BootUp PD 2022-2023

Student Knowledge and Perceptions of Coding

Evaluation conducted by

Peter Jacob Rich, PhD

Table of Contents

Executive Summary

Over 100,000 students were affected by BootUp's efforts in 2022-23. This study reports on the efforts of a handful of schools that collected student data to create a baseline for understanding elementary students' cognitive abilities and attitudes toward coding in BootUp classrooms. A high-level summary is provided in the following summary:

1. **Participation**
    a. Three districts administered the Computational Thinking test (CTt), while four districts gave the Elementary Student Coding Attitude Survey (ESCAS).
    b. Overall, 737 students completed the CTt (341 girls, 381 boys, 15 other). Most responses came from 3rd and 4th graders. Nearly two thirds (65.9%) of students identified as Black or African American. Another 17% identified as Hispanic, with only 8.8% of students identifying as White. Thus, the population for the ESCAS study is one made up in large part of students who are traditionally under-represented in computing.
    c. In total, 589 students completed the ESCAS (273 boys, 311 girls). Responses came from 237 3rd graders, 185 4th graders, and 164 5th graders.
    d. The majority of students have less than a year of coding experience, though just under a third (29.2%) reported 1-2 years' experience with coding.
2. **Performance**
    a. Students completed the CTt in under 30 minutes, with 4th graders finishing 5 minutes faster than 3rd graders. It took students about 12 minutes to complete the ESCAS.
    b. On average, 3rd grade students scored 9/28 points on the CTt. The average score tended to increase by 1 point per grade.
    c. On average, boys scored 1 point higher than girls in the same grade on the CTt. There was no noticeable difference between genders on the ESCAS.

d. There is a negative, but weak, correlation (r ~ -.232) between grade and attitude. This means that the older students are less positive toward coding than the younger students.

Recommendations

Inasmuch as the purpose of this study was to establish a baseline for student knowledge of and attitudes for coding/CT, the most important recommendation is to ensure that students at these schools:

(a) continue to participate in classes that integrate coding, and

(b) complete these same post assessments at the end of the 2024 school year. That data will enable us to see their growth.

Methods

Prior to the 2022 school year, no BootUp-aligned school districts were collecting either cognitive or affective data from students. In order to measure a longer-term effect, we first needed to establish a baseline for knowing where students are starting from. Thus, this year's data does not reveal the results of prolonged exposure to coding education, but rather provides the first point from which students will later be measured to indicate change. In this section, I first describe the measures used, the analysis methods, and the baseline findings.

Measures

We used the Computational Thinking test (CTt) and the Elementary Student Coding Attitude Survey (ESCAS) to measure students' cognitive abilities and affective leanings for coding. I describe each of these measures below so that the reader might have a better understanding of the types of questions students were asked.

The Computational Thinking test

The CTt was originally developed in Spain with the purpose of creating a test that can measure students' CT ability from elementary to high school (Román-González, et al., 2016). It was developed and tested on 5th-10th grade children and found to have good reliability. It scales well across grades, meaning that as students grow, they get progressively better at the test. The CTt has demonstrated convergent validity with other well-established problem-solving instruments, such as the Primary Mental Abilities (PMA) battery and the RP30 problem-solving test. While demonstrating convergence with other problem-solving tests, the CTt validation procedures also revealed that solving computational problems is different enough from general problem-solving that it is a related, but separate, type of thinking.

According to the authors, the CTt is highly focused on "computational concepts." Computational concepts tested include loops, conditionals and functions. These are tested in seven tetrads, totalling 28 questions overall. Each tetrad presents the student with a series of four related tasks, moving from easier to more difficult tasks. Computational tasks involve either sequencing, debugging, or completion. *Sequencing* tasks pose problems of ordering computational instructions. *Debugging* tasks focus on finding and resolving errors in a set of codes. And *Completion* tasks focus on adding in missing code. CTt problems do not require students to know a specific coding language in order to solve and can thus be completed even by students who have never coded previously. On average, it takes students about 30 minutes to complete the full 28-item test.

Figure 1

Sample item from the CTt

M. Román-González et al. / Computers in Human Behavior xxx (2016) 1–14

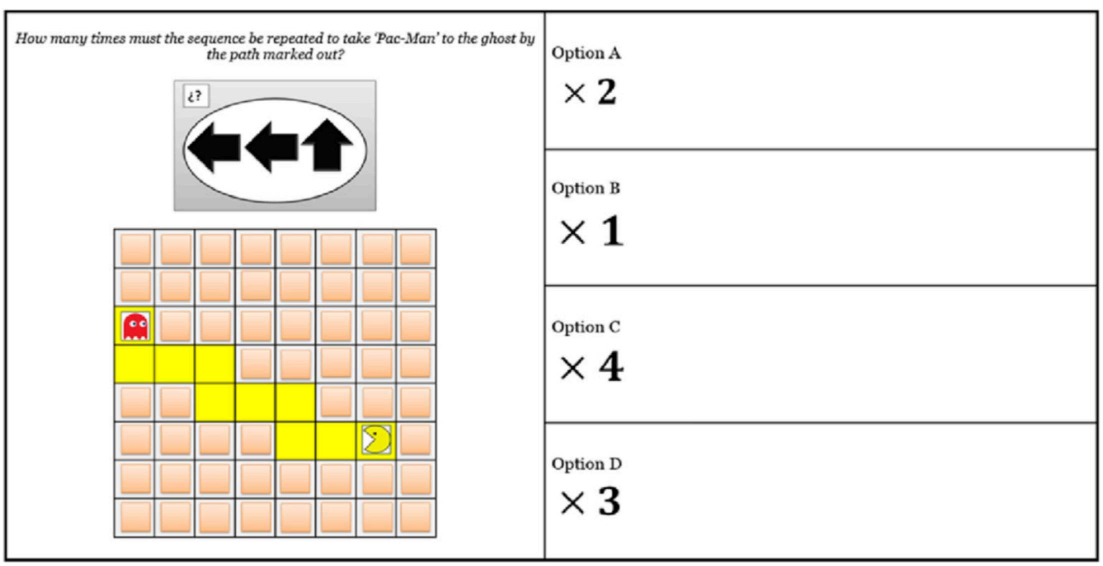


Fig. 1. CTt, item 6: loops–repeat times; 'The Maze'; visual arrows; no-nesting; completion.

The Elementary Student Coding Attitude Survey

While it's important to measure cognitive abilities affected by learning to code, it is equally important to measure affect, or attitudes. If students learn to like coding while young, they will be more likely to engage with coding as they grow older. Unfortunately, most coding attitude measures were either directed at older students (e.g., high school, college) or were too generic to be useful (e.g., "I like STEM"). Thus, Mason and Rich (2020) developed the ESCAS as a measure that could be used specifically with elementary-aged children. The ESCAS is written in simple language that elementary students can understand. It takes roughly 8-15 minutes to complete, on average.

The ESCAS consists of 23 Likert-type questions. Likert-type questions provide a statement and ask respondents the degree to which they agree or disagree with that statement. For example, for the statement, "Coding is Interesting" students can strongly disagree, disagree, somewhat disagree, somewhat agree, agree, or strongly agree. The ESCAS measures six key attitudes that have been shown to affect a student's likelihood to pursue coding. Namely: confidence, interest, usefulness, math confidence, perceptions of coders, and social value. *Confidence* measures their own confidence with coding. *Interest* measures how much they might like or be curious about coding. *Usefulness* gets at the practical aspects of coding, such as its value for employment, school or solving problems. *Math confidence* is a student's confidence with their math ability, which correlates with success in coding. *Perceptions of Coders* seeks to reveal a student's biases about what they believe a "person who codes" is like. Finally, "social value" reveals how a student believes their peers and parents view coding.

Procedures and Analysis Methods

In order to administer these surveys, we first sought Institutional Research Board (IRB) approval by each district. BootUp facilitators then worked with school teachers in districts that granted approval to distribute and collect parental permission forms. Only after a sufficient number of forms were collected were teachers allowed to administer the surveys. Thus, while the intent was that these would establish a baseline, different districts collected student data at different points of the year, depending on when they were able to secure all the appropriate permissions. Students did not provide their name or other personally-identifying information on these measures, but instead were asked to provide their school/lunch ID #. We will use this number in future studies to pair students and measure individual change over time.

Both the CTt and the ESCAS provide only quantitative data. To analyze this data, we used both descriptive and inferential statistics. Descriptive statistics provide a picture of the "shape" and nature of the data. It includes total participation counts, max/min scores, score ranges, etc. Much of the story of how students performed can be revealed with descriptive statistics. Inferential statistics, on the other hand, typically engage in comparisons. For example, we might use inferential statistics to determine the extent to which boys and girls or students from different grades performed on the test and whether or not these differences were due to natural variation or if there is likely a significant difference between groups.

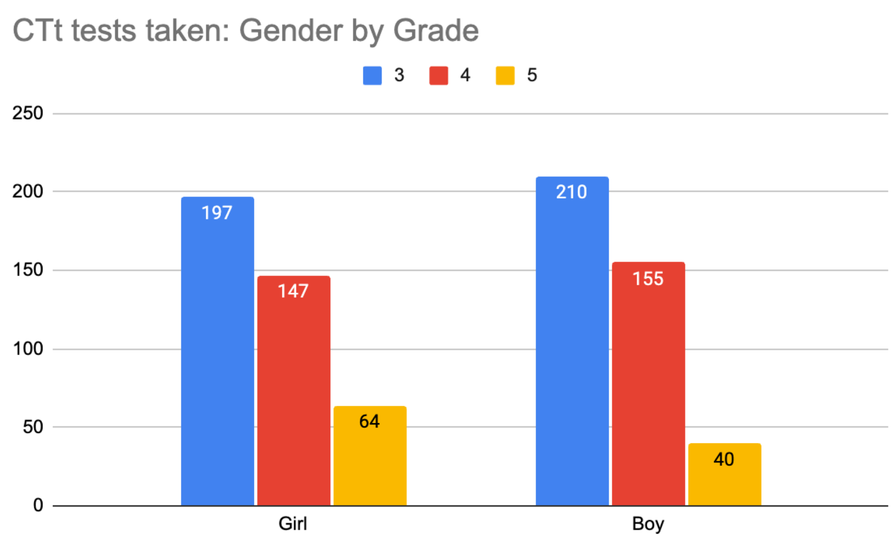
Findings

Three school districts, representing Georgia, Mississippi, and Oklahoma, completed the CTt. The same three districts administered the ESCAS, as well as a district from Alabama. I first present the findings from the CTt and then the ESCAS. This is followed by an analysis of correlations between these two measures.

CTt Findings

Overall, 933 students started the CTt. However, many of these did not attempt any questions or completed the exam fairly quickly. I contacted Dr. Román-González to ask what he thought the minimal amount of time needed to reasonably complete the CTt would be. He recommended not to consider any exams completed in fewer than 10 minutes, as those responses would likely only have been completed by guessing or marking answers at random. After eliminating such attempts, there were a total of 830 CTt exams, with students from 13 different schools (7 in Oklahoma, 5 in Georgia, and 1 in Mississippi). In total 410 usable responses came from girls, 406 came from boys and 14 students identified as 'other' (see Figure 2.1). There were 414 usable tests from 3rd graders, 314 from 4th graders, and 110 from 5th graders.

Figure 1



On average, students completed the CTt in about 28 minutes. There was a clear difference between grades, with 3rd and 5th graders taking an average of about 28 minutes and 4th graders each completing the test in 24 minutes. A little over 83% of students completed the exam.

CTt Performance Outcomes

Overall, scores were distributed normally, with a slight skew to the right (see Figure 2.2). Mean scores on the exam (see Table 1) showed about a .36-point increase by grade level. Third graders scored a 9.54, fourth graders scored 9.90 and fifth graders averaged 10.17 points. These scores are a few points lower than what has been observed by other studies that have administered the CTt to 3rd-5th grade students. The most common score across all grades and genders was 10/28 points. A t-test comparison of means revealed a statistically significant difference between overall boys' (10.148) and girls' (9.290) scores ($p < .002$). Research on the

CTt has consistently shown that boys tend to outscore girls by about 1 point. However, in this study there was one notable difference; 4th-grade girls performed on-par with 4th-grade boys. This may be more reflective of 4th grade boys under-performing, as 3rd-grade boys also outperformed both 4th-grade girls and boys. That is, one would normally expect 4th-grade boys to score in the 10-11 range. However, in this case, 4th grade boys averaged scores below 10 points.

Figure 2

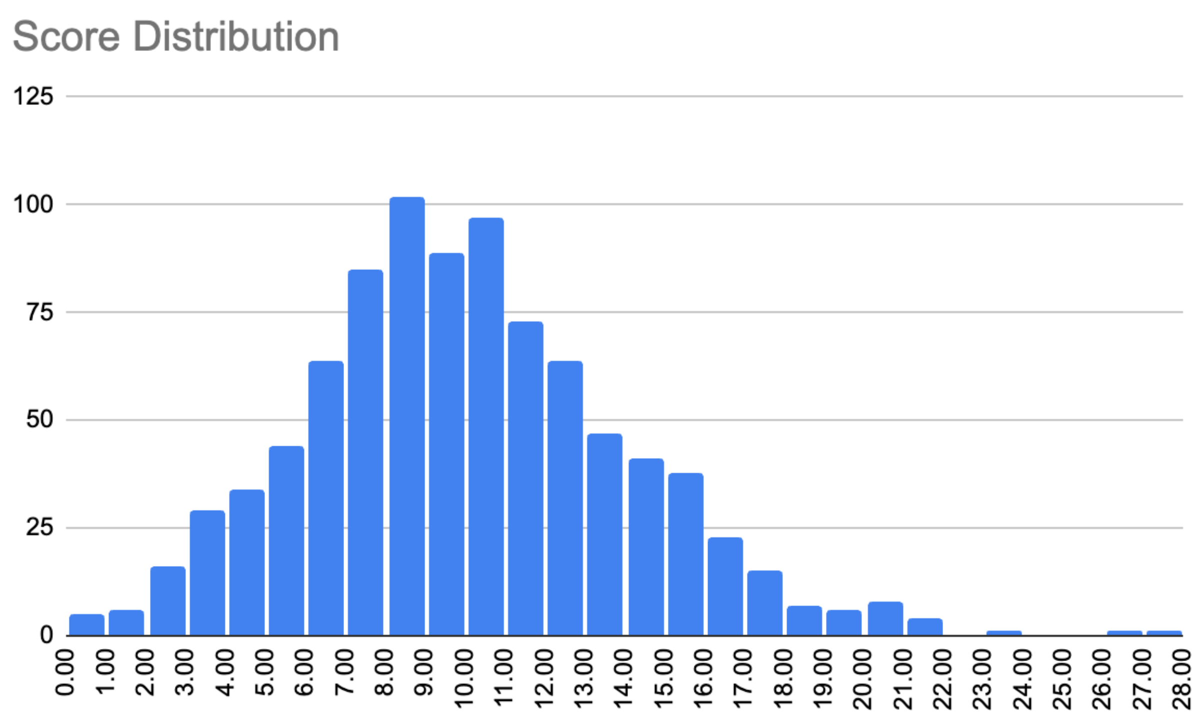


Table 1

CTt performance by grade and gender

	Overall	3rd		4th		5th	
Avg completion time	28.03 min	28.33 min		23.99 min		27.32 min	
Mean score	09.75	9.54		9.90		10.17	
Median score	09.00	9.00		10.00		9.00	
finished %	83.59	87.63		83.78		70.95	
		girl	boy	girl	boy	girl	boy
Mean score		8.84	10.13	9.82	9.89	9.50	11.18
Median score		9	10	10	10	9	10

CTt Performance Self-Assessment

In addition to coding problems, the CTt asks students to guess how many questions they thought they got right out of 28 as well as their facility with computers (on a 1-10 point scale). Estimation of performance has been shown to be an important indicator of students' ability to self-regulate their learning (Rebello, 2012).

Table 2

Students' self-estimations of score and comfort with computers

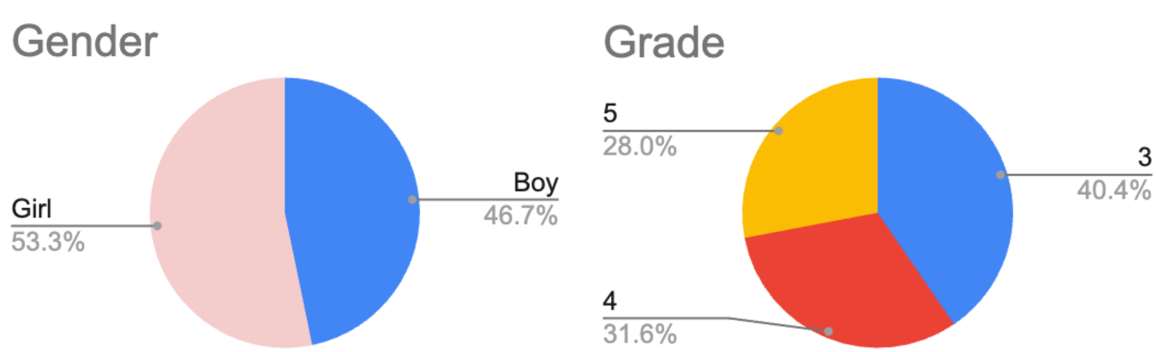
	Overall	3rd		4th		5th	
Estimated Correct # of 28	17.66	18.00		17.10		17.99	
Estimation vs. actual	4.85	6.01		4.27		2.16	
facility w/ computers	7.72	8.11		7.30		7.26	
		girl	boy	girl	boy	girl	boy
Estimated Correct # of 28		16.45	19.39	16.65	17.64	16.69	19.03
Estimation vs. actual		5.10	6.96	3.55	5.02	-0.11	5.48
facility w/ computers		7.81	8.42	7.03	7.49	6.75	7.71

On average, students overestimated their scores by nearly 5 points. Students improved their estimations by about 2 points for every grade they matured, meaning that older students made better estimates of their performance than younger students. In every grade, girls' estimations were closer to their actual scores than boys' estimations. The starkest difference occurred with 5th graders, where the difference between boys' and girls' estimations jumped to 5.5 points. In fact, the 64 5th grade girls who spent at least 10 minutes completing the CTt were surprisingly accurate on their estimations of their scores, actually underestimating slightly.

ESCAS Findings

Students in four districts (suburb-1, suburb-3, city-10, city-13) completed the ESCAS. Because this is a faster survey, we used completion as the criteria for inclusion in the analysis. While 649 students started the ESCAS, 589 completed it (311 girls, 273 boys). The survey was completed by 237 3rd graders, 185 4th grades, and 164 5th graders. On average, students completed the ESCAS in 12.22 minutes. Curiously, boys finished the survey 4 minutes faster than girls (10 min vs. 14 min). Older students finished the survey 1-1.5 minutes faster than those in the preceding grade.

Figure 3

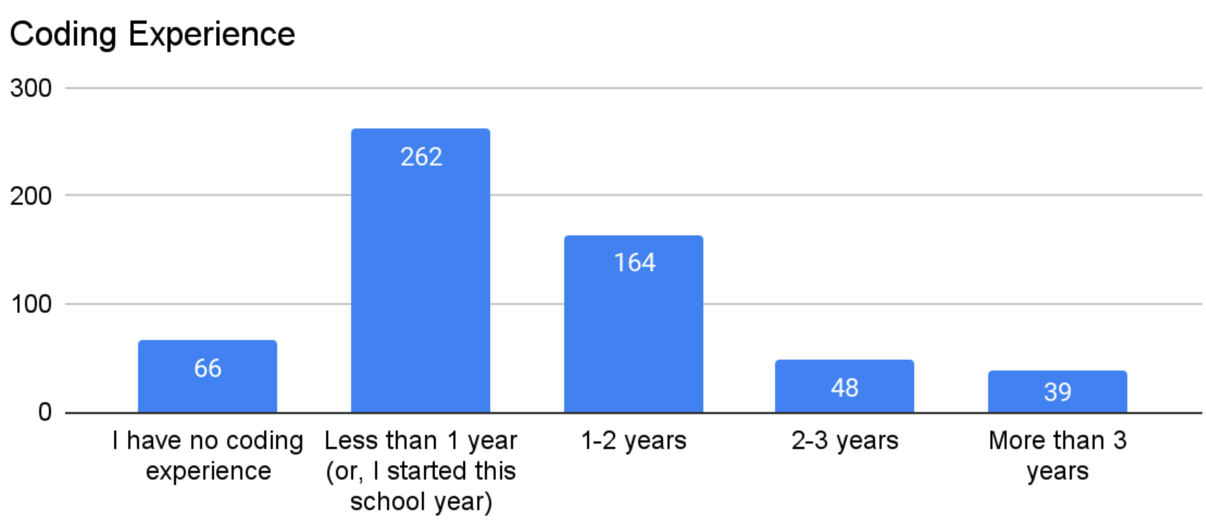


Students were asked to identify their race/ethnicity from a list of seven choices. They could select multiple races/ethnicities. Nearly two thirds (65.9%) of students identified as Black or African American. Another 17% identified as Hispanic, with only 8.8% of students identifying as White. Thus, the population for the ESCAS study is one made up in large part of students who are traditionally under-represented in computing.

Elementary Student Coding Experience

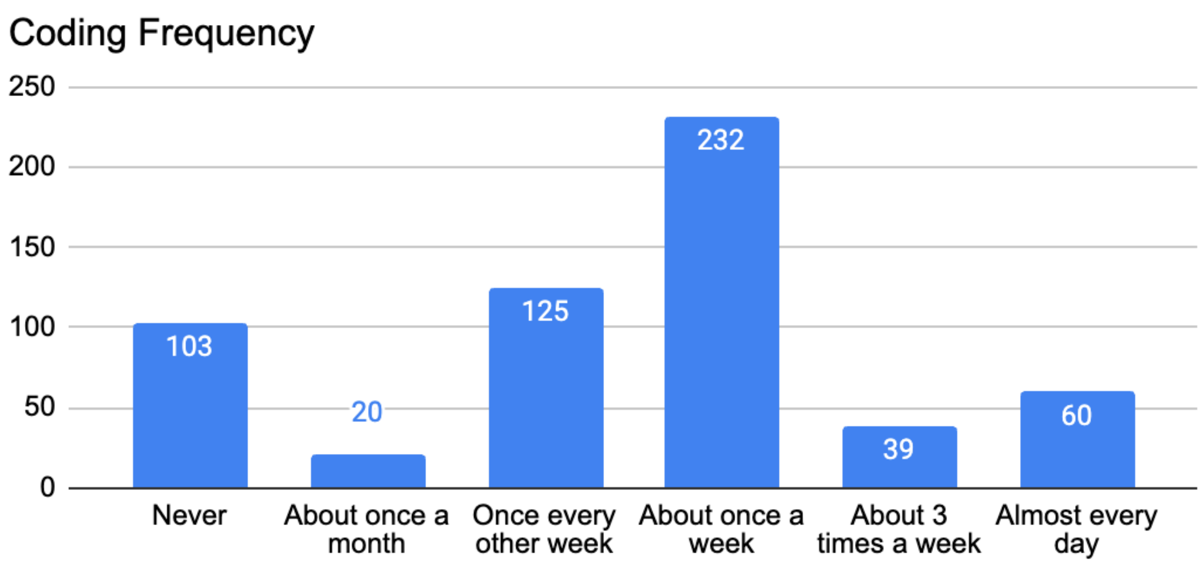
We asked students to indicate how much coding experience they had (see Figure 4), as well as how often they coded (see Figure 5). The majority of students (55.9%) indicated that they had less than a year of coding experience. Only 6.6% of students had more than 3 years' experience.

Figure 4



While students had little coding experience, the majority (60.7%) indicated that they were coding regularly. This means that, when they took the ESCAS, students were likely regularly engaging in coding lessons at school. About ⅕ (19.4%) of students were coding once a month or not at all.

Figure 5



Elementary Students' Attitudes for Coding

Students demonstrated slightly positive attitudes toward coding across most of the six measured attitudes. Their strongest attitude was actually not toward coding, but rather their confidence in their math abilities, which was rated ¾ of the way between "somewhat agree" and "agree" (see Table 3)

Table 3

Baseline Attitudes for Coding

ATTITUDE	score	count	boys	girls	3rd	4th	5th
Confidence	4.38	588	4.39	4.37	4.53	4.39	4.15
Interest	4.37	589	4.42	4.35	4.59	4.47	3.94
Utility	4.34	587	4.32	4.38	4.48	4.45	4.02
Perceptions	4.14	586	4.24	4.07	4.27	4.11	3.97
Math	4.75	585	4.82	4.71	4.91	4.70	4.56
Social Value	3.98	587	3.93	4.04	4.21	4.03	3.58

I used multiple regression to examine the predictive effect of demographics and experience on students' overall attitude score (i.e., the average of all their sub-scores). The model included grade, gender, coding experience, and coding frequency as predictors of students' attitudes. The model revealed a weak, but significant negative correlation between grade and attitude ($r = -.213$), but no significant effect of gender. In layman's terms, that means that older students in this population demonstrated slightly less confident attitudes toward coding than younger students (decreasing about .2 points per grade level) but that gender was not a significant factor between boys' and girls' attitudes. It is important to note that this grade-level effect is minimal, accounting for only 4% of the observed variance between grades.

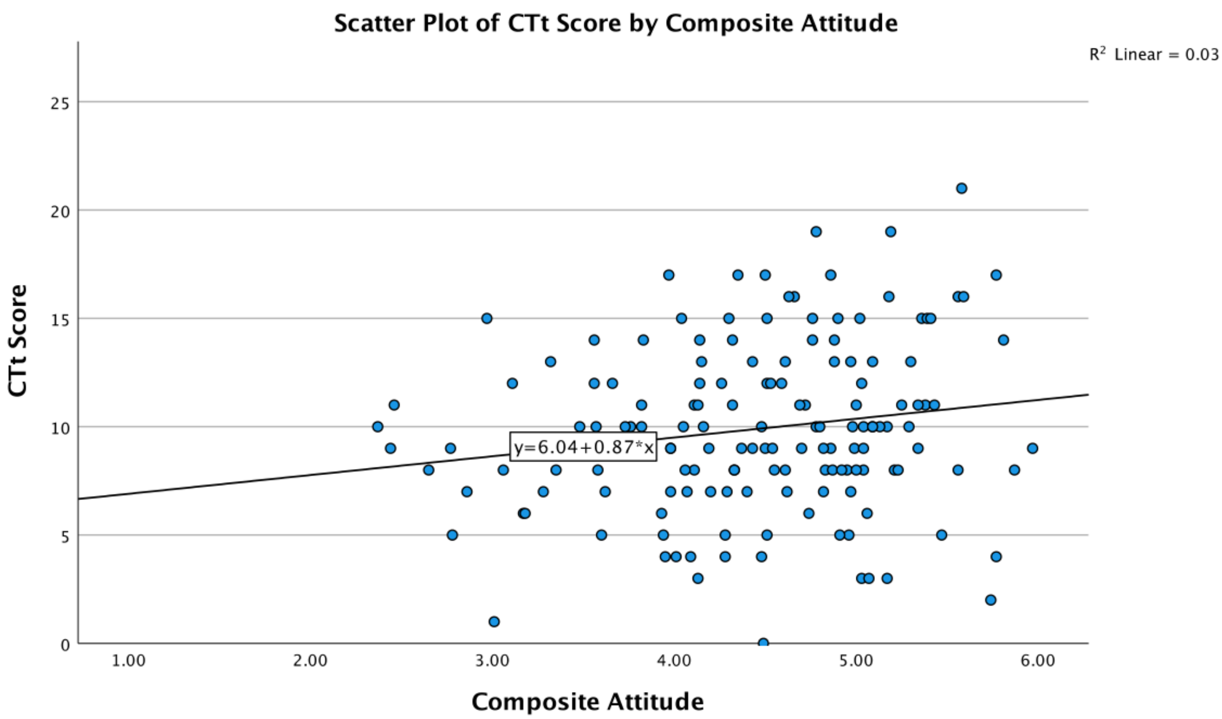
The model also revealed a weak, but significant, effect of coding frequency ($r = .207$, $p < .000$) on overall coding attitude. This means that, perhaps unsurprisingly, students who code more frequently have slightly more positive attitudes toward coding. That being said, this also only accounts for about 4% of the observed variance amongst students, meaning there are many other factors to consider. Taking into account all four predictor variables (i.e., grade, gender, coding experience and coding frequency) accounts for about 8% of the total variation in predicting a student's overall attitude toward coding.

It will be important to track these same students across years in future studies to see the extent to which their attitudes change. For example, will this same group of 3rd grade students carry over their more positive attitudes toward coding into their fourth grade year? If so, I would expect to see the observed effect of grade to decrease and disappear over two to three years. Furthermore, will that attitude become more positive with time as students more regularly engage with coding? We will be able to answer this question only if teachers in these same districts make sure to administer the ESCAS again at the end of the coming school years. Thus, it is vitally important that, at a minimum, these four districts continue to administer the CTt and ESCAS at the beginning and ending of the 2023-2024 school years.

Correlations between Cognitive and Affective Indicators

While the CTt and the ESCAS are separate measures, it might be helpful to know the extent to which they're correlated. In other words, do students who have more positive attitudes toward coding score higher on a cognitive test of computational thinking? Or vice versa? To test this, I first identified students who completed both measures. I did this by using students' self-provided school IDs. This resulted in 156 unique students, mostly representing Georgia (n = 113) and Oklahoma (n = 35), though there were a few students from Mississippi and Alabama, as well. A scatterplot shows students' attitude scores against their CTt scores (see Figure 6). This revealed a significant (p = .029) positive, but weak (r = .176) association between the two, accounting for only about 3% of the variance between scores.

Figure 6



Breaking this down by individual attitude (see Table 4), we can see that there were generally weak positive correlations across all attitudes. While both Interest and Utility were statistically significant, Interest was the only attitude to correlate above the .2 level, which is generally considered the level at which a correlation moves from nothing meaningful to weak. Two things are important to note here. First, correlation does not imply causation. Second, this may support the importance of collecting both cognitive and affective data for students, inasmuch as these two measures are clearly measuring something different.

Table 4

Correlations between students' CTt scores and measured attitudes

		Statistic			
Variable	Variable2	Correlation	Count	Lower C.I.	Upper C.I.
CTt	Confidence	.146	156	-.011	.297
	Interest	.257	156	.104	.398
	Utility	.157	156	.000	.307
	Perceptions	.091	156	-.067	.244
	Math	.088	155	-.070	.243
	Social	.041	156	-.117	.197

References

Mason, S. L., & Rich, P. J. (2020). Development and analysis of the Elementary Student Coding Attitudes Survey. *Computers & Education*, 103898. https://doi.org/10.1016/j.compedu.2020.103898

Rebello, N.S. (2012). How accurately can students estimate their performance on an exam and how does this relate to their actual performance on the exam. In AIP Cnference Proceedings, 1413, 315-318. https://doi.org/10.1063/1.3680058

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2016). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking test. *Computers in Human Behavior*. https://doi.org/http://dx.doi.org/10.1016/j.chb.2016.08.047