DOI: 10.34641/ctestem.2022.477

Log-Based Multidimensional Measurement of CT Acquisition

Rotem ISRAEL-FISHELSON, Arnon HERSHKOVITZ*
School of Education, Tel Aviv University
rotemisrael@tauex.tau.ac.il, arnonhe@tauex.tau.ac.il

ABSTRACT

Computational thinking (CT) has been proven challenging to conceptualize and assess. When assessing CT using problem-solving tasks, it is commonly measured based on achievements, that is, in a unidimensional summative way. However, this traditional measurement neglects to consider vital components of the learning progress, which may produce a richer, formative assessment. Using the log files drawn from an online learning platform for CT (Kodetu), we suggest a nuanced evaluation of CT acquisition which consists of four variables: number of attempts to solve a problem; time to solution; application of newly presented CT concept; and solution originality. The research population included 189 middle-school students who participated in a workshop aimed at promoting CT and creativity. Using a learning analytics approach, we analyzed data from a log file documenting 1478 student-task pairs. Findings suggest that these variables share some common features that make them suitable for assessing CT acquisition. Furthermore, the variables grasp different aspects of the learning progress; hence, taken together, they allow for a richer evaluation of CT acquisition. These results shed light on the importance of using diverse metrics to examine CT and contribute to the proliferation of assessment practices.

KEYWORDS

Computational thinking, assessment, CT concepts, achievements, originality

1. INTRODUCTION

A growing trend in educational systems looks to train students in vital skills such as problem-solving and computational thinking (CT). However, several aspects of CT make it challenging to quantify and evaluate it reliably (Blikstein, 2011). The many operational definitions available in the literature and the lack of consensus regarding CT's core features and competencies make it challenging to establish a uniform assessment approach (Cutumisu et al., 2019; Grover et al., 2015). Moreover, because CT is a relatively ill-defined and complex construct, various methods may focus on different dimensions of CT (Weintrop et al., 2021). Various assessment methods were developed and studied following the proliferation of CT-related initiatives. Román-González et al. (2019) proposed a valuable classification of assessment tools based on their evaluation approach. Tang et al. (2020), who reviewed 96 journal articles, offered a more concise classification. Four categories emerged from their analysis: traditional assessment composed of selected- or constructed-response questions, portfolio assessment, survey, and interview. Traditional tests are the most common evaluation method, and they mainly examine the correctness of items for summative purposes (Cutumisu et al., 2019; Metcalf et al., 2021). Such tests, along with surveys, interviews, or observations, are authentic and can lead to a deep understanding of the learning outcomes and required skills (Guenaga et al., 2021). However, focused on the scores achieved, they cannot capture the learning process and draw insights from it (Fields et al., 2019).

Portfolio assessment is used to evaluate CT skills mainly through projects and artifacts, using different rubrics for grading the level of achievement and understanding (Metcalf et al., 2021). Dr. Scratch, for example, draws insights on the application of CT concepts through the analysis of the coding blocks used (Moreno-León et al., 2015). Since CT is not a binary state which developed over time, it is crucial to analyze students' trajectories along the learning experience (Brennan & Resnick, 2012). Portfolio assessment supports such exploration. It enables the capture of the program iterations and identifies patterns and difficulties while solving various challenges (Metcalf et al., 2021). However, this method is often based on human ranking and performed manually (Tang et al., 2020).

Data mining and learning analytics methods are focused on the learning process and are based on automatic analysis of the learning platforms' logged data. They are practical approaches for predicting students' success (Emerson et al., 2019) and detecting difficulties while acquiring CT concepts over time (Román-González et al., 2019). In addition, such methods can help evaluate knowledge acquisition by aggregating students' achievements in learning tasks (Kong, 2019).

Moreover, different indicators that emerge from the logged data can be used to analyze CT's development and provide a multidimensional perspective of CT. Examples of such indicators used in recent studies are the duration and number of attempts to reach a solution, the length of the code, and the number of changes in the code (Eguíluz et al., 2017; Guenaga et al., 2021). Indeed, different indicators, such as score, completion rate, and completion time, provide additional layers of information. However, most of the studies have used such measures in a unidimensional manner and referred to them all as measuring the acquisition of CT. As was recently shown in another domain, this is not necessarily the case (Haleva et al., 2021).

Therefore, this study investigates the processes of CT concept acquisition in a game-based learning platform by performing a multidimensional analysis. We take a learning analytics approach to study the associations between four variables: number of attempts to solve a problem; time to solution; application of newly presented CT concept; and solution originality.

2. METHODOLOGY

2.1. The Learning Environment: Kodetu

Kodetu is a block-based online platform for acquiring CT. The platform is aimed primarily at elementary and middle school students with or without previous coding experience. It offers predefined challenges and enables the independent creation of challenges for the benefit of research or learning. Each challenge is comprised of levels in which the user has to route an astronaut on a given path to a marked destination by dragging coding blocks available in the workspace. Moving to the next level is possible only upon completing the current level, i.e., bringing the astronaut to the marked destination. Notably, users can repeat a level upon completing it and submit another solution. The platform logs all the actions performed by the user. See Guenaga et al. (2021) for further details on this platform.

For this research, we created a dedicated challenge comprised of eight levels that deal with three CT concepts: *Sequences, Loops,* and *Conditionals*. Levels 1-3 focused on the concept of *Sequences*. These levels include blocks representing instructions to move forward and turn right or left. Levels 4-6 present the concept of *Loops*. These levels have a “While” loop that repeats the operations until the astronaut's destination is reached. Finally, levels 7-8 present the concept of *Conditionals* (see Figure 1, for example). In these levels, the users are first presented with a block representing an "if" condition and then with a block representing an "if-else" condition. Each level is built on the previous concept presented.

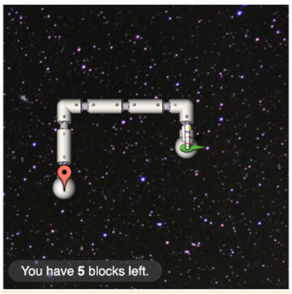

Figure 1. Example of Level 7

2.2. Population and Dataset

The sample comprised 189 ninth-grade students, 15-14 years old. Of them, 40% boys and 60% girls. The vast majority of the students (87%) had no prior coding experience, and 59% had a high affinity for technology. Students were given 80 minutes to solve eight dedicated levels created within the Kodetu platform. This was their first experience with Kodetu, as part of a broader study to examine the associations between CT and creativity (Israel-Fishelson & Hershkovitz, 2022). Data were collected anonymously, with a unique ID assigned to each student. The data were analyzed from the log files retrieved from the Kodetu platform. The log file included 21,784 rows, each representing an action taken by a student, including the users' unique ID, the level at which it was taken, the solution provided (both in Java code and the blocks used), its result [Success, Failure, Timeout, Error], and its timestamp. All statistical analyses were conducted using JASP version 0.16.1.

2.3. Research Variables

2.3.1. Computational Thinking

Three variables were used to measure the acquisition of CT, each computed first for each level separately and then averaged across all levels:

- *Solution Attempts* [#] – counting all solution attempts, including correct and incorrect ones (M=4.82, SD=2.27).
- *Completion Time* [min.] – calculated as the difference between the time of loading a level and the time of moving on to the next level (M=2.35, SD=1.52).
- *Concept Utilization* [0/1] – calculated by checking whether the concept-related blocks were used in the submitted solution. In levels 1-3, all the blocks were related to *sequences*, i.e., moving forward and turning right or left. In levels 4-5, it is examined whether *Loops* have been applied by extracting the command "FOREVER" from the code. Finally, in levels 6-8, it is examined whether *Conditionals* have been used by extracting the command "IF" from the code.

2.3.2. Computational Creativity

To measure the expression of creativity within the Kodetu platform, we have calculated *Solution Originality* as reflected by the frequency of a particular solution among all correct solutions, assessed on a scale of 0-1. In cases when an individual participant submitted several correct solutions, the average frequency of the solutions was taken. This measure was calculated for each level separately and then averaged across all levels (M=0.49, SD=0.1).

3. FINDINGS

To understand how the four research variables are associated with each other, we first checked their values along the game based on CT concepts. Then, we tested for correlations between pairs of them. Finally, we used cluster analysis to classify the participants into groups based on these variables.

3.1. Values of the Research Variables Along the Game

To better understand the acquisition of the three concepts, i.e., *Sequences, Loops,* and *Conditionals,* we conducted 12 pair-wise t-tests between each of the three concepts for each of the four research variables.

For the Solution Attempt variable, we have found an increase in the number of attempts submitted as the challenge progressed (see Figure 2), indicating that the concept of *Sequences* (levels 1-3) required the fewest attempts, followed by *Loops* (levels 5-6) and *Conditionals* (levels 7-8). These differences were significant (at $p<0.001$) with a medium-high effect size (*Sequences-Loops*: d=-0.45; *Loops-Conditionals*: d=-0.96; *Sequences-Conditionals*: d=-1.16). Note that the increase in Solution Attempts variables is not linear.

A similar trend was found for the *Completion Time* variable, as reflected by the increase in the average time to complete the levels along the challenge (see Figure 3). The time to complete the levels dealing with *Sequences* was significantly the shortest, followed by *Loops* and *Conditionals*. As in the case of *Solution Attempts*, this increase in values is also not linear.

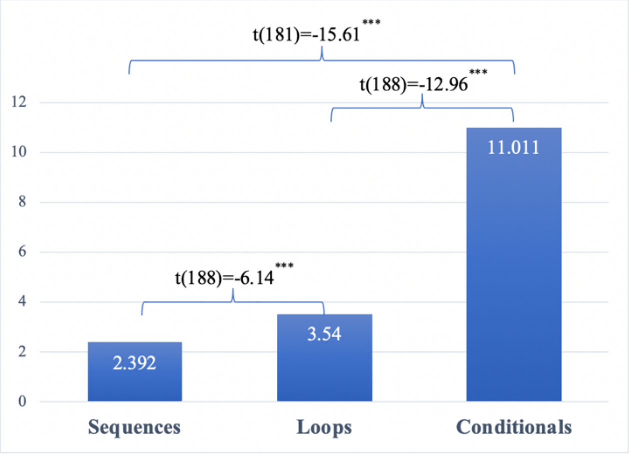

Figure 2. Comparing *Solution Attempts* by CT Concepts (*** p<0.001)

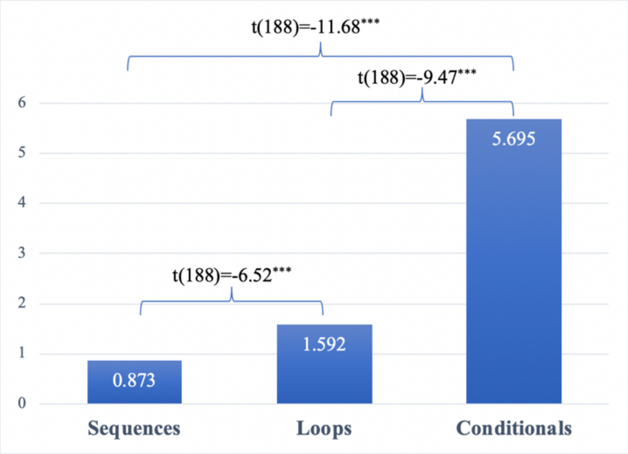

Figure 3. Comparing *Completion Time* by CT concepts (*** p<0.001)

As for *Concept Utilization*, we found that in levels related to the concept of *Loops*, there was a high usage rate of the designated *Loops* block (a "Do-While" loop) in the code. This rate was significantly higher compared to the usage of the designated *Conditionals* blocks ("IF" and "IF-ELSE" blocks) in the related levels (see Figure 4). Note that it is impossible to complete the levels dealing with *Sequences* without using the sequence-related blocks (moving forward and turning right or left), so all the solutions for these levels have implemented the concept of *Sequences*. Therefore, for testing for differences between *Concept Utilization* means in *Loops* and *Conditionals* with *Sequences*, we used a one-sample t-test, comparing them to 1; both were significantly lower than that value.

As for *Solution Originality*, we found that students provided more original solutions as the challenge progressed. The solutions for levels dealing with the concept of *Sequences* were found to be the least original ones, followed by *Loops* and *Conditionals.* These findings were significant (at p<0.001) with a medium-high effect size (*Sequences-Loops*: d=-0.57; *Loops-Conditionals*: d=-0.9; *Sequences-Conditionals*: d=-1.31), and depict a linear increase (see Figure 5).

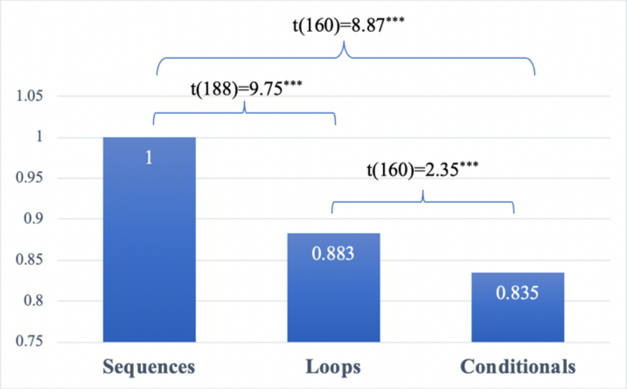

Figure 4 Comparing *Concept Utilization* by CT concepts (*** p<0.001)

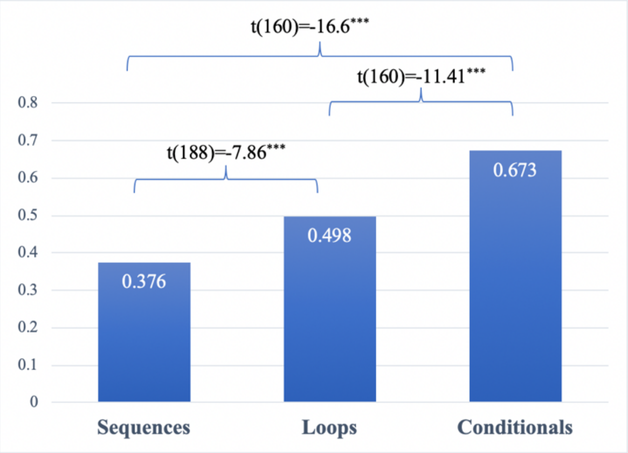

Figure 5. Comparing *Solution Originality* by CT concepts (*** p<0.001)

3.2. Correlations Between the Research Variables

Next, we examined the correlation between the four research variables for the concepts of *Loops* and *Conditional*. Observing these correlations points to three patterns (see Table 1). First, *Solution Attempts*, *Completion Time,* and *Solution Originality* were all positively correlated with each other. The more solutions students submitted, the more time it took them, and the more original their solutions were. For *Solution Attempts* and *Completion Time* specifically, we see very high coefficient values in *Loops* and *Conditionals* (0.85 and 0.69, respectively). This shows the strong connection between these two measures.

A positive correlation was also found between *Concept Utilization* and *Solution Originality* in the levels related to the concept of *Conditionals*. The more students applied the concept of *Conditionals*, the more original their solutions in these levels were. However, it is important to note that the coefficient value, in this case, was low, indicating a low connection between them.

In contrast, a significant negative correlation was found between *Concept Utilization* and *Solution Originality* in the levels related to *Loops*. The more students applied the concept of *Loops*, the less original their solutions were. Also, in this case, the coefficient value was low, indicating a low connection between the measures.

Notably, there were no significant correlations between *Concept Utilization* and *Solution Attempts*, and between *Concept Utilization* and *Completion Time*, neither for *Loops* nor for *Conditions*.

Table 1. Correlations between Research Variables, Per CT Concept

	Var 1	**Var 2**	**ρ**
Loops	Solution Attempts	Completion Time	0.85^{***}
	Solution Attempts	Concept Utilization	-0.05
	Solution Attempts	Solution Originality	0.26^{***}
	Completion Time	Concept Utilization	0.01
	Completion Time	Solution Originality	0.21^{**}
	Concept Utilization	Solution Originality	-0.17^{*}
Conditionals	Solution Attempts	Completion Time	0.69^{***}
	Solution Attempts	Concept Utilization	0
	Solution Attempts	Solution Originality	0.26^{***}
	Completion Time	Concept Utilization	0.04
	Completion Time	Solution Originality	0.22^{**}
	Concept Utilization	Solution Originality	0.21^{**}

* p < 0.05, ** p < 0.01, *** p < 0.001

3.3. Clustering Students by the Research Variables

After examining the different variables and their behavior, we analyzed the research population on a higher granularity level. To that end, we used an unsupervised hierarchical clustering algorithm based on the four research variables.

The hierarchical clustering algorithm aims to partition and group objects based on their similarities. The similarity between the cluster was measured using Pearson's distance, using Ward.D linkage (Ward, 1963), with variables scaled by a Z-score standardization of a mean of 0 and a standard deviation of 1. The elbow method indicated that five is the optimal number of clusters for our dataset. These clusters represent five sub-populations with distinct characteristics, as detailed below (see Figure 6 and Table 2).

Cluster 1 (N=41) includes students who quickly solved the challenge with the least number of attempts. Their application of the CT concepts was among the highest, but their solutions were the least original. Cluster 2 (N=37) includes students who required fewer attempts to solve the levels and the least time for completion. Additionally, they provided the most original solution, and their utilization of the concepts was relatively high. The students in this cluster had the best performance. Students in Cluster 3 (N=44) demonstrated mediocre performance. They were able to solve the levels in the shortest time with a low number of attempts. They were relatively original in providing the solutions but did little use of the concepts learned. Cluster 4 (N=39) included low-performing students. They solved the levels in the longest time and with the most attempts. They provided the second to lowest original solutions, and their implementation of the concepts was also low. Cluster 5 (N=28) also included students with moderate performance. They submitted many solutions but were able to solve the challenge in a short time. Their usage of the concepts was the highest, but their solutions were not so original.

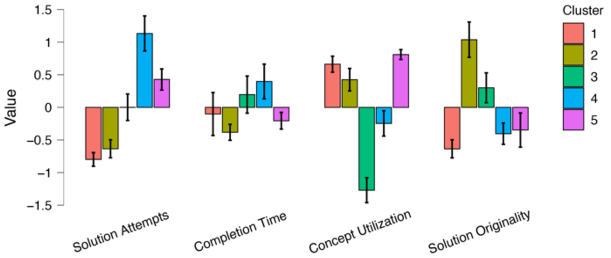

Figure 4. Clusters representing students' behavior according to the research variables

Table 2. Cluster Means. Grey background marks the value in each column that is indicative of the highest performance in CT acquirement or originality; numbers in italics indicate, for each column, the lowest performance in CT acquirement or originality

Cluster	N	Solution Attempts	Completion Time	Concept Utilization	Solution Originality
1	41	-0.798	-0.102	0.662	-0.636
2	37	-0.635	-0.383	0.424	1.038
3	44	0.001	0.196	*-1.271*	0.3
4	39	*1.133*	*0.397*	-0.246	-0.404
5	28	0.428	-0.206	0.81	-0.348

4. DISCUSSION

CT is most often measured by achievements in a unidimensional summative way. However, such an evaluation approach neglects to consider essential factors of the learning process that may produce a richer assessment. This study investigated a multidimensional evaluation of CT acquisition by 189 middle-school students who used an online gamed-based platform. We evaluated students' performance according to four dimensions: *Solution Attempts* (number of attempts to solve a problem); *Completion Time* (time to complete a challenge); *Concept Utilization* (application of newly presented CT concept); and *Solution Originality* (frequency of a correct solution in the set of all correct solutions). Our findings indicate complex relationships between these measures and suggest that they may capture different aspects of the learning process.

Superficially, as it seems from the exploratory analysis, the four variables demonstrate a similar learning behavior. *Solutions Attempts* and *Completion Time* increased as the game progressed, hence may be seen as proxies for difficulty. Moreover, both variables increase in a non-linear way. *Concept Utilization* decreased as the game progressed, demonstrating that correct solutions were often not implementing newly-taught concepts but rather relied on previous knowledge – which, again, reflects the increased difficulty. Finally, *Solution Originality* increased, which may be explained by the fact that the overall set of solutions within our research population increased along the game, echoing the behavior depicted by *Concept Utilization*. Indeed, both variables decrease or increase relatively linearly. This is in line with previous studies, which pointed out some difficulties and misconceptions regarding the concepts of *Loop*s and *Conditionals* (Grover & Basu, 2017; Israel-Fishelson & Hershkovitz, 2019; Weintrop & Wilensky, 2015). Sleeman et al. (1986) argued that such difficulties and misconceptions could stem from a limited understanding of the execution of “if” and “if-else” conditions, as well as from having a faulty understanding of which lines of codes would repeat themselves in “for” and “while” structures and the number of times the code would run.

However, further analysis has shown that the picture is more complex than this. First, correlations between pairs of variables slightly change when tested in different game levels. *For example, concept Utilization* and *Solution Originality* were negatively associated while engaging with the *Loops*-related levels and positively associated with the *Conditionals*-related levels. Additionally, *Solution Attempts* and *Completion Time* were more strongly correlated in *Loops*- than in *Conditionals*-related tasks (in both cases, the correlation was positive).

Second, when clustering students based on their behavior throughout the game, we observe even more complicated relationships. For example, we have a cluster where *Solution Attempts* and *Completion Time* are both low, on average (Cluster 2), a cluster where they are both relatively high (Cluster 4), and a cluster when one of them is high, and the other is low (Cluster 5). Similarly, for *Concept Utilization* and *Solution Originality*, we have clusters that demonstrate different behaviors (respectively): high-high (Cluster 2), high-low (Cluster 3), low-high (Clusters 1, 5), low-low (Cluster 4). Taken together, these findings suggest that CT acquirement measures depend on both personal and contextual characteristics. Indeed, previous studies have shown the importance of contextual factors in the acquisition of CT, and creativity, which is seemingly associated with personal characteristics, is also impacted by contextual factors (Hershkovitz et al., 2019; Israel-Fishelson & Hershkovitz, 2021).

As clearly evident by the cluster analysis, the misalignment between the different measures strengthens the notion that they may each grasp a different aspect of learning. Most easily explained is *Completion Time*, that is, time on task. This measure was shown in other contexts and settings to be impacted by factors other than “knowing” the subject matter, e.g., skill level or graphical user interface, hence it is not necessarily correlated with other, more traditional, measures like achievements or the number of attempts to solve a problem (Goldhammer et al., 2014; Haleva et al., 2021; Hershkovitz et al., 2019).

Our creativity measure, *Solution Originality*, is overall associated with *Solution Attempts* and *Concept Utilization.* This may be explained by the positive association previously suggested between creativity and difficulty (Espedido & Searle, 2018). It also echoes Epstein et al.’s (2008) claim that people can increase their production of new ideas and creative expression when facing challenging problem-solving situations. However, the associations between our creativity measure and our *Concept Utilization* were alternately positive and negative (when tested for each topic separately). This reflects that creativity and knowledge are not necessarily tied together (Edmonds & Candy, 2002). It is possible that students who had difficulty in solving the levels adopted a tinkering strategy which was found effective when learning to program (Berland et al., 2013). Thus, for *Loops* levels, which are considered easier, lower originality rates were observed compared to the *Conditionals* levels, which are considered harder.

This study contributes to the growing body of knowledge on CT assessment while emphasizing the importance of using diverse metrics to examine CT. Taking a log-based approach, we were able to identify nuanced relationships between the different measures throughout the learning process. These associations should be further investigated, on a larger scale, in other populations and contexts. Still, we hope that these findings will encourage researchers to consider the combination of different CT assessment indices to get a more fine-grained, rich assessment.

5. REFERENCES

Baker, R. S. J. d. (2007). Is gaming the system state-or-trait? Educational data mining through the multi-contextual application of a validated behavioral model. *Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*, 76–80. http://www.columbia.edu/~rsb2162/B2007B.pdf

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *Journal of the Learning Sciences*, *22*(4), 564–599.

Blikstein, P. (2011). Using learning analytics to assess students’ behavior in open-ended programming tasks. In P. Long, G. Siemens, G. Conole, & D. Gasevic (Eds.), *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK ’11* (pp. 110–116). ACM Press.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In C. A. Tyson & A. F. Ball (Eds.), *Proceedings of the 2012 Annual Meeting of the American Educational Research Association* (pp. 1–25). American Educational Research Association.

Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology*, *28*(6), 651–676. https://doi.org/10.1007/s10956-019-09799-3

Eguíluz, A., Guenaga, M., Garaizar, P., & Olivares-Rodríguez, C. (2017). Exploring the progression of early programmers in a set of computational thinking challenges via clickstream analysis. *IEEE Transactions on Emerging Topics in Computing*, *8*(1), 256–261. https://doi.org/10.1109/TETC.2017.2768550

Emerson, A., Smith, A., Smith, C., Rodríguez, F., Wiebe, E., Mott, B., Boyer, K., & Lester, J. (2019). Predicting early and often: Predictive student modeling for block-based programming environments. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 39–48).

Epstein, R., Schmidt, S. M., & Warfel, R. (2008). Measuring and training creativity competencies: Validation of a new test. *Creativity Research Journal*, *20*(1), 7–12. https://doi.org/10.1080/10400410701839876

Espedido, A., & Searle, B. J. (2018). Goal difficulty and creative performance: The mediating role of stress appraisal. In *Human Performance* (Vol. 31, Issue 3, pp. 179–196). https://doi.org/10.1080/08959285.2018.1499024

Fields, D. A., Lui, D., & Kafai, Y. B. (2019). Teaching computational thinking with electronic textiles: Modeling iterative practices and supporting personal projects in exploring computer science. In *Computational Thinking Education* (pp. 279–294). Springer Singapore. https://doi.org/10.1007/978-981-13-6528-7_16

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608–626. https://doi.org/10.1037/a0034716

Grover, S., & Basu, S. (2017). Measuring student learning in introductory block-based programming: examining misconceptions of loops, variables, and boolean logic. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education - SIGCSE '17*, 267–272.

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education*, *25*(2), 199–237.

Guenaga, M., Eguíluz, A., Garaizar, P., & Gibaja, J. (2021). How do students develop computational thinking? Assessing early programmers in a maze-based online game. *Computer Science Education*, *31*(2), 259–289. https://doi.org/10.1080/08993408.2021.1903248

Haleva, L., Hershkovitz, A., & Tabach, M. (2021). Students' activity in an online learning environment for mathematics: The role of thinking levels. *Journal of Educational Computing Research*, *59*(4), 686–712. https://doi.org/10.1177/0735633120972057

Hershkovitz, A., Sitman, R., Israel-Fishelson, R., Eguíluz, A., Garaizar, P., & Guenaga, M. (2019). Creativity in the acquisition of computational thinking. *Interactive Learning Environments*, *27*(5–6), 628–644. https://doi.org/10.1080/10494820.2019.1610451

Israel-Fishelson, R., & Hershkovitz, A. (2019). Persistence and achievement in acquiring computational thinking concepts: A large-scale log-based analysis. In S. Carliner (Ed.), *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Association for the Advancement of Computing in Education (AACE).

Israel-Fishelson, R., & Hershkovitz, A. (2021). Micro-persistence and difficulty in a game-based learning environment for computational thinking acquisition. *Journal of Computer Assisted Learning*, *37*(3), 839–850. https://doi.org/10.1111/jcal.12527

Israel-Fishelson, R., & Hershkovitz, A. (2022). One small step to man, a giant step to computational thinking: Improving student performances by promoting creativity. Seventeenth Chais Conference for the Study of Innovation and Learning Technologies, 36–46 [Hebrew].

Kong, S. (2019). Components and methods of evaluating computational thinking for fostering creative problem-solvers in senior primary school education. In S. Kong & H. Abelson (Eds.), *Computational Thinking Education* (pp. 119–142). Springer.

Metcalf, S. J., Reilly, J. M., Jeon, S., Wang, A., Pyers, A., Brennan, K., & Dede, C. (2021). Assessing computational thinking through the lenses of functionality and computational fluency. *Computer Science Education*, *31*(2), 199–223. https://doi.org/10.1080/08993408.2020.1866932

Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. *RED. Revista de Educación a Distancia*, *15*(46), 1–23. https://doi.org/10.6018/red/46/10

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In S. Kong & H. Abelson (Eds.), *Computational Thinking Education* (pp. 79–98). Springer.

Sleeman, D., Putnam, R. T., Baxter, J., & Kuspa, L. (1986). Pascal and high school students: A study of errors. *Journal of Educational Computing Research*, *2*(1), 5–23.

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103798

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Weintrop, D., & Wilensky, U. (2015). Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. *ICER*, *15*, 101–110.

Weintrop, D., Wise Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessing computational thinking: an overview of the field. *Computer Science Education*, *31*(2), 113–116. https://doi.org/10.1080/08993408.2021.191838