Understanding Students' Computational Thinking through Cognitive Interviews: A Learning Trajectory-based Analysis

Feiya Luo
University of Florida
College of Education
Gainesville, FL
sophia.f.luo@Gmail.com

Maya Israel
University of Florida
College of Education
Gainesville, FL
misrael@coe.ufl.edu

Ruohan Liu
University of Florida
College of Education
Gainesville, FL
liuruohan1990@ufl.edu

Wei Yan
University of Florida
College of Education
Gainesville, FL
wei1993yan@ufl.edu

Brian Gane
University of Illinois at Chicago
Learning Sciences Research Institute
Chicago, IL
bgane@uic.edu

John Hampton
University of Florida
College of Education
Gainesville, FL
ufxero@ufl.edu

ABSTRACT

For K-8 computer science (CS) education to continue to expand, it is essential that we understand how students develop and demonstrate computational thinking (CT). One approach to gaining this insight is by having students articulate their understanding of CT through cognitive interviews. This study presents findings of a cognitive interview study with 13 fourth-grade students (who had previously engaged in integrated CT and mathematics instruction) working on CT assessment items. The items assessed four CT concepts: sequence, repetition, conditionals, and decomposition. This study analyzed students' articulated understanding of the four CT concepts and the correspondence between that understanding and hypothesized learning trajectories (LTs). We found that 1) all students articulated an understanding of sequence that matched the intermediate level of the Sequence LT; 2) a majority of students' responses demonstrated the level of understanding that the repetition and decomposition items were designed to solicit (8 of 9 responses were correct for repetition and 4 of 6 were correct for decomposition); and 3) less than half of students' responses articulated an understanding of conditionals that was intended by the items (4 of 9 responses were correct). The results also suggested questioning the directional relationships of two statements in the existing Conditionals LT. For example, unlike the LT, this study revealed that students could understand “A conditional connects a condition to an outcome” before “A condition is something that can be true or false.”

CCS CONCEPTS

• **Social and professional topics → Computing education**; **Computational thinking**; **K-12 education**; *Student assessment*.

KEYWORDS

Computational thinking, Learning trajectories, K-8 Computer science education

ACM Reference Format:
Feiya Luo, Maya Israel, Ruohan Liu, Wei Yan, Brian Gane, and John Hampton. 2020. Understanding Students' Computational Thinking through Cognitive Interviews: A Learning Trajectory-based Analysis. In *The 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20), March 11–14, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3328778.3366845

1 INTRODUCTION

At the elementary school level, computer science (CS) is often presented as integrated computational thinking (CT) within the context of other academic disciplines, such as mathematics and science [11, 20, 21]. Reasons for this approach include both: (a) capitalizing on the affordances of authentic, applied CT within the disciplines, and (b) increasing equity by providing CS and CT within subject areas that are taught to all children as compared to electives or enrichment activities (e.g., [22]). As this method of teaching CS and CT expands, there is a growing need to better examine how students demonstrate the understanding of CT within these integrated experiences.

To add complexity, because elementary CT is still a new instructional area, there are few existing, tested learning progressions for what and how students should learn CT. The closest approximations that can be used for such work are learning trajectories (LTs) developed from extrapolating implicit learning goals from the literature (e.g., [18, 19]), the K-12 CS Framework [4], and resulting Computer Science Teachers Association (CSTA) standards [5]. While these perspectives provide potential standards and expectations that frame K-12 CS/CT instruction, they were developed based on common understanding rather than tested learning progressions.

This study used the CT definition provided by the K-12 CS Framework, which referred CT to be “the thought processes involved in expressing solutions as computational steps or algorithms that can be carried out by a computer” [4]. The LTs further helped the research team dissect CT into the four concept areas, sequence, repetition, conditionals, and decomposition. Therefore, this study

ACM ISBN 978-1-4503-6793-6/20/03…$15.00
https://doi.org/10.1145/3328778.3366845

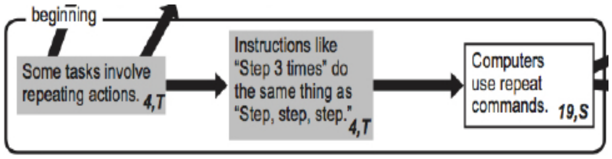

Figure 1: A snippet of the repetition learning trajectory. Adapted from Rich et. al. [19].

defined CT to be students' thought processes while expressing computational solutions in each of the aforementioned four concept areas.

Now that these LTs as well as the K-12 CS Framework and CSTA standards have been developed, a necessary next step is to examine how students demonstrate understanding of CT within these frameworks. This work will help teachers, curriculum developers, and tool designers understand what students should know as they progress from grade to grade, how the CT concepts intersect as students gain expertise, and the types of understandings and misconceptions students experience during elementary CS and CT instruction.

The purpose of this study, therefore, was to use the cognitive interview technique [6, 7] to qualitatively examine elementary students' CT understandings, leveraging the CT learning trajectories proposed by Rich et. al. [18, 19]. Given the wide range of variability among learners in elementary schools and the limited theoretical and empirical evidence of students' conceptual understanding of CT, this study focused on the range of understandings in one grade level rather than across grade levels. Inspired by Maloney, Confrey, and Nguyen's [12] work on learning trajectories, the research questions that guided this study were:

(1) How do 4th-grade students express and articulate CT understanding, in the areas of sequence, repetition, decomposition, and conditionals?
(2) How does students' CT understanding correspond to the hypothetical cognitive progression of the learning trajectories?

2 RELATED WORK

2.1 Theoretical Basis

An established construct in mathematics research and practice, LT (learning trajectory) refers to a hypothetical collection of landmarks students meet as they progress toward increasing sophistication during learning [2, 12]. An LT is hypothetical in nature because such progression is often conceptualized and created by working with a particular, usually small, groups of students. Since students' prior knowledge and engagement in learning cannot always be known in advance, an LT can only provide an a priori trajectory as to how learning may progress in a different group [2]. To study students' mathematical learning and skill acquisition, Clements and Sarama [3] dissected early math education into different topics and concepts such as quantity, counting, arithmetic, and spatial thinking and proposed a learning trajectory for each of the concepts.

Recently, the LT approach has been adopted by CS and CT researchers as a general way to gain insight into CT/CS teaching and learning. This study was guided and informed by the K-8 computational thinking LTs for four concepts: sequencing, repetition, conditionals, and decomposition [18, 19]. These computational thinking LTs are also hypothetical in nature as they were designed by examining literature not intended to be used for the specific aim of developing LTs [18, 19]. The four LTs are road maps with arrows that connect multiple consensus goals (CGs) placed in boxes that are either gray (indicating offline or unplugged goals) or white (computer-based goals). The CGs are the big ideas extracted and synthesized by Rich et al. [18, 19] from learning goals described in previous research studies.

For example (Figure 1), one CG of the repetition LT, "Some tasks involve repeating actions," is visually displayed in the LT as a gray box because it can be taught and demonstrated without the use of a computer, whereas "Computers use repeat commands" is visually displayed in a white box because of the specific reference to computer programs. By and large, the LTs provide a possible, but not linear, learning progression that students may traverse from their initial understanding of a topic to more sophisticated understanding [3].

For the sequence, repetition, and conditionals LTs, Rich et al. [19] categorized the CGs into beginning, intermediate, and advanced levels, indicating the increasing sophistication of a computing concept. In addition, the black or gray arrow between two CGs indicate "understanding of the source box is necessary" and "understanding of the source box is helpful, but not necessary" [19], respectively. For example, the conditionals CG "A conditional connects to a condition to an outcome" requires the understanding of two source CGs connected with black arrows, "A condition is something that can be true or false" and "Actions often result from specific causes." The decomposition LT does not specify any leveling of sophistication and the arrows show literature-supported connections instead of any prerequisite relationships between the CGs.

2.2 Computational Thinking Assessment

Design. Different from existing CT assessments such as Gonzalez's [9], the fourth-grade CT assessment items used in this study were conceptualized and designed by systematically mapping to the core ideas in the LTs. The items are intended to assess elementary students' understanding in five different CT domains: sequencing, repetition, conditionals, decomposition, and variables[8]. The assessment items were designed using evidence-centered design (ECD), an assessment design process that focuses on designing items that will elicit evidence of students' CT understanding. ECD helps one structure an assessment argument in which interpretive claims are specified (claims about what students know and can do) and the evidence for those claims (in terms of features of student responses that would provide support for the claims) are specified in tandem with designing the assessment items and scoring criteria/rubrics [15, 16]. In particular, different design patterns [14] were developed to structure the design of assessment items in each CT domain (e.g., sequencing, etc.). The set of assessment items include multiple response formats: true-or-false, multiple-choice, open-ended, and fill-in-the-blank questions. Additionally, some items use screenshots of the Scratch interface or specific Scratch bl ocks. For example, an item on conditionals asks for students

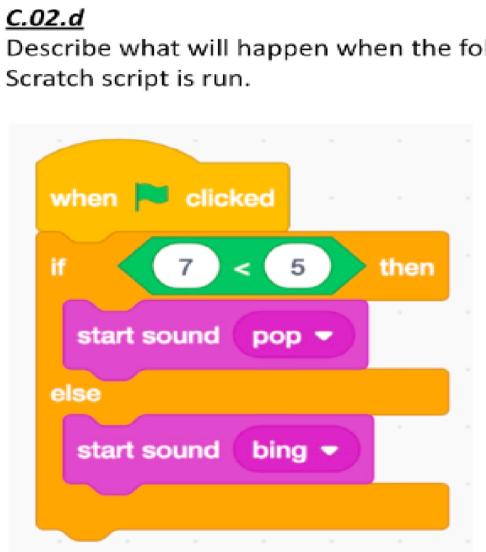

Figure 2: Sample conditional item

to describe the result of a program that involves conditional logic (Figure 2). In total, 11 items were sampled from the assessment item pool for the purpose of this study:

- Sequence: 3 items
- Repetition: 3 items
- Conditionals: 3 items
- Decomposition: 2 items

Although we had assessment items related to variables, there was no Variables LT in the literature. Therefore, we did not discuss the variables items in this paper.

Validity. Validation is a process of argumentation, in which one makes claims about the intended interpretive uses and evidence for those uses is collected and interpreted [1, 10]. The evidence can be multi-faceted, as different evidence is needed for different interpretive uses, and multiple pieces of evidence can provide stronger support for the assessment argument [1]. While the research team continues to marshal validity evidence of using these assessment items to understanding fourth-grade students' CT, existing evidence comes from following the principled design process and from conducting an internal review process prior to piloting the assessment items. An additional source of evidence comes from the analysis of student response data, in which statistical models are fit to the data [17]. After conducting preliminary analysis on data from multiple classrooms in which students completed the assessment items, it was found that the items range in difficulty, and are appropriate for measuring students with a range of CT abilities [8]. Further, the cognitive interviews reported in this paper will provide additional evidence for the cognitive aspect of validity [17].

3 METHODS

3.1 CT Instruction

This study used an integrated math-CT curriculum ("Action Fractions") aligned with the Common Core State Standards for Math (CCSS-M). The lessons provided fractions instruction for 4th-grade students. CT concepts such as sequencing, repetition, decomposition, conditionals, and variables are interwoven into the fractions lessons. The lessons include discussion prompts, reflections, unplugged activities, and hands-on coding exercises where students use the Scratch programming platform to build projects while building CT skills. Four of these lessons were taught to fourth-grade students prior to data collection to introduce the students to the CT content over the course of a month.

3.2 Data Collection

Participants. This study involved 13 fourth-grade students (four girls and nine boys) from a southeastern elementary school, whose student demographics reflect that of the state. Among the 13 students, there were eight white, two Hispanic, one African American, one Asian American, and one mixed ethnicity (as identified by their community learning leader).

Cognitive interview. Given the qualitative nature of this study, a small student sample size allowed the research team to interview each student individually and obtain information about their thinking while completing the CT assessment tasks. Cognitive interviews utilized a think-aloud protocol wherein the students were encouraged to verbalize their thinking as they addressed each assessment item. Each participant was presented with three assessment items that were randomly selected from the 11 sampled items. Each interview started with the interviewer's brief introduction of what the participant was expected to do and the interviewer modeling what "thinking aloud" looked like. A total of 31 responses were collected and analyzed.

3.3 Data Analysis

The data analysis was operationalized by adapting Maxwell's [13] categorizing strategies. The categorizing strategy to use the CGs [19][18] in each of the LTs as a priori codes to categorize students' think-aloud verbalization. For example, the sequence LT has a total of 10 CGs, four in the beginning level, four in the intermediate level, and two in the advanced level. All 10 CGs were used verbatim as code categories. Therefore, in this paper, the content in a "CG" and a "code category" was identical, only that "CG" was used as a general reference to the LTs, as "code category" was to data analysis.

The data analysis followed the following steps:

- For each excerpt of data, the participants'response to an item was examined and the code categories that best fit the student's understanding of the corresponding computing concept was assigned.
- Then, to understand how students' articulated CT understanding corroborate the hypothetical cognitive progression of the LTs, the researchers traced backward the connecting black arrows to see if the source CGs were manifested in the data. Recall that, theorized by Rich et. al. [19], the black arrows attached to a CG show what prior understanding is necessary.
- Lastly, to make sure that the assigned codes were accurate reflections of participants' verbalization, the destination CG connected to the assigned code was examined to see if that

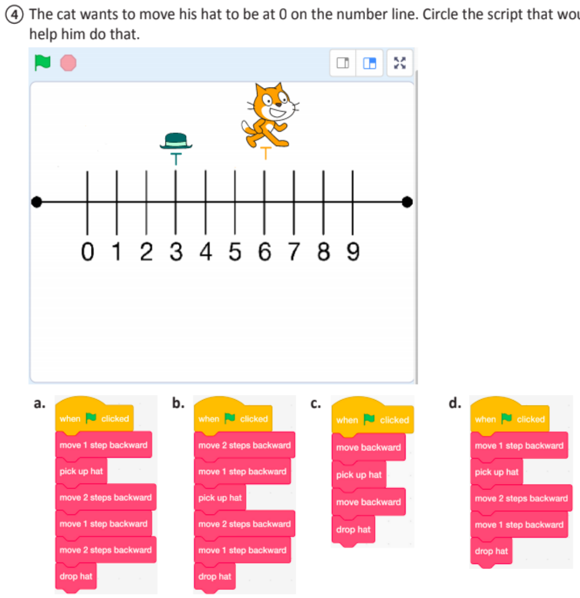

Figure 3: Sample sequence item

was also manifested in the data. If yes, that code category (the destination CG) was added, otherwise, the process stopped.

Note that since Rich and colleagues [18] used arrows in the decomposition LT differently than in the other LTs, the "trace-down" was not applicable when analyzing decomposition item responses. Rather, the adjacent CGs of the assigned code category were examined in a collective manner. During the data analysis process described above, the research team documented any agreement or disagreement between the theorized trajectory and what students actually demonstrated and articulated. Two researchers analyzed students' verbal data by engaging in collaborative coding and by rigorously discussing the codes for each excerpt of data. An inter-coder reliability with at least 80% of coding agreement was established.

Here is an example of what the coding of a participant's verbalization entails. When solving a sequence item (Figure 3), participant 11 traversed the four Scratch program choices and selected the answer with the correct sequence of code instructions. In this case, the code category, "The order in which instructions are carried out can affect the outcome," was initially assigned, suggesting that, in answering this item, the participant understood that the order of instructions could affect the outcome. Later, since there was no black arrow connected to this CG, no tracing back was needed. The only relevant CG was the one in the destination box (also in the intermediate level), "Computers have a default order of execution, so order matters in programming." It was decided that this aforementioned code was also relevant, given that the item was set in the context of coding in Scratch and the participant was able to mentally run the programs to select the correct answer. Therefore, in conclusion, this response demonstrated an intermediate level of understanding of sequencing.

After coding all participant verbal responses, data was aggregated by LT and students' understanding of each of the four concepts was presented in the Discussions and Results section below.

4 RESULTS AND DISCUSSION

The following section presented aggregated results by LT.

Sequence. Overall, participants' responses demonstrated that they understood that "Different sets of instructions can produce the same outcome,""The order in which instructions are carried out can affect the outcome," and that "Computers have a default order of execution, so order matters in programming." For example, when asked to provide two different ways for Aisha to carry 8 toys from the kitchen to the bedroom with a maximum of 3 toys each time, two students gave the following answer:

```
P1: "So, I did two, two, two, and two. And then, I
gotta thinking [think] another way. You can do one
at [for] eight times."
P2: "...you can do, she carries 2 toys and repeat[s] it
4 times, right? ... So he can also carry 1 at a time,
and then repeat it 8 times."
```

Another sequence item (Figure 4), which asked students to write instructions to move along a specific path, revealed interesting results. Participant 3 provided correctly ordered instructions "[go] right and up, then right." However, this answer lacked precision and completeness (i.e. to move how far right or up), which were two core ideas of the beginning-level CGs of the sequence LT. This students' answer meant that the participant had intermediate-level understanding of sequence ("The order in which instructions are carried out can affect the outcome") but not the beginning-level understanding ("Precision and completeness are important when writing instructions in advance"). While it may have seemed anti-intuitive for a student to have intermediate understanding without beginning understanding, it was indeed consistent with how the sequence LT was mapped. In the sequence LT, most beginning-level CGs and the intermediate CGs are parallel and are not connected by any arrows, indicating that, by theory, most intermediate CGs do not necessarily require an understanding of the beginning CGs and that learners can reach the intermediate level (understanding about order) without necessarily mastering the beginning level (understanding about precision and completeness of instructions). The only intermediate CG that requires prior understanding is, "Creating working programs requires considering both appropriate commands and their order," which is computer-based. However, given that none of the items sampled for sequence asked students to create a computer program, discussion about this CG is beyond the scope of this paper.

Repetition. Unlike the sequence items, where all students answering an item demonstrated the same level of understanding, the repetition items received responses with varying levels of understanding. More specifically, six of nine responses had intermediate-level understanding that "Repeating things can have a cumulative effect," two had a beginning-level understanding ("Instructions like 'Step 3 times' do the same thing as 'Step, step, step'"), and two did not have any understanding of repetition in the context of coding (programming).

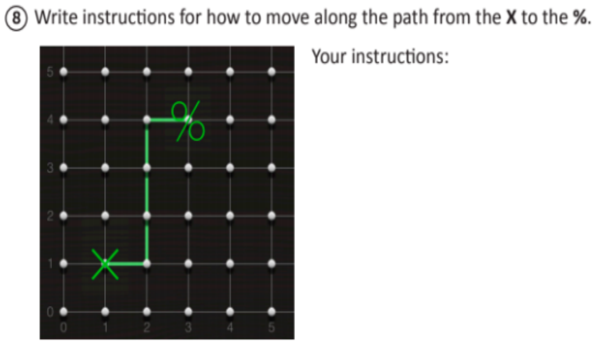

Figure 4: Sample sequence item 2

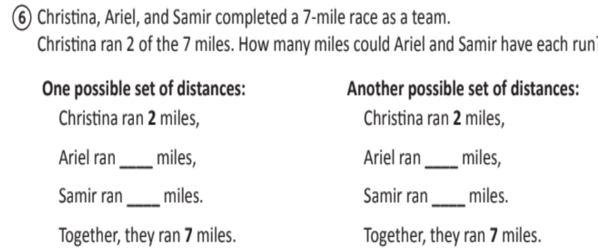

Figure 5: Sample decomposition item

For example, one repetition item asked students to decide the end result of both the "move 3 steps" and "sound meow" actions repeating five times. Students' responded "It takes 3 steps and meows 5 times" (P11) and "... the cat will take 3 steps and [meow] 5 times" (P2) while the answer should have been "move 15 steps and sound meow 5 times." The responses showed that both students had intermediate-level understanding, that "Repeating things can have a cumulative effect." However, that understanding was incomplete given that the responses were only partially correct.

Interestingly, when answering the item which asked students to "Circle the script that will make the computer count up by 4 from 0 to 12, and say the results each time (4,8,12)," participant 6 asked the researcher what the "repeat" and the "say" Scratch blocks meant. After getting an explanation, the participant was able to select the correct script out of the four choices given. It was decided that this participant was at the intermediate level ("Repeating things can have a cumulative effect") even though he did not initially understand the Scratch blocks. The reason was that the aforementioned CG was an offline or unplugged CG (in a gray box). This example showed how students could understand repetition and its cumulative effect outside the context of coding. Also, depending on what the purpose of the item is, an item outside the context of coding may be more appropriate than one specifically set in the coding context in soliciting students' understanding of repetition.

Conditionals. In this trajectory, items involved using the "if-else" conditional statements, both in the unplugged, everyday context and the coding context. The unplugged problem asked students to complete an "If...then" sentence by adding a condition or an action. For example, "a. If ______, then put on a jacket. b. If I see a spider, then ______." Four participants responding to this item all provided a reasonable condition or an action, such as, "If I see a spider, then run away" and "if it were cold, I would put on a jacket." Student responses showed that students understood "A conditional connects a condition to an outcome," which was a beginning-level CG.

The two other conditionals items were similar. One item in unplugged form asks for the end result of the code "If 5 < 8 then say 'pop' Else say 'bing'" and the other in Scratch script asked for the end result of the program "If 7<5, then start sound pop; Else start sound bing" (Figure 2). Data analysis showed that three responses demonstrated an understanding of "A conditional connects a condition to an outcome." However, no evidence of student understanding of the source CG "A condition is something that can be true or false" was found.

Here the response from participant 5 was used as an illustration. When solving the unplugged item, he gave the rationale saying,

```
"...you'd have to write another fraction that's
not 5<8 [for it to say 'bing']. So, for it to say
'else,' or for it to say this ['bing'], it has
to say another fraction like 4<8 or 3<8...For it
to say 'pop,' you need this fraction [5<8]..."
```

This answer suggested that while he knew that a condition was connected to an outcome, he did not understand that he needed to evaluate the number comparison statement (5<8) in order to come to the corresponding action. In fact, the other three respondents were in a similar blind spot, not knowing that the outcome action depended on the state of the condition (true or false). Therefore, it is suggested that the understanding that there are two states of a condition does not come intuitively to children and needs to be specifically taught before they know how to evaluate and connect the condition to its corresponding outcome. In fact, it is suggested that in the Conditionals LT, "A condition is something that can be true or false" should be the target, instead of the source, of "A conditional connects a condition to an outcome."

Decomposition. One of the items asked students to decompose a number (Figure 5) and the other asked students to decompose a problem wherein they needed to find the area of the backward L-shape outlined in white (Figure 6). In the case of the number decomposition, two responses were correct (distances added up to 7), both demonstrating a beginning-level understanding of decomposition, that "Systems are made up of smaller parts" ("system" refers to any object, concrete or abstract, that can be broken into parts). Participant 8, in his response, did not provide an answer for the distances to add up to 7 and confessed that he would need to at least know "how much Ariel ran" before being able to solve the problem. For the second decomposition item, two students responded that they "counted 14 squares" in the shape while the item instruction specifically prompted for the multiple steps involved to find the area of the outlined L-shape. However, it was nonetheless decided that the responses demonstrated an understanding of decomposition because the research team agreed that the 14 squares could be considered "smaller parts" of the "system"–the outlined L-shape.

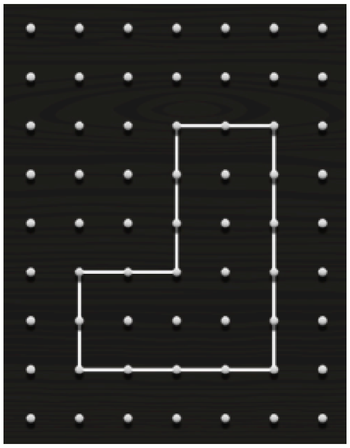

Figure 6: Sample decomposition item 2

In total, four responses demonstrated beginning-level understanding of decomposition, that "System are made up of smaller parts," and two did not demonstrate any understanding of decomposition.

5 CONCLUSION AND LIMITATIONS

This paper presented a qualitative account of how students demonstrated and articulated various levels of understanding of the four CT concepts: sequence, repetition, conditionals, and decomposition. The paper also discussed how CT understanding corresponded to the learning progressions within LTs proposed by Rich and colleagues [18, 19]. Data from the cognitive interviews revealed a close alignment between the fourth-grade students' CT understanding and the beginning and intermediate levels of the computational thinking LTs. This finding points to the utility of the LTs in ascertaining students' CT understanding. Additionally, the LTs were not grade specific. Rather, they proposed a general progression of CT understanding. Therefore, if a student shows misconceptions matched to a specific consensus goal, the LTs can be used to approximate potential activities that address those misconceptions.

One incongruity between the data from this study and the LTs was the directional relationships of two statements in the existing Conditionals LT. This finding was unsurprising as students' learning progression may not always be anticipated and may not be linear in nature. It is possible, for some students, that they may understand that "a conditional connects a condition to an outcome" even if they have not had instructional experiences to understand that "a condition is something that can be true or false." This finding also points to Clements and Sarama's [2] explanation that we cannot always know students' prior knowledge and experience; hence, we need to consider LTs in a flexible manner that accounts for learner variability. Future research should therefore delve into the incongruity discovered in this study to further understand whether the LT should be modified or whether the students in this sample had different experiences that would influence their understanding of conditionals.

As the research in computational thinking LTs is only emerging, this study presented several limitations. Firstly, the results of this study are highly sample-dependent. The participants sampled were from an elementary school where systematic CT instruction was not implemented at the school level. The results described only the likely landmarks these specific participants encountered after going through four CT integrated lessons. However, given that most schools do not have systematic CT instruction, the results of this study may, in fact, be representative of the scenarios in which CT is only beginning to be taught. Nonetheless, given the demographics and backgrounds of the participants, interpretations should be made with caution and future research is encouraged to include a more inclusive sample. Secondly, acknowledging that the LTs do not and should not dictate how every student accumulates increasingly complex knowledge and skills, this study presented how different learners may demonstrate CT understanding while solving assessment items. Lastly, the assessment items sampled for the purpose of this study cover only a specific portion of the knowledge for elementary CT. Therefore, the items may have limited what understanding a participant could have demonstrated in relation to the LTs. It was possible that participants had more advanced CT knowledge that was not captured due to the way items were sampled. Future research is encouraged to examine students' understanding with items that assess more advanced CT knowledge. Also, more empirical evidence is needed to help the field understand how these LTs can be used to guide CT instruction and learning and help teachers prepare instructional activities and set learning goals.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under project No. 1932920. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for educational and psychological testing.* Amer Educational Research Assn.

[2] Douglas H Clements and Julie Sarama. 2004. Learning trajectories in mathematics education. *Mathematical thinking and learning* 6, 2 (2004), 81–89.

[3] Douglas H Clements and Julie Sarama. 2014. *Learning and teaching early math: The learning trajectories approach.* Routledge.

[4] K-12 Computer Science Framework Steering Committee et al. 2016. K-12 computer science framework. (2016).

[5] CSTA. 2012. Computer science standards. *Computer Science Teachers Association* (2012).

[6] Dana Denick and Ruth Streveler. 2014. Qualitative analyses of students'conceptual reasoning. In *JW Pellegrino (Chair), Evaluating and Improving Concept Inventories as Assessment Resources in STEM Teaching and Learning. Symposium conducted at the meeting of the American Educational Research Association, Philadelphia, PA.*

[7] K Anders Ericsson and Herbert A Simon. 1980. Verbal reports as data. *Psychological review* 87, 3 (1980), 215.

[8] Brian Douglas Gane, Noor Elagha, Feiya Luo, Ruohan Liu, Wei Yan, Carla Strickland, Diana Franklin, Kathryn Rich, James W. Pellegrino, and Maya Israel. 2020. Developing Computational Thinking Assessments from Learning Trajectories: Design Approach and Preliminary Validity Evidence. *Annual Conference of Amer Educational Research Assn* (2020).

[9] Marcos Román González. 2015. Computational thinking test: Design guidelines and content validation. In *Proceedings of EDULEARN15 conference.* 2436–2444.

[10] Michael T Kane. 1992. An argument-based approach to validity. *Psychological bulletin* 112, 3 (1992), 527.

[11] Feiya Luo, Pasha Antonenko, and E Christine Davis. 2019. Exploring the evolution of two girls' conceptions and practices in computational thinking in science. *Computers & Education* (2019), 103759.
[12] Alan P Maloney, Jere Confrey, and Kenny H Nguyen. 2014. *Learning over time: Learning trajectories in mathematics education*. IAP.
[13] Joseph A Maxwell. 2012. *Qualitative research design: An interactive approach*. Vol. 41. Sage publications.
[14] Robert J Mislevy and Geneva D Haertel. 2006. Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice* 25, 4 (2006), 6–20.
[15] Robert J Mislevy, Linda S Steinberg, and Russell G Almond. 2003. Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives* 1, 1 (2003), 3–62.
[16] JW Pellegrino, LV DiBello, and SP Brophy. 2014. The science and design of assessment in engineering education. *Cambridge handbook of engineering education research* (2014), 571–598.
[17] James W Pellegrino, Louis V DiBello, and Susan R Goldman. 2016. A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist* 51, 1 (2016), 59–81.
[18] Kathryn M Rich, T Andrew Binkowski, Carla Strickland, and Diana Franklin. 2018. Decomposition: A K-8 Computational Thinking Learning Trajectory. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*. ACM, 124–132.
[19] Kathryn M Rich, Carla Strickland, T Andrew Binkowski, Cheryl Moran, and Diana Franklin. 2017. K-8 learning trajectories derived from research literature: Sequence, repetition, conditionals. In *Proceedings of the 2017 ACM conference on international computing education research*. ACM, 182–190.
[20] Kathryn M Rich, Aman Yadav, and Christina V Schwarz. 2019. Computational Thinking, Mathematics, and Science: Elementary Teachers'Perspectives on Integration. *Journal of Technology and Teacher Education* 27, 2 (2019), 165–205.
[21] José-Manuel Sáez-López, Marcos Román-González, and Esteban Vázquez-Cano. 2016. Visual programming languages integrated across the curriculum in elementary school: A two year case study using "Scratch" in five schools. *Computers & Education* 97 (2016), 129–141.
[22] David Weintrop, Elham Beheshti, Michael Horn, Kai Orton, Kemi Jona, Laura Trouille, and Uri Wilensky. 2016. Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology* 25, 1 (2016), 127–147.