Investigating the Role of Cognitive Abilities in Computational Thinking for Young Learners

Jean Salac
University of Chicago
Chicago, IL, USA
salac@uchicago.edu

Cathy Thomas
Texas State University
San Marcos, TX, USA
thomascat@txstate.edu

Chloe Butler
Texas State University
San Marcos, TX, USA
cbb64@txstate.edu

Diana Franklin
University of Chicago
Chicago, IL, USA
dmfranklin@uchicago.edu

ABSTRACT

With the global movement to incorporate computer science instruction into elementary education, learners are being introduced to computer science and computational thinking (CS/CT) ideas at increasingly younger ages. At these early ages, young learners are developing cognitive abilities foundational to their education. While other discipline-based education fields, such as math, science, and reading, have long studied the role of cognitive abilities, such as short-term working memory and long-term retrieval, in their respective fields, similar research in computer science education is relatively sparse.

In this exploratory study, we examined the relationship between cognitive abilities and CS/CT performance of fourth-grade students (ages 9-10) who underwent either an introductory CT curriculum based on Use–>Modify–>Create or the same curriculum with additional scaffolding from the TIPP&SEE metacognitive learning strategy. Our analysis revealed performance on CT assessments to be weakly correlated with working memory and long-term retrieval, with correlations increasing as the CT concepts grew more complex. This suggests that scaffolding beyond TIPP&SEE may be needed with more complex CT concepts. We also found that when using TIPP&SEE, students scoring below average on cognitive ability tests performed as well as students in the control condition with average cognitive ability scores. These results indicate TIPP&SEE's potential in creating more equitable computing instruction. We hope that results from this initial exploration can help encourage further study into the role of cognitive abilities in CS/CT education for young learners.

CCS CONCEPTS

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

KEYWORDS

cognitive abilities, computational thinking, memory, elementary education

ACM Reference Format:

Jean Salac, Cathy Thomas, Chloe Butler, and Diana Franklin. 2021. Investigating the Role of Cognitive Abilities in Computational Thinking for Young Learners. In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3446871.3469746

1 INTRODUCTION

All over the world, children are being exposed to the ideas of computer science (CS) and computational thinking (CT) at younger and younger ages as a result of nationwide movements to promote CS education [40]. Between the ages of 6 to 12, children develop the basic cognitive skills needed for learning [28]. Research into cognitive skills have a long tradition in related discipline-based education research fields, such as math [33, 55], science [55, 100], and reading [55, 59]. Few such studies have occurred in computer science. Further, the cognitive science research that does exist in computer science education is mostly with university-age or adult learners [74]. An understanding of the relationship between cognitive abilities and learning outcomes can help a new field such as elementary computer science create an appropriate developmental trajectory to guide standards [68–71], inform development of curriculum and assessment, and set the instructional pace for optimal learning outcomes [43].

To help address this gap, we investigated the cognitive abilities of fourth-grade students (age 9-10) who were introduced to CT either through a Use–>Modify–>Create (UMC) curriculum or the same curriculum with additional scaffolding from the TIPP&SEE learning strategy in a large urban school district in the United States. Use –> Modify –> Create is a learn-by-example approach, where students first observe and make small changes to sample code that demonstrates a new concept before incorporating the concept into a program they write from scratch [46]. TIPP&SEE is a metacognitive strategy that scaffolds the student exploration process in the Use –> Modify step [77].

In this study, we explore the following research questions:

ICER 2021, August 16–19, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8326-4/21/08…$15.00
https://doi.org/10.1145/3446871.3469746

(1) How are working memory, pattern recognition, and long-term retrieval associated with performance on the CS/CT concepts: events, sequence, and loops?
(2) How much does the TIPP&SEE learning strategy support students with differing cognitive abilities?
(3) For which computational thinking concepts does TIPP&SEE support students with differing cognitive abilities?

In the next section, we describe the TIPP&SEE learning strategy. We follow with a delineation of the theories underpinning our work in section 3 and an overview of the literature in section 4. We next describe our methods and results in section 5 and section 6, respectively. We conclude with a discussion of our results in section 7 and their broader implications in section 8.

2 TIPP&SEE LEARNING STRATEGY

Inspired by learning strategies in reading comprehension, the TIPP&SEE metacognitive strategy (Figure 1) guides students as they explore example programs in Scratch, a popular programming language and environment for elementary classrooms [30]. By providing extra guidance, TIPP&SEE further scaffolds the Use –> Modify step in Use –> Modify –> Create lessons.

The first half, *TIPP*, draws from previewing strategies in reading comprehension [97]. *TIPP* stands for Title, Instructions, Purpose, and Play, cuing students to concentrate on these informative aspects of a Scratch project before viewing any code. By focusing on the Title, Instructions, and Purpose of the project, students get a preview of the code, enabling them to set goals and recall previous knowledge. At the final step, Play, students execute the code while carefully observing its execution.

The second half, *SEE*, was designed based on text structure strategies [24]. *SEE* stands for Sprites, Events, and Explore, outlining a process through which students can explore example code. They focus on one sprite at a time and pay close attention to the events in that sprite's code, forming hypotheses on how each sprite's scripts contributed to the code's execution. Finally, they explore the code and learn how to effectively modify the code (e.g. adding, removing, and reordering blocks) to learn how a new code construct works. Further, drawing from work on problem-solving in broader STEM research, *Purpose*, *Play*, and *Explore* stages of the strategy provide prompts, models, and scaffolds for creativity and problem solving in student projects, helping students to plan and self-regulate [64].

Previous work has shown that students using TIPP&SEE outperformed students who used a less scaffolded Use–>Modify–>Create, both in CT assessments and project complexity [32, 77]. A more recent study revealed that students with academic challenges performed as well as students without challenges on CT summative assessments when using TIPP&SEE [76]. This study seeks to delve deeper by exploring the role of cognitive factors in CS/CT instruction for elementary age learners to further inform the developmental continuum of early computer science learning and formation of equitable curriculum.

3 THEORETICAL FRAMEWORK

This research examined the relationship between specific cognitive abilities and elementary computer science learning. Previous research has demonstrated that in related fields such as math [33, 55],

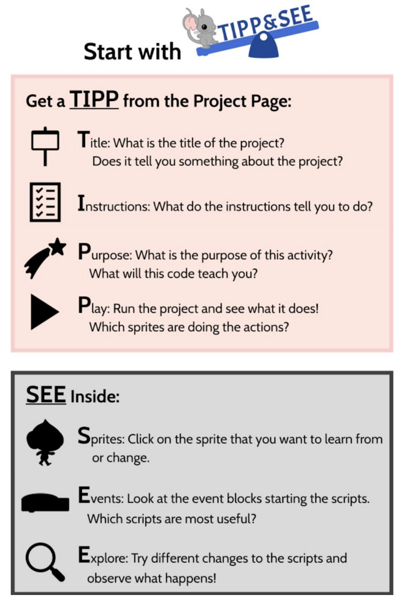


Figure 1: TIPP&SEE Learning Strategy

science [55, 100], and reading [55, 59], cognitive abilities, such as short-term memory and long-term retrieval, influence opportunity to learn [5, 9]. Therefore, cognitive science and intelligence theories [29, 51] should play a role in elementary computer science research, as well as in setting grade level standards, instructional design, and delivery.

3.1 Memory Capacity

Short-term working memory represents the cognitive ability to hold information and manipulate it for use. For example, working memory manages new information during learning and coordinates with long-term memory for tasks of storage and retrieval [3]. Working memory has limited capacity and can only hold information temporarily for use. Therefore, it is most efficient when learning is organized into meaningful chunks and when learners have developed fluency and automaticity for procedural knowledge. For diverse learners, including children with learning disabilities [89], multilingual learners [90], and children living in poverty [63], limitations in working memory influence learning outcomes [17]. Long-term memory is the capacity to encode, organize, store, retrieve, and utilize information; in other words, it is our store of knowledge. As with working memory, individual differences in these capacities impact learning outcomes, particularly for learners with academic challenges [17].

3.2 Cognitive Scaffolding

An important influence in the integration of computer science in elementary school curricula is constructionist theory [30, 37]. Constructionism posits that individuals learn best when they engage in self-directed learning, choosing their own project and learning independently in the process, and when the project produces an artifact that will be displayed to the public [37]. However, research in broader education has demonstrated that for diverse learners, a fully inquiry-based instructional approach does not lead to optimal learning outcomes [44, 79]. For this reason, our curriculum includes two levels of cognitive scaffolding.

The first level of cognitive scaffolding is the Use–> Modify structure in which students learn through example code [46]. At a simple level, access to models and opportunities for practice help to develop the procedural fluency required to reduce cognitive demands and also facilitate the development of mental models of knowledge in long-term memory [52]. Further, these experiences support development of the cognitive flexibility required for higher level thinking and complex tasks [9] that are inherent in the complex problem solving and creativity possible in computer science.

The second level of cognitive scaffolding is TIPP&SEE, a metacognitive strategy that scaffolds elementary computer science learning. In strategy instruction, the learners are taught memory devices for procedures to help learners guide themselves through full exploration and task completion [87]. In this case, it is a mnemonic that serves as a scaffold for executive function. Cognitive strategy instruction was developed to simulate expert information processing and self-regulation, teaching inefficient or less strategic learners to engage in those metacognitive processes to improve learning outcomes [65]. Strategy instruction promotes self-regulation in ways that manage information to optimize short-term memory and long-term storage and retrieval, thus automating procedural knowledge [75, 85]. Metacognitive strategies can facilitate problem solving, helping students to not only grasp the foundational knowledge and procedures, but to understand the conditions under which their knowledge will be useful for problem solving and innovation [16, 27, 51].

4 RELATED WORK

4.1 Landscape of Computational Thinking

While there are divergent views on what computational thinking should be [21, 22, 93, 98], there is broad consensus that computational thinking is a way of thinking used to develop solutions in a form executable by "information processing" or "computational agents" [19]. Similarly, there is wide agreement on these five main elements of computational thinking: algorithmic thinking, logical thinking, abstraction, generalization, and decomposition [83]. In this study, we specifically focus on one core element of CT: algorithmic thinking. Algorithmic thinking is the idea that solutions to problems are generalizable, namely that such solutions (or algorithms) are composed of instructions that, if followed precisely, will yield an answer [19]. Crucially, algorithmic thinking posits that one cannot have a true grasp of computation if one cannot give instructions using sequence, selection, and iteration, based on Turing's result on the essence of computation and Turing completeness [4, 8, 19].

As such, the curriculum used in this study covers three introductory CT concepts: sequence, loops, and events (one way to accomplish selection in Scratch programming). These three concepts and their instructional sequence were drawn from the K-8 learning trajectories [71] and frameworks for K-12 computing education used nationally in the United States [1] and United Kingdom [18].

4.2 Cognitive Abilities in Elementary Education

In education, the assessment of cognitive skills has been considered important in research and practice for purposes such as the identification of disabilities, as a contributor to the understanding of how children learn, and as a predictor of learning outcomes with the intention of improving instruction and outcomes [15]. A significant amount of research over the years has examined the relationships between cognitive skills such as short-term working memory, long-term memory, processing speed, fluid reasoning, auditory and visual processing, and general knowledge [15]. While a thorough summary is beyond the scope of this paper, select examples from disciplines with potential relevance to computer science will be presented.

In reading, expert readers employ metacognitive processes before, during, and after reading in order to comprehend text [65]. Good readers rely on short-term memory to compare new information to information previously read in the text, and to relate and integrate it with information that is already known. Executive functions of self-monitoring guide good readers to know when they don't know, need to re-read, rehearse, slow down, or speed up. Good readers are metacognitive and strategic in their reading processes, intentionally, unconsciously, and intuitively. Experts in math are highly strategic and have learned or developed many strategies from which to draw [84]. In addition to expertise in calculation, these individuals have strong reasoning skills, cognitive flexibility, processing speed, advanced skills in estimation, and elaborated visual representations of numeric information, although measurable limitations in visuo-spatial information processing can mediate this [84]. Additionally, problem solving opportunities and necessities are present in reading, math, and all academic/disciplinary content areas, including computer science. Problem solving is the cognitive process that occurs when the learner must address an issue for which there is not an evident solution at hand to attain a goal [53]. In general, expert problem solvers have a fund of domain knowledge from which to draw. They can easily break complex problems into their sub-components, can hold and manipulate useful information in working memory, can relate patterns and critical features to other known problem/solutions, and focus on the big picture rather than details. They are faster in employing their procedural knowledge and integrating familiar information toward the novel solution [88].

Since reading, math, and problem-solving skills have been demonstrated to be associated with computer science learning [34, 35, 48, 66, 78], research in these related disciplines can help to inform research and practice in computer science as well. To date, limited research into the relationships between cognitive abilities and computer science, especially at the elementary level, exists.

4.3 Cognitive Science in Computing Education

Elementary computer science is a new, developing field. Research into the relationships between cognitive science and computer science learning have garnered increasing interest in the community and exploration is underway. While there has been little work in cognitive abilities in computer science education, there has been research on cognitive load theory. Cognitive load theory seeks to support the learning of complex tasks by using knowledge of cognitive architecture to guide instructional design. Specifically, cognitive load theory seeks to address limitations in working memory and optimize capacity of long-term memory [60]. As early as 2000, computer science researchers began to consider how cognitive load theory would apply to computer science education [96]. Further, Dr. John Sweller, the theorist who first conceptualized Cognitive Load Theory, presented an invited talk at the 2016 SIGCSE technical symposium, *Cognitive Load Theory and Computer Science Education* [91].

Morrison and Guzdial [56] drew on work by Leppink et al. [47] to develop and test a subjective, self-report measure of cognitive load specifically for computer science education. Morrison and colleagues [57] also began exploring instructional design properties from math and science, such as the use of subgoal labels to reduce cognitive load. They found mixed results, suggesting that computer science may require different solutions than those that have been effective for math and science learning. Nonetheless, these studies were conducted at the university level.

At the pre-university level, Seufert presented a conceptual paper, bridging research across self-regulated learning and cognitive load [85]. Seufert includes discussion of strategy instruction to reduce cognitive load, and is germane to our work with metacognitive strategy instruction for this purpose. Sands also argued for applying cognitive science and attending to cognitive load and working memory in the K-12 computer science classroom [80]. While not a full study, Sands' article recommended instructional strategies, such as modeling, worked examples, and peer collaborations, as scaffolds [80]. More recently in 2019, Mutlu-Bayraktar and colleagues published a systematic review of studies that explored cognitive load in multimedia learning [58].

None of the previous studies, however, addressed elementary computer science or studied cognitive abilities specifically. By examining children's cognitive abilities, particularly working and long-term memory, we can better understand the best ways to optimize instructional design to maximize opportunities for developing complex thinking and problem solving skills [61]. This study addresses this research gap by focusing specifically on elementary computer science education and potential associations between cognitive abilities that are relevant and believed to underlie cognitive load in computer science learning [61].

5 METHODS

5.1 Scratch Act 1

Over approximately six months, students were instructed with Scratch Act 1 [2], an introductory computational thinking (CT) curriculum modified from the Creative Computing curriculum [13], in 45-60 minute sessions every 1-2 weeks. One version of the curriculum was scaffolded using the TIPP&SEE learning strategy. Scratch Act 1 is comprised of three modules, covering the CT concepts of sequence, events, and loops. Each module started with Use–>Modify projects to introduce the CT concept, where they learn from example code. Each module led up to a Create project, where students programmed from a blank slate (see Table 1). Sands' recommendations for attending to cognitive load in CS education were integral components of our curriculum [80]. All materials were available in both English and Spanish.

Module	Project	Use-Modify-Create
Sequence	Name Poem	Use–>Modify
	Ladybug Scramble	Use–>Modify
	5 Block Challenge	Create
Events	Events Ofrenda	Use–>Modify
	Parallel Path	Use–>Modify
	About Me	Create
Loops	Build a Band	Use–>Modify
	Interactive Story	Create

Table 1: Scratch Act 1 Modules

5.2 Study Design

Fifteen teachers of fourth-grade students (ages 9-10) were recruited from a large, urban school district in the United States with a high percentage of students from marginalized backgrounds. The teachers were trained with the same professional development to teach the Scratch Act 1 curriculum A total of 16 classrooms participated in the study, six of which were bilingual classrooms. Each classroom was assisted by an undergraduate CS researcher.

Teachers were randomly assigned to either the treatment or the control condition, resulting in five English-only and three bilingual English and Spanish classrooms in each condition. The eight teachers in the treatment condition were taught the TIPP&SEE learning strategy, which scaffolds student exploration of example programs for Use –> Modify activities. Classrooms in the control condition were taught Scratch Act 1 and completed the same Use–>Modify tasks but without the TIPP&SEE worksheets guiding them through familiarization with and exploration of the example code.

There were a total of 92 and 101 students in the TIPP&SEE (TS) and control (C) conditions, respectively. Due to the student population in this study, we tried our best to ensure that both conditions had as similar proportions as possible of students with: economic challenges (76.1% TS vs 89.2% C), multilinguality (designated "Limited English Proficiency" by the school district; 27.2% TS vs 51.5% C), disabilities (17.4% TS vs 14.9% C), and below-grade level proficiencies in reading (58.7% TS vs 45.5% C) and math (59.8% TS vs 57.8% C). These factors were chosen based on research on what factors influence cognitive function in general and computer science performance in particular. More specifically, factors of economic challenge and poverty such as nutrition [67], stress [25], trauma [11], neighborhood-wide poverty [36], rural and urban poor environments [95], refugee status [11], family factors [36], food insufficiency, housing, and employment [20] all impact cognitive development and function. Similarly, economic challenges [26, 31, 50, 79], disabilities [41], English proficiency [66], and reading and math skills [35, 48, 78] all influence computer science performance.

5.3 Computational Thinking Assessment Design

Students took two pen-and-paper assessments, the first one after the Events & Sequence module (E&S) and the second one after the

Loops module (L). Each assessment consisted of a mix of multiple-choice, fill-in-the-blank and open response questions, and were designed to take 20-30 minutes to complete.

Following the Evidence-Centered Design framework [54], assessments were designed based on K-8 learning trajectories for elementary computing [71]. For face validity, questions were first tested with students in the previous school year and then revised based on teacher and student feedback for use in this current study. At each stage, questions were iteratively reviewed by a team of researchers and practitioners from CS and education. We conducted Cronbach's alpha and exploratory factor analyses on the questions in this current study; we do not include analyses from the student trial in the previous school year because questions were revised between the school years.

Cronbach's alpha (α) was calculated for internal reliability between questions on the same topic. Between the questions and sub-questions on both assessments, 5 items targeted events (α=.72), 4 items targeted sequence (α=.7), and 9 items targeted loops (α=.85). A question with parallel loop execution was removed from the reliability calculation because it reduced the reliability of the loops questions to α=.82, indicating that it did not cover the same concepts as the other questions.

Due to the limitations of Cronbach's alpha [92], we complement it with an exploratory factor analysis (EFA) on student scores to characterize the underlying structure of our questions, i.e. which questions tested the same concept and the same level of Bloom's Taxonomy, a framework for classifying learning objectives [7]. We conducted an EFA for two reasons: (1) to further verify the similarity of questions with another method due to the limits of Cronbach's alpha [92] and (2) to discuss similar questions collectively in our results because an individual question is not sufficient to show an understanding of a concept. The EFA was conducted on all student scores in order to have enough responses per question for statistical analysis.

Questions with multiple parts were treated as separate items. We excluded two questions from this analysis: a question on parallelism because of the Cronbach's alpha results, and an extra credit question on nested loops because that concept was not explicitly covered in the curriculum. A maximum likelihood factor analysis was conducted with six factors, the minimum number of factors that was deemed sufficient, and the varimax rotation, which rotates the orthogonal basis so that the factors are not correlated. The minimum number of factors was determined using both the Kaiser's eigenvalue-greater-than-one criterion [42] and the scree plot elbow [10]. Based on the factor loadings from this analysis, we drafted a test blueprint. Table 2 shows the questions accounted by each of the five factors, their loading and their variance. We only included five of the six factors, as the last factor only accounted for one question. The remaining five factors accounted for 12 of the 18 questions included in the factor analysis.

	Remember	Understand
Scratch Basics	E&S Q2, Q3 (Loading=1.07; Variance = 0.06)	—
Events	—	E&S Q4a, Q4b (Loading=1.90; Variance = 0.11)
Sequence	—	E&S Q6, Q7 (Loading=2.08; Variance = 0.12) L: Q5a,b,c (Loading=1.84; Variance = 0.11)
Loops	—	L: Q1, Q2, Q4 (Loading=1.90; Variance = 0.11) L: Q5a,b,c (Loading=1.84; Variance = 0.11)

Table 2: Test Blueprint with Concept & Bloom's Level for Events & Sequence (E&S) and Loops (L) Assessments

5.4 Woodcock-Johnson IV Tests of Cognitive Abilities

The Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV) [81, 82] is a standardized, norm-referenced test of cognitive abilities that was developed based upon the Catell-Horn-Carroll theory of intelligence [29], and is appropriate for measuring cognitive abilities in persons from age two to 80+ years of age. These cognitive tests are not malleable to instruction, but are malleable to child development, maturity, and age. The purpose of these assessments is to gather information that allows comparison of an individual to others of similar age on important cognitive abilities. Together with other sources of information, these types of assessments contribute to the identification of exceptionalities, including "giftedness" and disability. When used ethically and properly, the WJ IV cognitive tests are less flawed, more theoretically grounded, and more fair than other methods of diagnoses [6].

For the purposes of this study, graduate students in school psychology who had been trained to administer this test with fidelity conducted individual assessments of participants in their school settings. Students in our sample were tested before or early in the computer science instructional instruction. All assessments were audio-recorded and were dual scored. Disagreements in scoring were resolved through discussion, with resulting inter-rater agreement of 100%. Inter-rater agreement was supervised by the second author, a licensed diagnostician. Four subtests were administered to each participant.

5.4.1 Numbers Reversed. Numbers Reversed is a 34 item subtest of auditory, short-term working memory. For Grade 4, administrators begin with Sample Item A, and items are presented by audiotape. Participants listen to an increasing series of numbers that do not follow a predictable sequence. For example, Sample Item A includes 2 unrelated numbers, and item 34 includes 8. They are asked to repeat the numbers in reverse order. This test assesses auditory memory that requires both attention and manipulation (recoding) of new information and is a complex span task. This test was selected because research in math and reading has demonstrated that short-term memory is highly predictive of performance [14, 55]

5.4.2 Verbal Attention. Verbal Attention is a 36 item subtest of auditory, short-term working memory. For standardization, each item is presented using an audiotape included in the test kit. Participants listen to an increasing series of words that include animal names and numbers, and are then asked to answer a question. For example, item 9 includes 1 word, while item 36 includes a combination of 5 animal names and numbers. This subtest assesses the capacity of auditory memory, with a focus on attention. In this test, participants are asked to hold information in working memory, and use their executive search skills to identify the correct information to answer a question, assessing abilities in directing attention to needed information that is present in working memory. This subtest was selected because research in math and reading has demonstrated that short-term memory one of the strongest predictors of performance [14, 55].

5.4.3 Pair Cancellation. Pair Cancellation is a 49 item, 3-minute, timed subtest of accuracy in pattern recognition and scanning abilities. In this test, participants scan lines of pictures to identify specific patterns, for example, a picture of a dog, followed by a picture of a ball, and directed to circle each instance. This subtest measures aspects of visual/spatial perception, information processing speed, attention and concentration. This subtest was selected because of its relationship to basic reading skills such as rate and fluency [62] and math calculation [14].

5.4.4 Visual-Auditory Learning. Visual-Auditory Learning is a 7 item subtest of paired associates memory, one aspect of long-term storage and retrieval. In this subtest, participants are shown black and white rebuses and asked to associate each with a word/name. Initially, they are asked to name single rebuses, but as the assessment proceeds, are asked to “read” sentences of sequences of rebuses. This task represents learning, in that it requires short-term working memory to hold and organize the novel information, testing abilities in encoding, storage and retrieval of the new learning. This test requires students to organize, story and retrieve information during learning. The child must remember the word they are taught for each rebus in order to read the sentence. This subtest was chosen because of the importance of encoding to math and reading [62], and because research has demonstrated contributions of long-term memory in math and reading learning and outcomes [86], including for problem-solving [51]. Maximizing long-term memory is a goal of cognitive load theorists, so understanding children’s abilities is important [60]. This test is less common in research, but is commonly used in assessment of children with and at risk for a learning disability.

5.5 Data Analysis

To understand how different cognitive abilities relate to CS/CT performance (our first research question), we first separated our data by groups of TIPP&SEE and control students because of previous work showing that TIPP&SEE was associated with better CT performance [77]. We then ran Spearman correlations between the cognitive abilities subtest scores and scores on their end-of-module assessments. We chose the non-parametric Spearman correlation because not all of the assessment scores met assumptions of normality and linearity. We provide ρ values for correlation strength and p values for statistical significance, with $p < .05$ as our threshold. We also interpreted ρ values based on guidelines from Hinkle et al [39], where $\rho = 0 - 0.3$ is very weak, $\rho = 0.3 - 0.5$ is weak, $\rho = 0.5 - 0.7$ is moderate, $\rho = 0.7 - 0.9$ is strong, and $\rho = 0.9 - 1$ is very strong.

To examine how much TIPP&SEE supports students with various levels of cognitive ability and in which CT concepts (our second and third research questions), we first ranked student scores according to classifications from the Woodcock-Johnson IV test manual (Table 3). The distribution of all student scores followed a normal distribution, with the number of students in each classification decreasing the farther the scores were from the mean. For some subtests, classifications on either tails of the distribution only had one student (see Table 4). As a result, we combined ranks of students with scores in the "Very Superior" and "Superior" into one "Superior" classification and students with scores in the "Very Low" and "Low" into one "Low" classification, in order to have cell sizes large enough for analysis. For the pair cancellation subtest, there was only one student in the "Superior" classification and was therefore excluded from analysis. While the correlations in the previous analysis allow for a more fine-grained picture, this classification allowed us to better describe students’ relative standing among same aged peers and identify students most at-risk.

Scores 40 and below were considered to be outliers and were removed from analysis as they may not represent a fair example of their cognitive abilities. Possible reasons for outliers include students reaching their ceiling before maintaining the minimum score required for each test and students being unable to pay attention to the task for an adequate amount of time (3 minutes). Outliers could also be due to the test environment. For example, if a student was not able to hear the audio the first time it was presented, it could result in them having a lower score because they could not hear the item, nor can the examiner repeat the items. Therefore, these extremely low scores may not be due to limitations in cognitive ability, but instead reflective of test administration issues, including audio difficulty, noise interference, or hearing issues. Further, significant under-performance may be a result of student disengagement with the test conditions. Low student motivation and interest or lack of rapport with the test administrator can influence test results. Additionally, the WJ IV Tests are culturally and linguistically loaded, meaning that children who have limited English proficiency may struggle with the instructions of the test and may score lower than their actual ability due to an inability to understand the test directions and task. Table 4 shows the total number of students, as well as students with disabilities, English Language Learners (as designated by the school district), and students with economic disadvantages in each WJ IV classification.

Comparing across conditions (TIPP&SEE and Control) and cognitive ability classifications (Low, Low Average, Average, High Average, and Superior), we transformed both aggregate and individual question assessment scores with the Aligned Rank Transform (ART), which enables non-parametric factorial analyses, before running an ANOVA F-test [38, 99]. A non-parametric transformation was chosen because of small cell sizes in the WJ IV classifications. Type III sum of squares was employed to account for unequal cell sizes and estimated marginal means were used for post-hoc comparisons. For statistical significance, we provide F and p values for both condition and WJ IV classification. For practical significance [45, 72], we also provide the partial eta squared (η_p^2) effect size. The effect size specifies the magnitude of the observed effect or relationship between variables [49]. η_p^2 measures the proportion of the total variance in a dependent variable (DV) that is associated with the membership of different groups defined by an independent variable (IV) [12]. For example, if an IV has a η_p^2 of 0.25, that means that 25% of a DV's variance is associated with that IV.

WJ IV	Standard Score	Percentile Rank
Very Superior	131 & above	98 to 99.9
Superior	121 to 130	92 to 97
High Average	111 to 120	76 to 91
Average	90 to 110	25 to 75
Low Average	80 to 89	9 to 24
Low	70 to 79	3 to 8
Very Low	41 to 69	0.1 to 2
Extremely Low	40 & below	Outliers

Table 3: Woodcock Johnson IV (WJ IV) Classifications

6 RESULTS

To address our first research question, we first outline the results from our analysis of the correlations between cognitive ability scores and performance on question sets covering the same CT concepts. We next delineate the outcomes from comparing the computational thinking performance of students in different Woodcock-Johnson IV classifications, addressing our second and third research questions.

6.1 Correlations between Cognitive Abilities & CT Performance

We detail the correlations found between cognitive abilities and performance based on the test blueprint of questions and CT concepts developed through the exploratory factor analysis (EFA) described in Section 5. EFA enables us to discuss questions covering the same CT concept as a collective; the following results are organized based on CT concept. A summary of the correlations is shown in Table 5.

Finding 1: Pair Cancellation, a measure of pattern recognition, was not correlated with better performance on any CT concept.

There were almost no correlations between scores on the Pair Cancellation subtest, which measures pattern recognition and scanning abilities, and scores on CT assessment questions. There was only a very weak correlation between Pair Cancellation scores and scores on one question on Events (Q4b) ($\rho = .232, p = .03$) for the TIPP&SEE students. However, given that none of the other questions were correlated and that classification over Pair Cancellation subtest scores were not statistically-significant (see Section 6.2), this very weak correlation on the single question likely does not imply a relationship between the skills measured by the Pair Cancellation subtest and learning the CT concepts covered in this curriculum.

Finding 2: Measures of working memory and long-term retrieval were weakly correlated with better performance on CT questions, with the correlations increasing with more complex CT concepts.

6.1.1 *Scratch Basics.*

There were two questions on the Events & Sequence assessment that covered the basics of Scratch (Table 2). Q2 asked students to identify the last block in script, while Q3 asked students to identify all the scripts that ran when the sprite was clicked. There was a weak correlation between the Numbers Reversed subtest (a measure of working memory) and scores on Q2 for TIPP&SEE students ($\rho = .323, p = .0022$). For control students, there were very weak correlations between Q3 scores and both measures of working memory, Numbers Reversed ($\rho = .270, p = .0077$) and Verbal attention subtests ($\rho = .277, p = .0063$). There was a greater correlation between Q3 scores and scores on the Visual-Auditory Learning subtest, which measures long-term retrieval ($\rho = .431, p = 1.18 \times 10^{-5}$).

6.1.2 *Events.*

Q4a and Q4b in the Events & Sequence assessment covered an understanding of events (Table 2). Looking at a Scratch stage with two sprites that resulted from a green flag click, students were asked to identify the script that ran for each sprite. For TIPP&SEE students, performance on both events questions were very weakly correlated with one of the working memory measures, Numbers Reversed (Q4a:$\rho = .218, p = .043$; Q4b:$\rho = .237, p = .027$). They were more correlated with the other working memory measure, Verbal Attention (Q4a:$\rho = .335, p = .0015$; Q4b:$\rho = .391, p = .00018$), and the long-term retrieval measure, Visual-Auditory Learning (Q4a:$\rho = .420, p = 5.09 \times 10^{-5}$; Q4b:$\rho = .416, p = 6.14 \times 10^{-5}$). For control students, only the long-term retrieval measure was very weakly correlated with scores on Events questions (Q4a:$\rho = .219, p = .031$; Q4b:$\rho = .235, p = .021$).

6.1.3 *Sequence.*

The two questions on sequence from the Events & Sequence assessment (Q6 and Q7) asked students to describe the order in which the blocks in an example script would run. For TIPP&SEE students, there was a very weak correlation between one of the working memory measures, Numbers Reversed, and performance on Q6 ($\rho = .263, p = .014$). Control students showed a very weak correlation between the other working memory measure, Verbal Attention, and performance on Q7 ($\rho = .235, p = .021$). Similar to the questions on Events, they were more correlated with the Visual-Auditory Learning subtest in both TIPP&SEE (Q6:$\rho = .222, p = .039$; Q7:$\rho = .294, p = .0057$) and control conditions (Q6:$\rho = .223, p = .022$; Q7:$\rho = .361, p = .0003$).

6.1.4 *Loops.*

Q1 from the Loops assessment asked students to identify the number of times an example loop would repeat. Q2 and Q4 from the same assessment asked students to unroll a loop, but with different answer choices. Q2 asked about a single-block loop repeating 4 times and had the answer choices of the block in the loop repeated 1, 2, 3, or 4 times. Q4 asked about a double-block loop repeating 3 times and had the answer choices of the two blocks alternating 3 times (the correct execution) and a script with the first block repeated 3 times followed by the second block repeated 3 times (a common misconception) [35]. Q5a, b, and c from the Loops assessment covered both sequence and loops, asking students to identify code that ran before, in, and after a loop.

For questions that only covered loops (Q1, Q2, and Q4), there was only one very weak correlation between Q4 scores and scores on one of the working memory measure, Verbal Attention, for

WJ IV	Total				Disability				ELLs				Econ Disadvantage			
	NR	VA	PC	VAL	NR	VA	PC	VAL	NR	VA	PC	VAL	NR	VA	PC	VAL
Very Superior	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Superior	8	7	1	3	0	1	0	0	2	2	1	1	7	5	0	3
High Average	27	20	6	12	3	2	1	2	8	2	1	1	20	14	6	10
Average	96	107	107	110	11	10	12	16	33	40	46	41	83	93	92	92
Low Average	33	29	33	33	6	10	4	5	15	15	14	14	30	29	32	30
Very Low	7	5	7	0	3	5	2	0	3	4	1	0	7	5	4	0
Extreme Low	2	2	1	0	1	0	1	0	1	1	1	0	2	2	1	0

Table 4: Students in Each WJ IV Classification for all 4 cognitive subtests (Numbers Reversed (NR), Verbal Attention (VA), Pair Cancellation (PC), & Visual-Auditory Learning (VAL))

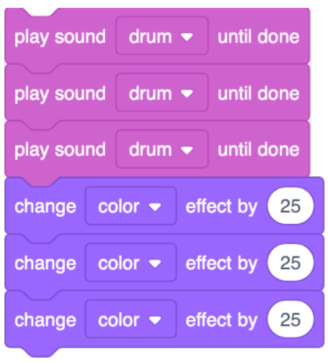


Figure 2: Common Misconception of Multi-Step Loop Execution

TIPP&SEE students ($\rho = .240, p = .027$). In contrast, for control students, performance on Q2 and Q4 were weakly correlated with both working memory measures, Numbers Reversed (Q2:$\rho = .306, p = .0024$; Q4:$\rho = .238, p = .019$) and Verbal Attention (Q2:$\rho = .399, p = 5.75 \times 10^{-5}$; Q4:$\rho = .317, p = .0017$). Visual-Auditory Learning, a measure of long-term retrieval, was weakly correlated with all loops questions for control students (Q1:$\rho = .258, p = .011$; Q2:$\rho = .372, p = .00019$; Q4:$\rho = .381, p = .00013$). For questions covering both sequence and loops (Q5a-c), there were weak correlations between scores on these questions and measures of both working memory and long-term retrieval in both TIPP&SEE (Q5a:$\rho = .347, p = .0011$; Q5b:$\rho = .342, p = .0013$; Q5c:$\rho = .365, p = .00059$) and control conditions (Q5a:$\rho = .358, p = .00034$; Q5b:$\rho = .468, p = 1.52 \times 10^{-6}$; Q5c:$\rho = .360, p = .00032$).

6.1.5 Discussion. With the exception of the Pair Cancellation subtest measuring pattern recognition, the correlations between cognitive skill measures and CT performance grew with the complexity of CT concepts, with more correlations of very weak magnitude ($\rho < .3$) for questions covering Scratch Basics, Events, and Sequence to majority of weak correlations ($\rho = .3 - .5$) for questions with loops.

It is worth noting that for the questions on events, TIPP&SEE student scores were correlated with both measures of working memory and the measure of long-term retrieval, unlike the control students, whose scores were only correlated with the measure of long-term retrieval. The questions on events were the only ones that used the vocabulary word "Stage" to refer to the area in the Scratch interface where students see the output of their code and showed students an image of the Scratch stage. Recalling domain-specific vocabulary may have loaded on students' working memory and long-term retrieval, independent of the computational thinking concept covered by that question. Acquiring disciplinary vocabulary is often a challenge that impedes learning for diverse learners in STEM content [94]. Unlike their typically developing peers, diverse learners may benefit from pre-teaching, explicit instruction, and increased exposure to learn new words well enough for them to be useful. Further, while related images may be paired with key information to enhance learning, visual information that requires interpretation can be more challenging for many learners [94]. Related to Scratch, students need opportunities to see and practice with the graphical representations in the platform, and to pair those with their meaning. In Scratch, the opportunities to use these for their own purposes should also enhance their vocabulary. For questions 4a and 4b, it is possible that neither the word "Stage" nor its image conveyed meaning to students, given limited exposure to this concept in their instruction [73].

It is also noteworthy that for the questions on loops, there was only one very weak correlation between one working memory measure and one question in the TIPP&SEE condition, while in the control condition, both working memory and long-term retrieval measures were correlated with all but one question, which was still correlated with long-term retrieval. Further, for the most advanced questions that required knowledge of both sequence and loops, there were weak, with some bordering on moderate, correlations with both working and long-term retrieval measures in both conditions. While TIPP&SEE may have provided enough additional scaffolding to Use–>Modify–>Create for loops and easier CT concepts, more support may be needed for more complex CT concepts.

6.2 TIPP&SEE Support across Cognitive Abilities

We first report the results from comparing total scores from two end-of-module assessments across WJ IV classifications for each of the cognitive subtests. We follow with results from analyzing different sets of questions that cover different CT concepts to understand which concepts TIPP&SEE provides support for. Results

Concept	Q	Numbers Reversed		Verbal Attention		Visual-Auditory Learning	
		TS	C	TS	C	TS	C
Scratch	E&S Q2	.323**	—	—	—	—	—
Basics	E&S Q3	—	.270**	—	.277**	—	.431**
Events	E&S Q4a	.218*	—	.335**	—	.420**	.219*
	E&S Q4b	.237*	—	.391**	—	.416**	.235*
Sequence	E&S Q6	.263*	—	—	—	.222*	.223*
	E&S Q7	—	—	—	.235*	.294**	.361**
	L Q1	—	—	—	—	—	.258*
Loops	L Q2	—	.306**	—	.399**	—	.372**
	L Q4	—	.238*	.240*	.317**	—	.381**
Sequence	L Q5a	.442**	.321**	.410**	.258*	.347**	.358**
& Loops	L Q5b	.432**	.334**	.268*	.340**	.342**	.468**
	L Q5c	.285**	.285**	.276*	.331**	.365**	.360**

$^{*}p < .05;^{**}p < .01$

Table 5: Correlations between Cognitive Skills & CT Performance on Questions from Events & Sequence (E&S) and Loops (L) Assessments

from different questions are discussed collectively based on the CT concepts they cover (Table 2).

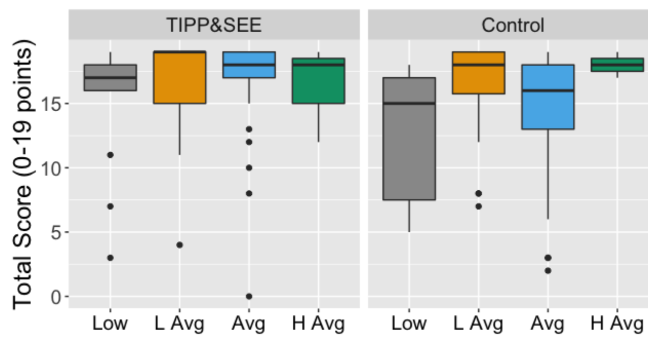
Finding 3: For both TIPP&SEE and control conditions, there was no statistically-significant effect of the Pair Cancellation subtest, a measure of pattern recognition, on CT performance.

Figures 3a and 3b illustrate the distribution of scores in each Pair Cancellation subtest classification within condition, in ascending order of classification from "Low" to "High Average". The "Superior" classification was omitted in the analysis as there was only one student in that category.

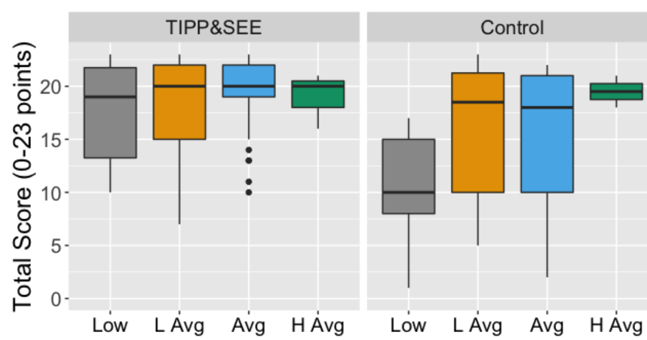
Analysis across classifications for the Pair Cancellation subtest revealed no statistically-significant effect on total scores on either assessment (E&S: $F(1, 161) = 2.19, p = .0918$, L: $F(1, 159) = 2.36, p = .0739$). There was only a statistically-significant effect of condition on aggregate scores on both the Events & Sequence and Loops assessments (E&S: $F(1, 161) = 8.63, p = .0038, \eta_p^2 = .0509$, L: $F(1, 159) = 8.08, p = .0051, \eta_p^2 = .048$). Because of this, further analysis comparing Pair Cancellation classifications and CT questions was not conducted.

Finding 4: When using TIPP&SEE, students classified as having low scores on measures of working memory and long-term retrieval performed equal or better than control students classified as having average scores.

6.2.1 *Numbers Reversed.* Comparing across classifications based on the Numbers Reversed subtest, there were statistically-significant effects of both condition and classification on the scores from both the Events & Sequence (Condition: $F(1, 163) = 8.32, p = .0045, \eta_p^2 = .0486$; Classification: $F(4, 163) = 6.20, p = .00011, \eta_p^2 = .132$) and Loops assessments (Condition: $F(1, 161) = 14.97, p = .00016, \eta_p^2 = .0851$; Classification: $F(4, 161) = 8.99, p = 1.39 \times 10^{-6}, \eta_p^2 = .183$). Post-hoc analyses revealed no statistically-significant differences in performance between TIPP&SEE students with low Number Reversed scores and control students with low average (E&S: $t = -.924, p = .357$, L: $t = -.70, p = .485$), average (E&S: $t = -.836, p = .405$, L: $t = -.345, p = .731$), high average (E&S: $t = -.997, p = .320$, L: $t = -.479, p = .633$), and superior (E&S: $t = -1.38; p = .170$, L: $t = -.253, p = .80$) scores on the Numbers Reversed subtest. Figures 4a and 4b depict the distribution of scores for each Numbers Reversed subtest classification nested within each condition, in increasing order from "Low" to "Superior".

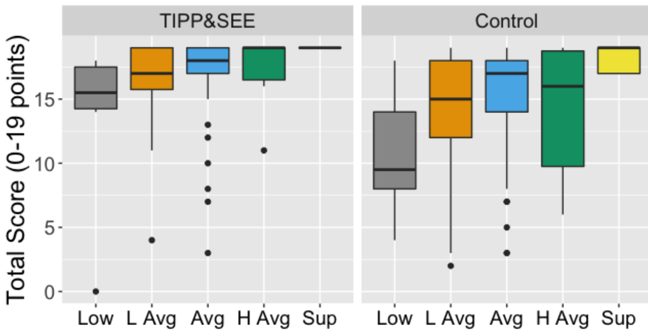


(a) Events & Sequence Assessment

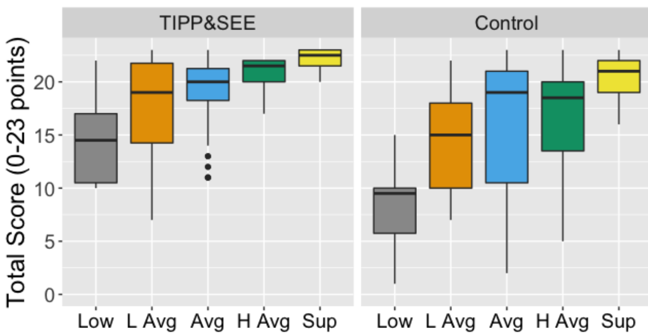


(b) Loops Assessment

Figure 3: Performance across Pair Cancellation Classification

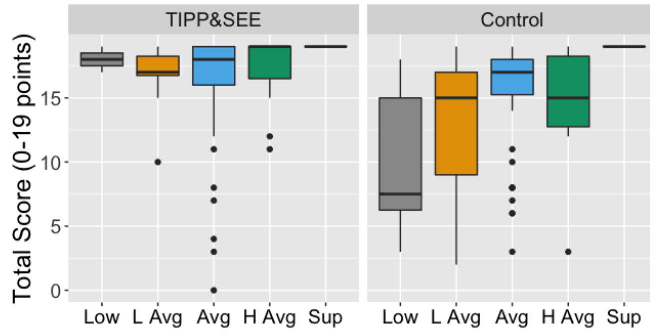


(a) Events & Sequence Assessment

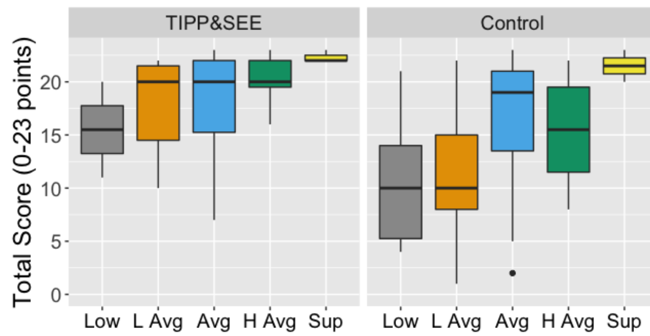


(b) Loops Assessment

Figure 4: Performance across Numbers Reversed Classification



(a) Events & Sequence Assessment



(b) Loops Assessment

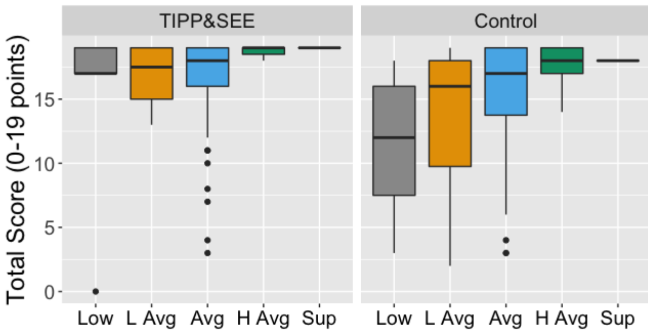
Figure 5: Performance across Verbal Attention Classification

6.2.2 Verbal Attention. Our analysis across Verbal Attention subtest classification showed statistically-significant effects of both condition and classification on aggregate scores on Events & Sequence (Condition: $F(1, 162) = 4.14, p = .043, \eta_p^2 = .0249$; Classification: $F(4, 162) = 6.59, p = 6.05 \times 10^{-5}, \eta_p^2 = .140$) and Loops assessments (Condition: $F(1, 160) = 9.49, p = .0024, \eta_p^2 = .0559$; Classification: $F(4, 160) = 7.67, p = 1.12 \times 10^{-5}, \eta_p^2 = .161$). Unlike in the previous working memory measure Numbers Reversed, TIPP&SEE students with low Verbal Attention scores did not perform as well as control students with average scores on the Events & Sequence assessment. Instead, they *out-performed* them, performing better than control students with low average ($t = -2.57; p = .011$), average ($t = -2.073; p = .039$), and high average ($t = -2.39; p = .018$) scores on the Verbal Attention subtest. Results were more similar to Numbers Reversed in the Loops assessment, with TIPP&SEE students who had low Verbal Attention scores performing as well as control students with low average ($t = -.867, p = .387$), average ($t = -.560, p = .576$), and high average ($t = -.640, p = .523$) Verbal Attention scores. Figures 5a and 5b show the distribution of scores for each Verbal Attention subtest classification for each condition.

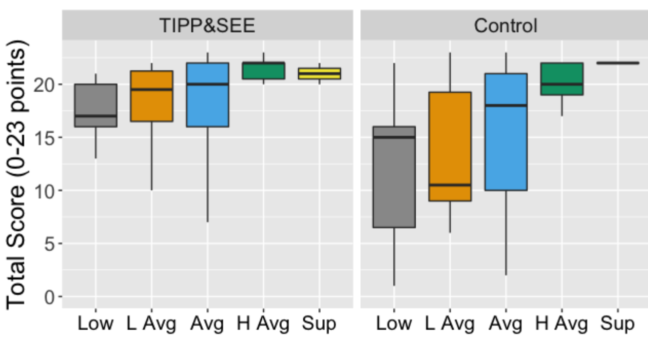
6.2.3 Visual-Auditory Learning. In our analysis across Visual-Auditory Learning subtest classifications, there were statistically-significant effects of both condition and classification on performance on both Events & Sequence (Condition: $F(1, 164) = 5.21, p = .0237, \eta_p^2 = .0308$; Classification: $F(4, 164) = 5.41, p = .00041, \eta_p^2 = .117$) and Loops assessments (Condition: $F(1, 162) = 5.97, p = .016, \eta_p^2 = .0355$; Classification: $F(4, 162) = 5.77, p = .00023, \eta_p^2 = .12$). Similar to the working memory measures, post-hoc comparisons indicated that TIPP&SEE students with low Visual-Auditory scores performed as well as control students with low average (E&S: $t = -1.18, p = .239$, L: $t = -.976, p = .330$), average (E&S: $t = -1.01, p = .316$, L: $t = -.632, p = .528$), and high average (E&S: $t = -1.79, p = .0755$, L: $t = -.488, p = .626$) scores. Figures 6a and 6b portray the distribution of scores for each Visual Auditory-Learning classification in each condition.

Finding 5: For questions on events, there was a statistically-significant effect of Visual-Auditory Learning classification, not condition, on CT performance.

We now turn our attention from aggregate performance to performance on specific concepts. There was only a statistically-significant effect of visual-auditory learning, a measure of long-term retrieval, on performance on both questions covering events (Table 6). There was no effect of condition on performance on either question (Table 6). This may be due to the use of the vocabulary word "Stage" to describe the graphical output of Scratch code, which could have relied on students' long-term retrieval. In contrast, comparisons across the two measures of working memory were ambiguous, where the two Events questions had divergent outcomes.



(a) Events & Sequence Assessment



(b) Loops Assessment

Figure 6: Performance across Visual-Auditory Learning Classification

Finding 6: There were statistically-significant effects of both condition and Verbal Attention classification on performance on questions covering loops.

When contrasting across one of the working memory measures, Verbal Attention, we found statistically-significant effects of both condition and classification on performance on Loops questions (Table 6). This may be early evidence for TIPP&SEE support for the concept of loops and for the role of working memory in learning this concept. On the other hand, comparisons across the other working memory measure, Numbers Reversed, and the long-term retrieval measure, Visual-Auditory Learning, were mixed, with varying outcomes for each question.

Finding 7: Results for the rest of the CT concepts were inconclusive.

Results were mixed for the questions covering the other CT concepts, with disparate outcomes for questions on the same concept. Of the questions covering the basic syntax and semantics of Scratch, the effects of both conditions and cognitive score classifications were not consistent (Table 6). As for questions on sequence, there were no effects of both condition and classification on question scores in comparisons across Verbal Attention classification, while results were mixed for Numbers Reversed and Visual-Auditory Learning (Table 6). Lastly, outcomes were ambiguous for the questions combining sequence and loops for comparisons across subtests of both working memory and long-term retrieval (Table 6).

6.2.4 Discussion. Our analysis of TIPP&SEE support for students with differing cognitive abilities indicated that when using TIPP&SEE, students with low scores on cognitive tests perform similarly on summative, end-of-module assessments as students with average scores who underwent a less scaffolded curriculum. In some assessments, TIPP&SEE students with low cognitive scores even outperformed control students with average scores or performed as well as control students with superior scores, as was the case in our comparison across both the working memory measures, Verbal Attention and Numbers Reversed, on the Events & Sequence assessment.

Results were less definitive when questions were broken down by CT concept. There was a statistically-significant effect of long-term retrieval (Visual-Auditory Learning subtest), not condition, on performance on questions covering events only. In contrast, there were statistically-significant effects of condition and one measure of working memory (Verbal Attention subtest) on performance on questions covering loops. The rest of the CT concepts had mixed outcomes, with questions on the same concept having mismatched outcomes.

7 DISCUSSION

We now return to our overarching research questions:

How are working memory, pattern recognition, and long-term retrieval associated with performance on the CS/CT concepts: events, sequence, and loops?

The correlations found overall were weaker than would be expected based on prior work which showed that cognitive abilities affect learning opportunities in math, science, and reading [33, 55, 59, 100]. This may be because the scaffolding in the curriculum, either through Use–>Modify–>Create alone or Use–>Modify–>Create with TIPP&SEE, supported students in learning CS/CT. While the correlations were smaller than expected, this study presents a critical first step in exploring cognitive capacity in the learning of CS/CT in young students.

As the CT concepts grew more advanced, the correlations increased between measures of cognitive skill and performance on CT questions. For the basic CT concepts of events and sequence, there were various correlations of very weak to weak magnitude between measures of working memory and long-term retrieval and performance on questions covering these concepts. With respect to a more complex CT concept covered in this curriculum, loops, TIPP&SEE student performance was only very weakly associated with one working memory measure for one question. With only one very weak correlation with only one question, it is unlikely that a relationship between working memory and TIPP&SEE students learning of loops. In contrast, the performance of control students were correlated with both measures of working memory, as well as long-term retrieval in all but the simplest loops question. Lastly, for both groups, measures of working memory and long-term retrieval were the most correlated with performance on the most complex set of questions requiring a combined knowledge of both sequence and loops. Taken together, this may be early evidence that this curriculum was manageable for all students with scaffolding from Use–>Modify–>Create for simpler CT topics such as events and sequence. This also implies that with increased complexity, the burden on cognitive abilities might require additional scaffolding, such as TIPP&SEE, or adapted curriculum to remain accessible.

		Condition			Classification		
Numbers Reversed		$F(1,163)$	p	η_p^2	$F(4,163)$	p	η_p^2
Scratch Basics	E&S: Q2	1.57	.213	—	4.13^{**}	.0032	.0921
	E&S: Q3	9.20^{**}	.0028	.0534	6.67^{**}	5.36×10^{-5}	.0141
Events	E&S: Q4a	2.14	.145	—	3.19	.0149	.0725
	E&S: Q4b	1.42	.235	—	1.11	.352	—
Sequence	E&S: Q6	1.46	.228	—	3.382^{**}	.00541	.0856
	E&S: Q7	11.42^{**}	.000911	.0655	3.28^{*}	.0128	.0746
Loops	L: Q1	15.62^{**}	.000116	.0884	2.38	.0543	—
	L: Q2	.556	.453	—	3.08^{*}	.0177	.0717
	L: Q4	5.54^{*}	.0198	.0333	2.68^{*}	.0338	.0624
Sequence & Loops	L: Q5a	3.64	.0580	—	6.28^{**}	.000101	.135
	L: Q5b	28.39^{**}	3.28×10^{-7}	.149	8.25^{*}	4.42×10^{-6}	.170
	L: Q5c	7.79^{**}	.00590	.0461	5.54^{**}	.000333	.121
Verbal Attention		$F(1,162)$	p	η_p^2	$F(4,162)$	p	η_p^2
Scratch Basics	E&S: Q2	4.96^{*}	.027	.0297	.658	.622	—
	E&S: Q3	.909	.342	—	2.34	.058	—
Events	E&S: Q4a	.332	.566	—	1.54	.192	—
	E&S: Q4b	2.47	.118	—	2.24	.0670	—
Sequence	E&S: Q6	.320	.572	—	2.29	.0620	—
	E&S: Q7	.516	.473	—	2.18	.0737	—
Loops	L: Q1	4.29^{*}	.0398	.0261	2.48^{*}	.0464	.0583
	L: Q2	38.27^{**}	4.94×10^{-9}	.193	5.93^{**}	.000178	.129
	L: Q4	23.76^{**}	2.61×10^{-6}	.129	3.51^{**}	.00892	.0807
Sequence & Loops	L: Q5a	2.49	.117	—	4.38^{**}	.00218	.0987
	L: Q5b	11.83^{**}	.000745	.0688	5.49^{**}	.000361	.121
	L: Q5c	2.23	.137	—	6.04^{**}	.000149	.131
Visual-Auditory		$F(1,164)$	p	η_p^2	$F(4,164)$	p	η_p^2
Scratch Basics	E&S: Q2	20.09^{**}	1.38×10^{-5}	.109	1.73	.146	—
	E&S: Q3	.386	.535	—	2.69^{*}	.033	.0617
Events	E&S: Q4a	.0398	.842	—	2.61^{*}	.0375	.0598
	E&S: Q4b	.408	.524	—	2.69^{*}	.0328	.0616
Sequence	E&S: Q6	2.81	.0951	—	2.32	.0589	—
	E&S: Q7	.552	.458	—	2.69^{*}	.0327	.0617
Loops	L: Q1	4.48^{*}	.0358	.0269	1.46	.216	—
	L: Q2	2.49	.116	—	2.36	.0552	—
	L: Q4	6.89^{**}	.00950	.0408	4.21^{**}	.00288	.0941
Sequence & Loops	L: Q5a	4.78^{*}	.0301	.0286	3.10^{**}	.0172	.0711
	L: Q5b	3.64	.0582	—	5.31^{**}	.000483	.115
	L: Q5c	2.68	.104	—	5.19^{**}	.000579	.114

$^{*}p < .05$; $^{**}p < .01$

Table 6: Test Statistics from Concept-Level Analysis of Events & Sequence (E&S) and Loops (L) Assessments

How much does the TIPP&SEE learning strategy support students with differing cognitive abilities?

There were more correlations between CS/CT performance and the Numbers Reversed subtest for students in the TIPP&SEE condition. Numbers Reversed can be considered a measure that focuses on working memory capacity. It is a complex span task in that the operation of reversing the sequence remembered requires active engagement to hold and manipulate the information, and reflects ability to control attention during tasks, an executive function. Weaknesses in these skills would impact an individuals' ability to follow multi-step and complex directions and the quantity of materials that could be managed at a time. Teaching strategies such as verbal rehearsal and visualization can support student learning. Further accommodations would include short, simple directions, visual cues, chunking of information to be learned, and monitoring performance.

For students in the Control condition, there were more correlations between CS/CT scores and the Verbal Attention and Visual-Auditory Learning subtests. The skills assessed by the Verbal Attention subtest, similar to the Numbers Reversed subtest, measure working memory capacity. Uniquely, it also examines skills in holding and finding information for needed purposes, and skills in focusing attention to sort distractors. The ability to update information and find it in a timely fashion is important. Performance on this subtest is predictive of academic performance, and students with weaknesses in these capacities might require additional repetition, limiting distractions, simplification of directions, and also strategy instruction, including mnemonics for learning rules, patterns, and lists of words [81]. This subtest was selected for its role in working memory, which is predictive of performance in other academic tasks. Further, learning, remembering, and retrieving, and using new verbal information may be important to success in learning CS/CT.

The Visual-Auditory Learning subtest may also be related to learning CS/CT. This skill requires associated memory, in which more than one type of information is learned. For example, in Scratch programming, new vocabulary and images are organized and stored together. The term "sprite" and the images of various sprites provide mental models of the concept of sprites. In this subtest, the novel information is both encoded and retrieved. For weaknesses in this skill set, rehearsal, overlearning, shorter and more frequent sessions, and the use of visual images may strengthen learning. Given the results across concepts, while correlations were weak, TIPP&SEE students appeared to be more impacted by short-term memory capacity and attention, while students in the Control group more often were impacted by skills in holding and "looking up" needed information for use, both in short-term working memory (Verbal Attention) and in the associative memory functions in the encoding process that facilitate both storage and retrieval. While this research is exploratory, it is possible that TIPP&SEE is providing the recommended rehearsal and scaffolded practice needed to perform the CS/CT tasks and supporting students.

We also found that on many tasks, students in the TIPP&SEE group with low scores on cognitive markers of short-term memory and long-term retrieval performed as well as their average peers on CS/CT tasks. Students who experience poverty and deprivation tend to have low scores on theses cognitive markers, which are often linked with academic underperformance [11, 20, 25, 36, 67]. Students who demonstrate a weakness in these areas may experience difficulty with developing strategies independently while studying, difficulty with vocabulary development, and difficulty simultaneously remembering a comprehension question and integrating previously learned information. TIPP&SEE can provide the strategic scaffold that "levels the playing field". This "leveling of the playing field" has been demonstrated with explicit teaching of meta-cognitive strategies in academic areas including reading, math, and science [23]. In this study, we see the potential for strategy instruction as an effective scaffold for young learners in CT/CS.

For which computational thinking concepts does TIPP&SEE support students with differing cognitive abilities?
Results were less clear when we took a more detailed look into specific CT concepts. Most questions had inconclusive results, with questions covering the same CT concept having inconsistent outcomes. While students with low cognitive scores were better served with TIPP&SEE in aggregate, we cannot yet tell in which concepts TIPP&SEE was more useful for these students. A larger suite of questions covering a wider variety of learning goals within each CT concept would be necessary to get a more definitive picture.

Across all our research questions, it was surprising that the Pair Cancellation subtest, which measures scanning and pattern recognition, was not associated with CS/CT performance and only weakly demonstrated a correlation with one Events question. This leads us to hypothesize that this cognitive skill is not related to the tasks in this curriculum. Perhaps, these skill is less necessary at the elementary level but would be more important for more advanced CS/CT concepts. It may also be the case that this skill does not apply for younger learners, the CT concepts were not complex enough to engage this skill, or something else entirely. Future research will be needed to further investigate this relationship.

8 IMPLICATIONS, LIMITATIONS, & FUTURE WORK

We discuss the implications, limitations, and future directions with respect to both research and practice.

8.1 Research

This research is exploratory in nature, with a small number of items measuring each CT concept and small cell sizes, resulting in non-normal distributions that required non-parametric analyses. We performed data analysis on separate question items, but we acknowledge that a better approach would be to analyze item clusters based on factor analysis. We did not have enough participants or enough items per construct in order to cluster items with validity. This limitation is not surprising in computing education research for young learners. With computing's largely elective status in elementary education, there is limited instructional time to devote to assessments; more question items increase the time spent testing. Future research should improve upon this instrument with more question items and more participants. Our results should also be replicated beyond this study's context and by other researchers.

While our sample was diverse, all participants attended under-resourced schools serving marginalized communities, so larger samples and broader diversity within those samples should be included. Additional tests of cognitive abilities may be of interest as well, including those that investigate oral vocabulary and general knowledge to consider vocabulary and prior knowledge, and analysis-synthesis to examine problem solving skills, and other measures of cognitive abilities beyond the WJ IV. It is important to remember that one test of cognitive ability gathered at a single time point is not a global representation of a student's working memory, long-term retrieval, or intelligence. Further, there are several factors that are not measured by this assessment or other cognitive skill tests. These factors include curiosity, creative talent, work habits, and study skills.

Cognitive abilities should be investigated further. Given the results of this study that demonstrated increased correlations between memory tasks and CS/CT learning, future study should work to define the parameters for which TIPP&SEE provides sufficient scaffolding for learning. Future research should also identify additional challenges and barriers within computer science for this age group and create and match new strategies to scaffold those needs. A potential avenue would be cognitive load for students in this age group, given prior work on cognitive load in university computing courses [56, 57]. Finally, while TIPP&SEE functioned very well for students aged 9-10, understanding its applicability to other grades and developmental levels is imperative. The development of additional strategies may be useful to scaffold other CS/CT concepts/content.

8.2 Practice

TIPP&SEE is a novel strategy that has been developed and tested for use in elementary CS/CT learning. Measures of teacher fidelity of implementation were not explored, leaving the possibility that some results may be attributed to teacher effects. TIPP&SEE is presented in combination with a Use–>Modify–>Create approach, and some outcomes may be influenced by the contributions of this model rather than TIPP&SEE alone. The components of TIPP&SEE have not been disaggregated and tested. Therefore, it is unknown which aspects of TIPP&SEE make the greatest contributions and if any aspects are not critical to the model. Previous research has demonstrated that when instructed using a Use–>Modify–>Create instructional sequence scaffolded by TIPP&SEE, students who underperformed academically on state assessments, including diverse learners such as students with disabilities, students from marginalized communities, and multilingual students, performed as well as typically developing peers [76]. The findings of this study provide additional support for the use of TIPP&SEE for elementary student CS/CT learning given the predictive qualities of tests of cognitive abilities for under-performance in academic tasks. The results of this study also demonstrates that for most students of lower measurable cognitive abilities, TIPP&SEE helped to level the playing field, allowing them to participate fully in the general CS/CT curriculum.

ACKNOWLEDGMENTS

We would like to thank the teachers and students who participated in this study, as well as the undergraduates who assisted in the computer science instruction. This project was funded by National Science Foundation (NSF) Grant No. 1660871 and DGE-1746045.

REFERENCES

[1] [n.d.]. K–12 Computer Science Framework. https://k12cs.org/
[2] [n.d.]. Scratch Act 1. https://www.canonlab.org/scratchact1modules
[3] Eryn J Adams, Anh T Nguyen, and Nelson Cowan. 2018. Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools* 49, 3 (2018), 340–355.
[4] Alfred V Aho. 2012. Computation and computational thinking. *The computer journal* 55, 7 (2012), 832–835.
[5] American Psychological Association et al. 2016. Coalition for Psychology in Schools and Education.(2015). *Top 20 principles from psychology for preK–12 teaching and learning* (2016).
[6] Mark Benisz, John O Willis, and Ron Dumont. 2018. 16 Abuses and Misuses of Intelligence Tests: Facts and Misconceptions. *Pseudoscience: The conspiracy against science* (2018), 351.
[7] Benjamin S Bloom et al. 1956. Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay* (1956), 20–24.
[8] Corrado Böhm and Giuseppe Jacopini. 1966. Flow diagrams, Turing machines and languages with only two formation rules. *Commun. ACM* 9, 5 (1966), 366–371.
[9] J Bransford, N Vye, R Stevens, P Kahl, D Schwartz, P Bell, A Meltzoff, B Barron, R Pea, B Reeves, et al. 2005. Learning theories and education: toward a decade of synergy. The LIFE Center: The University of Washington.
[10] Raymond B Cattell. 1966. The scree test for the number of factors. *Multivariate behavioral research* 1, 2 (1966), 245–276.
[11] Alexandra Chen, Catherine Panter-Brick, Kristin Hadfield, Rana Dajani, Amar Hamoudi, and Margaret Sheridan. 2019. Minds Under Siege: Cognitive Signatures of Poverty and Trauma in Refugee and Non-Refugee Adolescents. *Child development* 90, 6 (2019), 1856–1865.
[12] Jacob Cohen. 1988. Statistical power analysis for the behavioural sciences.
[13] Creative Computing. [n.d.]. An introductory computing curriculum using Scratch.
[14] Damien C Cormier, Okan Bulut, Kevin S McGrew, and Deepak Singh. 2017. Exploring the relations between Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics achievement. *Applied Cognitive Psychology* 31, 5 (2017), 530–538.
[15] Damien C Cormier, Kevin S McGrew, Okan Bulut, and Allyson Funamoto. 2017. Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment* 35, 8 (2017), 731–754.
[16] National Research Council et al. 2000. *How people learn: Brain, mind, experience, and school: Expanded edition.* National Academies Press.
[17] Nelson Cowan. 2014. Working memory underpins cognitive development, learning, and education. *Educational psychology review* 26, 2 (2014), 197–223.
[18] Andrew Csizmadia, Paul Curzon, Mark Dorling, Simon Humphreys, Thomas Ng, Cynthia Selby, and John Woollard. 2015. Computational thinking-A guide for teachers. (2015).
[19] Paul Curzon, Tim Bell, Jane Waite, and Mark Dorling. 2019. Computational thinking. *The Cambridge Handbook of Computing Education Research* (2019), 513–546.
[20] Junhua Dang, Shanshan Xiao, and Siegfried Dewitte. 2015. Commentary:"Poverty impedes cognitive function" and "The poor's poor mental power". *Frontiers in psychology* 6 (2015), 1037.
[21] Peter J Denning. 2017. Remaining trouble spots with computational thinking. *Commun. ACM* 60, 6 (2017), 33–39.
[22] Peter J Denning and Matti Tedre. 2019. *Computational thinking.* MIT Press.
[23] Anouk S Donker, Hester De Boer, Danny Kostons, CC Dignath Van Ewijk, and Margaretha PC van der Werf. 2014. Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review* 11 (2014), 1–26.
[24] S Dymock. 2005. Teaching Expository Text Structure Awareness. *The Reading Teacher* 59, 2 (2005), 177–181.
[25] Gary W Evans and Thomas E Fuller-Rowell. 2013. Childhood poverty, chronic stress, and young adult working memory: The protective role of self-regulatory capacity. *Developmental science* 16, 5 (2013), 688–696.
[26] Cheri Fancsali, Linda Tigani, Paulina Toro Isaza, and Rachel Cole. 2018. A Landscape Study of Computer Science Education in NYC: Early Findings and Implications for Policy and Practice. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education.* ACM, 44–49.
[27] Febrian Febrian et al. 2016. The Instruction To Overcome The Inert Knowledge Issue In Solving Mathematical Problem. *Jurnal Gantang* 1, 1 (2016), 16–25.
[28] Kurt W Fischer and Daniel Bullock. 1984. Cognitive development in school-age children: Conclusions and new directions. *Development during middle childhood: The years from six to twelve* (1984), 70–146.
[29] Dawn P Flanagan and Shauna G Dixon. 2013. The Cattell-Horn-Carroll theory of cognitive abilities. *Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals* (2013).
[30] Louise P Flannery, Brian Silverman, Elizabeth R Kazakoff, Marina Umaschi Bers, Paula Bontá, and Mitchel Resnick. 2013. Designing ScratchJr: support for early childhood learning through computer programming. In *Proceedings of the 12th International Conference on Interaction Design and Children.* ACM, 1–10.
[31] Baker Franke, Jeanne Century, Michael Lach, Cameron Wilson, Mark Guzdial, Gail Chapman, and Owen Astrachan. 2013. Expanding access to K-12 computer science education: research on the landscape of computer science professional development. In *Proceeding of the 44th ACM technical symposium on Computer science education.* ACM, 541–542.
[32] Diana Franklin, Jean Salac, Zachary Crenshaw, Saranya Turimella, Zipporah Klain, Marco Anaya, and Cathy Thomas. 2020. Exploring Student Behavior Using the TIPP&SEE Learning Strategy. In *Proceedings of the 2020 ACM Conference on International Computing Education Research.* 91–101.
[33] Lynn S Fuchs, Amelia S Malone, Robin F Schumacher, Jessica Namkung, and Amber Wang. 2017. Fraction intervention for students with mathematics difficulties: Lessons learned from five randomized controlled trials. *Journal of Learning Disabilities* 50, 6 (2017), 631–639.
[34] Annagret Goold and Russell Rimmer. 2000. Factors affecting performance in first-year computing. *ACM SIGCSE Bulletin* 32, 2 (2000), 39–43.

[35] Shuchi Grover, Roy Pea, and Stephen Cooper. 2016. Factors influencing computer science learning in middle school. In *Proceedings of the 47th ACM technical symposium on computing science education*. ACM, 552–557.

[36] Daniel A Hackman, Laura M Betancourt, Robert Gallop, Daniel Romer, Nancy L Brodsky, Hallam Hurt, and Martha J Farah. 2014. Mapping the trajectory of socioeconomic disparity in working memory: Parental and neighborhood factors. *Child development* 85, 4 (2014), 1433–1445.

[37] Idit Ed Harel and Seymour Ed Papert. 1991. *Constructionism*. Ablex Publishing.

[38] James J Higgins, R Clifford Blair, and Suleiman Tashtoush. 1990. The aligned rank transform procedure. (1990).

[39] Dennis E Hinkle, William Wiersma, and Stephen G Jurs. 2003. *Applied statistics for the behavioral sciences*. Vol. 663. Houghton Mifflin College Division.

[40] Peter Hubwieser, Michail N Giannakos, Marc Berges, Torsten Brinda, Ira Diethelm, Johannes Magenheim, Yogendra Pal, Jana Jackova, and Egle Jasute. 2015. A global snapshot of computer science education in K-12 schools. In *Proceedings of the 2015 ITiCSE on working group reports*. ACM, 65–83.

[41] Maya Israel, Quentin M Wherfel, Jamie Pearson, Saadeddine Shehab, and Tanya Tapia. 2015. Empowering K–12 students with disabilities to learn computational thinking and computer programming. *TEACHING Exceptional Children* 48, 1 (2015), 45–53.

[42] Henry F Kaiser. 1960. The application of electronic computers to factor analysis. *Educational and psychological measurement* 20, 1 (1960), 141–151.

[43] Paul A Kirschner. 2002. Cognitive load theory: Implications of cognitive load theory on the design of learning.

[44] Paul A Kirschner, John Sweller, and Richard E Clark. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist* 41, 2 (2006), 75–86.

[45] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* 4 (2013), 863.

[46] Irene Lee, Fred Martin, Jill Denner, Bob Coulter, Walter Allan, Jeri Erickson, Joyce Malyn-Smith, and Linda Werner. 2011. Computational thinking for youth in practice. *Acm Inroads* 2, 1 (2011), 32–37.

[47] Jimmie Leppink, Fred Paas, Cees PM Van der Vleuten, Tamara Van Gog, and Jeroen JG Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behavior research methods* 45, 4 (2013), 1058–1072.

[48] Colleen M Lewis and Niral Shah. 2012. Building upon and enriching grade four mathematics standards with programming curriculum. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. ACM, 57–62.

[49] Jessica Middlemis Maher, Jonathan C Markey, and Diane Ebert-May. 2013. The other half of the story: effect size analysis in quantitative research. *CBE—Life Sciences Education* 12, 3 (2013), 345–351.

[50] Jane Margolis. 2010. *Stuck in the shallow end: Education, race, and computing*. MIT Press.

[51] Richard E Mayer. 2011. *Applying the science of learning*. Pearson/Allyn & Bacon Boston, MA.

[52] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.

[53] Richard E Mayer and Merlin C Wittrock. 2006. Problem solving. *Handbook of educational psychology* 2 (2006), 287–303.

[54] 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.

[55] P. L. Morgan, G. Farkas, Y. Wang, M. M.. Hillemeier, and S. Maczuga. [n.d.]. Executive function deficits in kindergarten predict repeated academic difficulties across elementary school. *Early childhood research quarterly* 46 ([n. d.]). https://doi.org/10.1016/j.ecresq.2018.06.009

[56] Briana B. Morrison, Brian Dorn, and Mark Guzdial. 2014. Measuring Cognitive Load in Introductory CS: Adaptation of an Instrument. In *Proceedings of the Tenth Annual Conference on International Computing Education Research* (Glasgow, Scotland, United Kingdom) *(ICER '14)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2632320.2632348

[57] Briana B. Morrison, Lauren E. Margulieux, and Mark Guzdial. 2015. Subgoals, Context, and Worked Examples in Learning Computing Problem Solving. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research* (Omaha, Nebraska, USA) *(ICER '15)*. Association for Computing Machinery, New York, NY, USA, 21–29. https://doi.org/10.1145/2787622.2787733

[58] Duygu Mutlu-Bayraktar, Veysel Cosgun, and Tugba Altan. 2019. Cognitive load in multimedia learning environments: A systematic review. *Computers & Education* 141 (2019), 103618.

[59] Suzan Nouwens, Margriet A Groen, and Ludo Verhoeven. 2017. How working memory relates to children's reading comprehension: the importance of domain-specificity in storage and processing. *Reading and writing* 30, 1 (2017), 105–120.

[60] Fred Paas, Tamara Van Gog, and John Sweller. 2010. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational psychology review* 22, 2 (2010), 115–121.

[61] Fred Paas and Jeroen JG van Merriënboer. 2020. Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks. *Current Directions in Psychological Science* 29, 4 (2020), 394–398.

[62] Peng Peng and Douglas Fuchs. 2016. A meta-analysis of working memory deficits in children with learning difficulties: Is there a difference between verbal domain and numerical domain? *Journal of learning disabilities* 49, 1 (2016), 3–20.

[63] Seth D Pollak and Barbara L Wolfe. 2020. How developmental neuroscience can help address the problem of child poverty. *Development and Psychopathology* 32, 5 (2020), 1640–1656.

[64] Michelle Popham, Simone Adams, and Janie Hodge. 2020. Self-regulated strategy development to teach mathematics problem solving. *Intervention in School and Clinic* 55, 3 (2020), 154–161.

[65] Michael Pressley and Karen R Harris. 2006. Cognitive strategies instruction: From basic research to classroom instruction. (2006).

[66] Yizhou Qian and James D Lehman. 2016. Correlates of Success in Introductory Programming: A Study with Middle School Students. *Journal of Education and Learning* 5, 2 (2016), 73–83.

[67] Amy C Reichelt, R Fred Westbrook, and Margaret J Morris. 2017. Impact of Diet on Learning, Memory and Cognition. *Frontiers in behavioral neuroscience* 11 (2017), 96.

[68] 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*. 124–132.

[69] Kathryn M Rich, Diana Franklin, Carla Strickland, Andy Isaacs, and Donna Eatinger. 2020. A Learning Trajectory for Variables Based in Computational Thinking Literature: Using Levels of Thinking to Develop Instruction. *Computer Science Education* (2020), 1–22.

[70] Kathryn M Rich, Carla Strickland, T Andrew Binkowski, and Diana Franklin. 2019. A k-8 debugging learning trajectory derived from research literature. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 745–751.

[71] 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.

[72] John TE Richardson. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 2 (2011), 135–147.

[73] Kathryn L Roberts, Rebecca R Norman, Nell K Duke, Paul Morsink, Nicole M Martin, and Jennifer A Knight. 2013. Diagrams, timelines, and tables—Oh, my! Fostering graphical literacy. *The Reading Teacher* 67, 1 (2013), 12–24.

[74] Anthony V Robins, Lauren Margulieux, and Briana B Morrison. 2019. Cognitive sciences for computing education. (2019).

[75] Carly Rosenzweig, Jennifer Krawec, and Marjorie Montague. 2011. Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. *Journal of learning disabilities* 44, 6 (2011), 508–520.

[76] Jean Salac, Cathy Thomas, Chloe Butler, and Diana Franklin. 2021. Supporting Diverse Learners in K-8 Computational Thinking with TIPP&SEE. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 246–252.

[77] Jean Salac, Cathy Thomas, Chloe Butler, Ashley Sanchez, and Diana Franklin. 2020. TIPP&SEE: A Learning Strategy to Guide Students through Use-Modify Scratch Activities. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 79–85.

[78] Jean Salac, Cathy Thomas, Bryan Twarek, William Marsland, and Diana Franklin. 2020. Comprehending Code: Understanding the Relationship between Reading and Math Proficiency, and 4th-Grade CS Learning Outcomes. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 268–274.

[79] Jean Salac, Max White, Ashley Wang, and Diana Franklin. 2019. An Analysis through an Equity Lens of the Implementation of Computer Science in K-8 Classrooms in a Large, Urban School District. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 1150–1156.

[80] Philip Sands. 2019. Addressing cognitive load in the computer science classroom. *ACM Inroads* 10, 1 (2019), 44–51.

[81] Fredrick A Schrank, Scott L Decker, and John M Garruto. 2016. *Essentials of WJ IV cognitive abilities assessment*. John Wiley & Sons.

[82] Fredrick Allen Schrank, Kevin S McGrew, Nancy Mather, Barbara J Wendling, and Erica M LaForte. 2014. *Woodcock-Johnson IV tests of cognitive abilities*. Riverside.

[83] Cynthia Selby and John Woollard. 2013. Computational thinking: the developing definition. (2013).

[84] Francesco Sella and Roi Cohen Kadosh. 2018. What expertise can tell about mathematical learning and cognition. *Mind, Brain, and Education* 12, 4 (2018), 186–192.

[85] Tina Seufert. 2018. The interplay between self-regulation in learning and cognitive load. *Educational Research Review* 24 (2018), 116–129.
[86] Mikyung Shin and Diane Pedrotty Bryant. 2015. A synthesis of mathematical and cognitive performances of students with mathematics learning disabilities. *Journal of learning disabilities* 48, 1 (2015), 96–112.
[87] Jack P Shonkoff, Greg J Duncan, Philip A Fisher, Katherine Magnuson, and Cybele Raver. 2011. Building the brain's "air traffic control" system: How early experiences shape the development of executive function. *Contract* 11 (2011).
[88] Robert J Sternberg and Talia Ben-Zeev. 2001. *Complex cognition: The psychology of human thought.* Oxford University Press.
[89] H Lee Swanson. 2015. Intelligence, working memory, and learning disabilities. In *Cognition, intelligence, and achievement*. Elsevier, 175–196.
[90] H Lee Swanson, Michael J Orosco, and Catherine M Lussier. 2015. Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology* 132 (2015), 155–188.
[91] John Sweller. 2016. Cognitive Load Theory and Computer Science Education *(SIGCSE '16)*. Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/2839509.2844549
[92] Keith S Taber. 2018. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education* 48, 6 (2018), 1273–1296.
[93] Matti Tedre and Peter J Denning. 2016. The long quest for computational thinking. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. 120–129.
[94] Cathy Newman Thomas, Delinda Van Garderen, Amy Scheuermann, and Eun Ju Lee. 2015. Applying a universal design for learning framework to mediate the language demands of mathematics. *Reading & Writing Quarterly* 31, 3 (2015), 207–234.
[95] Michele Tine. 2014. Working memory differences between children living in rural and urban poverty. *Journal of Cognition and Development* 15, 4 (2014), 599–613.
[96] Juhani E. Tuovinen. 2000. Optimising Student Cognitive Load in Computer Education. In *Proceedings of the Australasian Conference on Computing Education* (Melbourne, Australia) *(ACSE '00)*. Association for Computing Machinery, New York, NY, USA, 235–241. https://doi.org/10.1145/359369.359405
[97] Sharon Vaughn, Janette Klingner, Elizabeth Swanson, Alison Boardman, Greg Roberts, Sarojani Mohammed, and Stephanie Stillman-Spisak. 2011. Efficacy of Collaborative Strategic Reading With Middle School Students. *American Education Research Journal* 48, 4 (august 2011), 938–964.
[98] Jeannette M Wing. 2006. Computational thinking. *Commun. ACM* 49, 3 (2006), 33–35.
[99] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.
[100] Kun Yuan, Jeffrey Steedle, Richard Shavelson, Alicia Alonzo, and Marily Oppezzo. 2006. Working memory, fluid intelligence, and science learning. *Educational Research Review* 1, 2 (2006), 83–98.