Multicollinearity (strong correlations over 0.85) did not exist between items except 0.856 between Items 18 and 19, implying that most of the items do not measure the same aspect of engineering computational thinking ability. Four eigen values (10.4, 1.5, 1.2, and 1.1) were over 1.0, but we extracted the number of factors underlying the data based on the point of inflection of the curve in the scree plot [12]. This yielded one factor considered for inclusion in a putative factor structure for the ECTD Gamma version. According to Stevens' guideline about the relationship between sample size and cutoff factor loading, we considered items with a factor loading greater than 0.40 significant for the designated factor [16]. This cutoff usually functions to suppress any irrelevant items that do not fit well into the designated factor. This resulted in all 20 items that had statistically significant factor loadings onto the one factor, general computational thinking ability incorporating the five categories of (a) Abstraction, (b) Algorithmic Thinking, (c) Decomposition, (d) Data Representation and Organization, and (e) Impact of Computing, indicating each item's unique contribution to the factor. Table IV shows factor loadings of 20 items and their item characteristics, such as item difficulty and item discrimination based on the classical test theory.

B. Item Analysis

Item Difficulty. Here, item difficulty index is defined as the ratio of the number of correct responses to the number of total responses on each item. A higher item difficulty value indicates the easiest item. Among the 20 ECTD Gamma items, Item 1 was the easiest question as 90% of the participants got the question correct and Item 16 was the hardest as only 39% of the participants got the question correct.

Table IV: Exploratory Factor Analysis

Category	Question	Factor Loading	Item Difficulty	Item Discrim.
Abstraction	1. Number of lines	0.640	0.90	0.37
	2. Factorials	0.589	0.66	0.41
	3. Sigma	0.570	0.68	0.41
	4. Arrays	0.651	0.59	0.51
Algorithmic Thinking	5. Find value of w	0.765	0.88	0.52
	6. Variable max	0.877	0.80	0.63
	7. Factor and max	**0.505**	0.73	0.34
	8. Printed sum	0.683	0.52	0.50
Decompose	9.Dashboard	0.796	0.85	0.51
	10. Program	0.738	0.76	0.52
	11. Rubik's cube	0.783	0.53	0.56
	12. Test engineer	0.719	0.78	0.53
Represent Data	13. Natural number	0.765	0.70	0.56
	14. New phone app	0.919	0.82	**0.67**
	15. Weather data	0.527	0.67	0.39
	16. Bridges	**0.434**	0.39	**0.29**
Impact of Computing	17. Self-driving cars	0.554	0.82	0.32
	18. Social media	0.840	0.79	0.60
	19. Smart phones	0.904	0.86	0.60
	20. Website	0.781	0.74	0.57

Item Discrimination: Item discrimination index, which is referred to as item-test correlation or as point-biserial correlation for dichotomously scored items, means the correlation between the item score and the total score. Even though the correlation is dependent on item difficulty, high item correlation is desired because it indicates that high ability respondents tend to get the item correct while low ability respondents tend to get the item incorrect [17]. Therefore, among the 20 items on ECTD Gamma, Item 14 has the highest discrimination and Item 16 has the lowest discrimination.

Internal Consistency Reliability: The overall reliability of the ECTD Gamma with 20 items was Cronbach's $\alpha = 0.878$. All ECTD Gamma items were worthy of inclusion because removal of any items would not increase the score reliability for any construct and ECTD Gamma as a whole [18].

Subgroups: Table VI presents results from independent samples t-test between several subgroups. There was a statistically significant sex difference on the ECTD Gamma with a medium effect size (0.324) that favors male students. Similarly, there was a statistically significant difference in representation on the ECTD Gamma that favors White and Asian students over those from other systemically marginalized racial/ethnic groups with a large effect size. However, there was no difference by student level (first year versus all other years).

Table VI: Subgroup Comparisons of Scores on ECTD Gamma

Sub-cate-gory	N	M	SD	Hedges g	t	df	P
Sex							
Female	72	13.50	4.98	0.324	-2.2	186	**0.031**
Male	116	15.04	4.58				
Representation							
URM	47	12.98	5.38	0.785	-2.4	69.236	**0.017**
ORM	135	15.11	4.45				
Student Level							
Year 1	133	14.80	4.63	0.245	1.6	189	0.120
Year 2 +	58	13.64	5.01				

Note: URM = underrepresented racial/ethnic group, ORM = overrepresented

V. Discussion

Results show items clustering in one factor establishing that such factor is the desired computational thinking construct. We recognize the difficulty of discriminating factors within the overall construct since they are highly correlated; these factors are closely intertwined.

We plan to conduct another iteration of validation during the Fall of 2021. Fall semesters tend to have better response rates, especially since the three institutions are planning to have in-person instruction instead of online instruction this year. We expect that this next validation will show the five desired factors. If it does not, we will begin another round of refining questions until the desired model is achieved.

Our survey results are consistent with the many previous studies that show White/Asian advantage over students from systemically marginalized racial/ethnic groups in computational thinking skills. Similarly, male students were found to have an advantage over female students. The absence of difference in computational thinking skills by student levels is an indication that the instrument may capture an ability that does not change simply because of exposure to college classes, even in engineering.

The limitations of the study include a high chance of sampling bias, considering the low response rates across all three institutions. Second, due to the small sample size of 191, we were only able to conduct EFA. CFA modeling is planned using the data collected in this fall. Third, even though ECTD was designed to assess the five categories, they were not captured as latent factors from the EFA. This could be a result from a low power due to the small sample size. Fourth, for the generalizability of the ECTD, there is a need to test psychometric characteristics of the ECTD using the data collected from additional institutions.

References

[1] J. Krauss and K. Prottsman, *Computational thinking and coding for every student: the teacher's getting-started guide*, 1 edition. Thousand Oaks, California: Corwin, 2016.

[2] Computer Science Teachers Association, "CSTA K-12 Computer Science Standards, Revised 2017," 2017. https://www.csteachers.org/standards (accessed Jun. 30, 2020).

[3] K. Brennan and M. Resnick, "New frameworks for studying and assessing the development of computational thinking," *Proc. 2012 Annu. Meet. Am. Educ. Res. Assoc.*, pp. 1–25, 2012.

[4] College Board, "AP computer science principles." https://apstudents.collegeboard.org/courses/ap-computer-science-principles (accessed Jun. 30, 2020).

[5] C. L. Isbell *et al.*, "(Re)defining computing curricula by (re)defining computing," *ACM SIGCSE Bull.*, vol. 41, no. 4, pp. 195–207, Jan. 2010, doi: 10.1145/1709424.1709462.

[6] P. J. Denning, "The profession of IT: beyond computational thinking," *Commun. ACM*, vol. 52, no. 6, pp. 28–30, Jun. 2009, doi: 10.1145/1516046.1516054.

[7] "Criteria for Accrediting Engineering Programs, 2021 – 2022 | ABET." https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2021-2022/ (accessed Jul. 09, 2021).

[8] C. J. Finelli, M. Borrego, and G. Rasoulifar, "Development of a taxonomy of keywords for engineering education research," *J. Eng. Educ.*, vol. 104, no. 4, pp. 365–387, 2015, doi: 10.1002/jee.20101.

[9] "EER Taxonomy Version 1.2." http://taxonomy.engin.umich.edu/taxonomy/eer-taxonomy-version-1-2/ (accessed May 14, 2021).

[**10**] **N.V.** Mendoza Diaz et al., (2018). Exploring Enculturation in the First-Year Engineering Program-Second Year. *Proceedings of the American Society for Engineering Education 2018*. Salt Lake City, Utah, USA.

[11] **N.V.** Mendoza Diaz et al., Computational Thinking Growth During a First-Year Engineering Course. *Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE)*. Uppsala, Sweden

[12] R. B. Cattell, "The scree test for the number of factors," *Multivar. Behav. Res.*, vol. 1, pp. 245–276, 1966.

[13] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educ. Psychol. Meas.*, vol. 20, pp. 141–151, 1960.

[14] L. K. Muthén and B. O. Muthén, *Mplus User's Guide*, 7th ed. Los Angeles, CA: Muthén and Muthén, 1998.

[15] IBM Corp., *IBM SPSS statistics for Windows Version 25.0*. Armonk, NY: IBM Corp., 2018.

[16] J. P. Stevens, *Applied Multivariate Statistics for the Social Sciences*, 4th ed. Hilsdale, NJ: Erlbaum, 2002.

[17] R. J. de Ayala, *The theory and practice of item response theory.* New York: The Guildford Press, 2009.

[18] A. Field, *Discovering Statistics Using SPSS*, 3rd ed. London: SAGE Publications, Inc.