

Accounting for Pedagogical Content Knowledge in a Theory of Change Analysis

Tom McKlin
The Findings Group
Decatur, Georgia USA
tom@thefindingsgroup.org

Brian Magerko
Digital Media Program
Georgia Institute of Technology
Atlanta, Georgia USA
magerko@gatech.edu

Emily Bryans
The Findings Group
Decatur, Georgia USA
emily@thefindingsgroup.org

Taneisha Lee
The Findings Group
Decatur, Georgia USA
taneisha@thefindingsgroup.org

Doug Edwards
Center for Education Integrating
Science Mathematics and Computing
Georgia Institute of Technology
Atlanta, Georgia USA
doug.edwards@ceismc.gatech.edu

Jason Freeman
Center for Music Technology
Georgia Institute of Technology
Atlanta, Georgia USA
Jason.freeman@gatech.edu

Dana Wanzer
The Findings Group
Decatur, Georgia USA
dana@thefindingsgroup.org

Sabrina Grossman
Center for Education Integrating
Science Mathematics and Computing
Georgia Institute of Technology
Atlanta, Georgia USA
sabrina.grossman@ceismc.gatech.edu

ABSTRACT

Educators have long claimed that pedagogical content knowledge (PCK), ways of presenting a subject that make it comprehensible to others, is a critical element of student academic success. This paper presents an exploratory study finding that PCK significantly correlates with students' computer science (CS) content knowledge acquisition while teacher CS content knowledge does not. This study analyzes systemic factors comprising the theory of change for the EarSketch learning intervention including classroom- and student-level factors that contribute to shifts in students' attitudes toward computing and ultimately to students' intentions to persist in future computing endeavors and students' CS content knowledge. We present the results of a multi-level modeling analysis and offer an exploratory approach to measuring CS PCK along with recommendations to improve the sensitivity of the method.

CCS CONCEPTS

- J.5 [Arts and Humanities], K.3.2 [Computer Education]: Computers and Information Science Education – *computer science education, curriculum*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICER '19, August 12–14, 2019, Toronto, ON, Canada
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6185-9/19/08...\$15.00
<https://doi.org/10.1145/3291279.3339412>

KEYWORDS

Computer Science Education, Pedagogical Content Knowledge, Teacher Professional Development, STEAM, Music, Broadening Participation

ACM Reference format:

Tom McKlin, Taneisha Lee, Dana Wanzer, Brian Magerko, Doug Edwards, Sabrina Grossman, Emily Bryans and Jason Freeman. 2019. Accounting for Pedagogical Content Knowledge in a Theory of Change Analysis. In *ICER '19: 2019 International Computing Education Research Conference. August 12–14, 2019, Toronto, Canada*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3291279.3339412>

1 Introduction

As the number of computer science (CS) students increases, the number of teachers supporting those students also grows. Along with the growth in the number of CS teachers comes the need to provide professional learning opportunities to equip those teachers with pedagogical practices that help learners comprehend the content and resolve misconceptions. This paper analyzes the theory of change underlying EarSketch [28][30], a CS learning environment, curriculum, and professional development program deployed within Computer Science Principles high school classrooms, and in an exploratory study, finds that pedagogical content knowledge correlates with improved student content knowledge, while teachers' CS content knowledge does not. We outline the theory of change, analyze the connections among predictor and outcome variables in the theory of change,

and discuss the classroom-level factors related to student content knowledge growth.

Although the study described in this paper is specific to EarSketch, the findings with respect to our theory of change suggest the importance of a) equipping all CS teachers with the pedagogical content knowledge to help student learners comprehend complex content and address common misconceptions and b) considering the role of teacher pedagogical content knowledge in research studies on student CS learning outcomes.

1.1 EarSketch

EarSketch strives to engage diverse student populations in introductory computer science by emphasizing the personally expressive role of computing in the domain of music. EarSketch students learn basic elements of computing and sample-based music composition (i.e. composition using musical beats, samples, and effects). For more information on EarSketch, see [28].

The design of EarSketch leverages music's potential as a hook to engage students in computing, as well as the natural conceptual parallels between the two disciplines. It affords an "immediate opportunity to act" [5] in that students with no prior musical or computational background can quickly become musically expressive, writing a few lines of code to assemble audio clips from the loop library. EarSketch is also perceived to be authentic [28][39] by students; it utilizes industry-relevant programming languages, borrows heavily from music production paradigms in its user interface and API, and incorporates popular musical styles and content created by well-known engineers and producers.

1.2 EarSketch Professional Learning

Teachers in this study participated in a summer professional learning workshop that consisted of a month-long asynchronous online component followed by three consecutive, day-long, face-to-face meetings. The asynchronous learning component included videos, text, quizzes, and short programming assignments focused on content knowledge in CS and EarSketch. The face-to-face sessions focused on pedagogical content knowledge and more advanced CS and EarSketch concepts. Pedagogical topics were led by experienced CSP teachers as well as EarSketch staff, and included pair programming, live problem coding, project-based learning, and a group activity in which teachers devised and delivered sample lesson plans. During the subsequent academic year, teachers were supported with two half-day in-person training sessions, an online community where they could ask questions of the research team, and an online library of teaching materials and training videos. All of the professional learning events concentrated on the Computer Science Principles programming unit (Big Idea 5)[7][8].

1.3 Theory of Change

The EarSketch research agenda is guided by a theory of change. In general, theories of change outline the central mechanisms by which change occurs; it is often a diagram revealing what program

participants are expected to experience and the behaviors or internal characteristics that may change for participants along with the anticipated outcomes that result from those changes [13]. The EarSketch theory of change (Figure 1) shows the theorized characteristics expected to occur in classrooms taught by teachers participating in EarSketch professional learning. It also reveals anticipated outcomes from students of participating teachers.

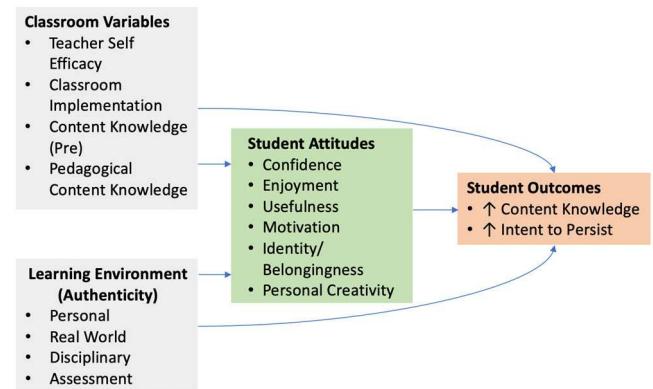


Figure 1. EarSketch Theory of Change

This paper uses the EarSketch theory of change as a backdrop to better understand how teachers' PCK connects to student content knowledge. The EarSketch team modified its original theory of change [10][28], which only included student-level variables, to account for classroom-level variables (*teacher self-efficacy*, *classroom implementation*, *teacher CS content knowledge*, and *pedagogical content knowledge* prior to entering the program). In working with 18 teachers across five school districts, we found that a theory of change focused solely on student-level factors was not enough to capture the significant differences in outcomes we found across the various classroom implementations. Thus, we added teacher-level variables in an effort to better understand the variance from classroom to classroom that affects student outcomes. The EarSketch theory of change posits that classroom variables combined with a learning environment conducive to learning music and computing contribute to changes in student attitudes, which then lead to changes in two important student outcomes: increased computer science content knowledge and an intention to persist in future computing education and computing-centric careers. The following sections describe each element of the theory of change. In the remainder of this section, we address the frameworks and strategies we use to understand each variable in the theory of change. In the following section, we detail the specific approaches to adapting and/or creating instruments to measure them.

1.3.1 Classroom Variables. We introduce four classroom-level variables into the theory of change. The first is CS teaching self-efficacy [44], which asks teachers to self report their beliefs that their content knowledge is sufficient and useful, their ability to facilitate CS learning, and that they can engage students and motivate them to succeed. Classroom implementation asks teachers about the degree to which they used curriculum resources

tied to each area of the 5E model of instruction: engagement, exploration, explanation, elaboration, and evaluation [1]. A fuller description of classroom implementation and its connection to EarSketch outcomes may be found in [28]. We also assessed teachers' content knowledge at two time points. The content knowledge assessment (see section 2.2.3) was designed to address the learning objectives in "Big Idea 5" of the Computer Science Principles Framework, and the assessment that teachers took is the same assessment that their students took.

1.3.2 Learning Environment. We take a constructionist perspective [19] that acknowledges that learning is most effective when students are actively engaged in their learning environment. Shaffer and Resnick [39] along with Lee and Butler [26] offer a framework for studying the learning environment from the students' perspective by asking them whether the learning environment is personally meaningful to them (Personal), contains learning opportunities that relate to the world outside of school (Real World), provides opportunities to think in the modes of a particular discipline (Disciplinary), and equips students to assess their own learning and what comes next in their learning (Assessment). The environment refers to both the online EarSketch environment and the classroom structure.

1.3.3 Student Attitudes. The theory of planned behavior holds that the most important precursor to intending to do a behavior (i.e., persist in computing) is predicted by perceived control over the behavior, subjective norms (e.g. the support of friends and family), and attitudes toward the behavior [1][3]. We focus on student attitudes in this study, which are operationalized as students' confidence in, enjoyment of, importance of, and motivation to succeed in computing, as well as their perceived identity as someone who engages in computing and their perceptions of personal creativity. See [10][28][29] for a more extensive treatment of these constructs.

1.3.4 Student Outcomes. The EarSketch theory of change contains two outcome variables: students' intentions to persist in future computing endeavors and students' increases in computing content knowledge. The EarSketch project has primarily focused on using music as a hook that promotes student interest leading to the intentional pursuit of future computing endeavors, and this paper treats both intention to persist and students' CS content knowledge as outcome variables. Therefore, it is important to describe our approach to assessing content knowledge.

We measure content knowledge using distractor-driven multiple-choice (DDMC) tests. DDMCs are assessment tools in which students select one correct response from a group of response options containing distractors that represent common misconceptions. Students typically select one response option out of four: one response option is correct, and the remaining three represent a common misconception [12][16]. Perhaps the most popular of these is the Force Concept Inventory [17] in physics. While this approach is fairly commonplace, it has faced criticism. Some complain that students may only select from among a

previously-identified set of ideas [10] and that popular distractors may "trap even very knowledgeable students" [33].

Sadler et al. [36][37] reason that good assessments accurately elicit students' ideas whether they are right or wrong and that we do not wish students to arrive at correct answers by eliminating a set of weak distractors. Sadler's team contends "multiple-choice items that do not employ distractors based on known misconceptions are of questionable utility for measuring conceptual understanding." The DDMC approach also allows educators to measure shifts in understanding, preferably pre-to-post movement from misconception to correctly understanding a concept. In section 2.2.3, we discuss our approach to developing a DDMC content knowledge assessment for this and other EarSketch studies.

1.3.5 Pedagogical Content Knowledge (PCK). As summarized in section 1, this study finds a direct link between student content knowledge in CS and teachers' pedagogical content knowledge. Teachers demonstrating higher PCK tend to have students who demonstrate higher content knowledge. While this connection may seem obvious, it is important to know that this connection occurs under conditions in which teachers' content knowledge is not correlated with students' content knowledge. Along with other researchers [2][23][24][40][46], we surmise that PCK plays an important role in the development of student content knowledge, and that it therefore demands further consideration in the design of teacher professional development and further focus in the design and analysis of CS research studies.

PCK was originally defined in Shulman's presidential address to the American Educational Research Association in 1985. Shulman sought to legitimize teaching as a profession by demonstrating that teaching requires a specific form of knowledge, specifically pedagogical content knowledge which he defined as the "subject matter knowledge for teaching" [41]. Shulman elaborates: PCK is "The most useful forms of representation..., the most powerful analogies, illustrations, examples, explanations, and demonstration -- in a word, the ways of representing and formulating the subject that make it comprehensible to others."

One approach [36][37] to measuring PCK asks teachers to identify and describe common misconceptions among their students. Educational researchers view misconceptions in different ways: some perceive misconceptions as ideas to be confronted and replaced while others see misconceptions as productive building blocks for re-shaping students understanding [9][18][42]. We take the latter constructivist perspective and see 'misconceptions' as knowledge elements to be refined, reorganized, and built upon. Our methods section describes an approach to measuring teachers' pedagogical content knowledge that asks teachers to both identify the most popular misconception from a distractor-driven multiple-choice test and to describe instructional interventions to re-shape student understanding.

2 Methods

2.1 Procedures

Twenty teachers attended professional learning during the summer before the 2017-18 academic year, and two dropped out. The 18 remaining teachers taught 882 students during the 2017-18 academic year, and 473 (53.6%) of those students consented to participate in the study. As described in section 1.2, teachers prepared for the three-day summer professional learning session by participating in a month-long asynchronous online component.

2.2 Assessment Instruments

This study examines the theory of change to discern how classroom-level variables relate student attitudes and the two student outcome variables: intention to persist and content knowledge. While the theory of change suggests causality, the statistical approaches show correlation and cannot prove causality.

2.2.1 Self-Efficacy. We modified the Teaching Engineering Self-efficacy Scale [47] and administered the 16-item self-efficacy survey to teachers that measures five subconstructs: *Computing Pedagogical Knowledge Self Efficacy*, *Instructional Self Efficacy*, *Engagement Self Efficacy*, *Disciplinary Self Efficacy*, and *Outcome Expectancy*. However, self-efficacy was not examined at the subconstruct level and the overall mean composite of self-efficacy was used instead because the subconstructs are highly correlated and would have issues of multicollinearity in regression models. The instrument was administered upon entry (pre) into the program and then after teachers finished teaching all EarSketch units (post). For this analysis, we used responses from the post administration with the expectation that the expressed self-efficacy while teaching more closely aligns with the post administration.

2.2.2 Classroom Implementation. We measured classroom implementation through an enactment checklist administered to teachers for three lessons over the course of the EarSketch module. The self-report enactment checklist was adapted from two prior projects [14][31] that focused on curriculum development and implementation. The enactment surveys were constructed as an efficient method to measure curriculum implementation with multiple teachers in diverse geographic settings. The self-reported enactment checklist was used during curriculum implementation and supported by teacher interviews and observations. Data from the checklist aligned with teacher responses from interviews, encouraging continued use of this tool when evaluating enactment on a large-scale curricular intervention.

The first step in development of the checklists was to identify the critical components, or essential program elements of the curriculum [6][31]. In EarSketch, these critical components were defined by the learning goals of each lesson. These learning goals were organized on the enactment checklist using the 5E instructional model [4] to match the lesson plan format provided to the teachers. The enactment checklist asks teachers whether and

to what degree they used curriculum resources and the percentage of students engaged in each stage of the lesson.

2.2.3 Content Knowledge. We assessed computer science content knowledge for both teachers and students using a 20-item, language-agnostic instrument developed by the EarSketch team with support from an Advanced Placement Computer Science (AP CS) teacher. The assessment is aligned to the AP Computer Science Principles Framework [7][8], specifically objectives 5.1.1 to 5.4.1 (Big Idea 5: Programming). The assessment was administered as a traditional pre- and post-test: students took the test before and after the EarSketch units in their CSP course; and teachers took the test at the end of professional learning and again after completing the EarSketch unit in their CSP course. The assessment covers knowledge areas appropriate for a multiple-choice format. The research team also conducted think-aloud cognitive interviews and assessed item difficulty and item discrimination with high school computer science principles students to verify that students correctly interpreted the items. Findings from the assessment were further verified through six focus groups with a representative sample of students from the study.

2.2.4 Pedagogical Content Knowledge. This study uses a similar approach to the one used by Sadler et al. [36] who administered a subject matter knowledge assessment (SMK) along with two additional questions as measures of pedagogical content knowledge (PCK): 1) What proportion of your students will answer this item correctly? and 2) Of the students who answer incorrectly, which distractor are they most likely to choose? The team reasoned that knowing the most popular distractor helps teachers reconstruct student misunderstandings [36].

Similarly, we extended the content knowledge assessment by modifying the CKA form for teachers. We extended five items of the content knowledge assessment administered to teachers. After answering the item, the teacher was asked three follow-up questions:

1. Out of all of the incorrect responses, which response are students most likely to select?
2. Why do you think students would select this option? In other words, what student misconception leads them to answer this way? (one or two sentence response)
3. Please describe how you would correct the students' misconception. (two or three sentence response).

The highest possible score on the pedagogical content knowledge assessment is a 20. For each of the five items, teachers were awarded one point for correctly answering the following sub-items:

1. Answering the content knowledge question correctly.
2. Selecting the distractor students were most likely to select.
3. Correctly identifying why students would select that distractor.
4. Describing an appropriate intervention.

Each of the four parts of the PCK assessment is measured as follows. The first was graded like most content knowledge assessments and asks respondents to select one correct answer from four response options. One option is correct and the other three represent a common misconception surrounding that particular item. The second was determined by analyzing all available student responses to each item to determine the most popular incorrect response option and then comparing the actual most popular misconception from the student data to the teacher's answer to this item. The third and fourth parts of the assessment required comparing teacher responses against an expert teacher's response. One EarSketch team member is also an experienced CS teacher and provided "acceptable" responses against which teacher responses were awarded either a 1 if it aligned with the expert response or a 0 if it did not.

2.2.5 Learning Environment and Student Attitudes. The EarSketch student survey measures students' perceptions of the learning environment (whether they perceive it to be authentic to making music and to computing) along with attitudes (comprised of *confidence*, *enjoyment*, *importance*, *motivation*, *identity*, *intent to persist*, and *personal creativity*). See [10], [29], and [44] for more on the student survey, particularly *authenticity* and *personal creativity*. The student survey was administered at the end of the unit as a retrospective pre-post. This approach accounts for response-shift bias, changes in participants' understanding of the items as a result of the intervention [20][21][35]. The authenticity scale was administered as a post-only scale because students cannot rate the *authenticity* of a learning environment prior to experiencing the learning environment.

2.2.6 Qualitative Analyses. Teachers and administrators participated in 30- to 60-minute semi-structured interviews after teaching the EarSketch module. Interviews were audio recorded and transcribed verbatim. The first cycle of coding involved creating attribute (school district, number of years teaching CSP, etc.) and structural codes [38]. Structural codes identify content-based phrases related to implementation, decision making in selecting CS curriculum and courses, experiences with EarSketch, differing interest and engagement by subgroups, factors that enhance teaching, and types of support needed to teach CSP with EarSketch. During the second cycle of coding, pattern coding was used to develop major themes. We also interviewed students and reported those findings in prior work [10][28][29][44]; this analysis focuses more specifically on classroom-level factors.

2.3 Multi-Level Modeling

Multi-level modeling is a statistical approach that accounts for grouped data [22] and is suitable for this analysis because individuals in a group tend to be more similar than individuals across groups. We analyze the data at two levels: students (level 1 data) grouped within classrooms (level 2 data). *Teaching efficacy*, *classroom implementation*, *teacher content knowledge (pre)*, and *teacher pedagogical content knowledge* are classroom-level data (level 2) while student ratings of the learning environment (*authenticity*), the six student attitudinal constructs, and student outcomes (content knowledge and intent to persist) are student-

level data (level 1). We performed three sets of three MLMs: one set each for *student attitudes*, *student content knowledge*, and *intent to persist*. The first step is to determine what percentage of the outcome at the student-level is predicted by classroom-level variables. The second step includes all student-level predictors. The third step adds the classroom-level predictors to see how well they predict the outcome above and beyond the student-level predictors.

3 RESULTS: PCK

Our analysis reveals a significant relationship between teacher PCK and students' content knowledge but reveals no significant connection between teachers' content knowledge and students' content knowledge. Therefore, we present the findings from the teachers' PCK assessment as context for understanding how PCK and content knowledge fit into the larger analytic structure. We also acknowledge the limitations of the PCK assessment and student CKA and offer this as a viable approach to measuring professional learning that could be improved with more rigorous PCK and content knowledge assessments and a larger sample of teachers. Recall that the PCK assessment has four parts, and each is addressed below.

Table 1. Response Distribution for Teachers and Students

Item	Student Responses (n = 473)			
	A	B	C	D
13	9.9%	26.1%	19.9%	44.1%
15	53.6%	<u>19.9%</u>	17.2%	9.3%
16	<u>38.3%</u>	22.1%	29.5%	10.2%
17	17.1%	38.7%	<u>32.5%</u>	11.7%
18	<u>33.0%</u>	22.6%	35.1%	9.3%
Item	Teacher Responses (n = 20)			
	A	B	C	D
13	0%	0%	<u>15.0%</u>	85%
15	90%	<u>10.0%</u>	0%	0%
16	<u>15.0%</u>	5.0%	65.0%	<u>15%</u>
17	10.0%	55.0%	<u>25.0%</u>	10%
18	<u>25.0%</u>	0%	75.0%	0%

Note. Bold percentages represent the correct response; underlined percentages represent the most popular incorrect response.

3.1 Part One: Answering the Content Knowledge Item

Overall, 20 teachers took the PCK assessment, and those teachers range in general teaching experience from one to 25 years ($M = 10.8$; $SD = 7.8$) and in computer science teaching experience from one to 21 years ($M = 4.3$; $SD = 5.1$). The average score on the PCK assessment was 48% correct. Table 1 shows both student and teacher results from the content knowledge assessment, and Table 2 shows how teachers performed by experience level. Given there were relatively few teachers, we analyze teaching experience in two broad categories: those who are in their first year of teaching CSP and those who have taught CSP for one or more years. Not surprisingly, teachers with one or more years teaching experience scored higher (57% correct) than first-year teachers (26% correct).

Hubbard [23] notes that teachers approach professional learning with a unique perspective: they acquire content knowledge through the lens of reteaching that content. Knowing that teachers and students approach CS content from different perspectives, we sought to understand whether teachers and students have the same misconceptions; that is, do teachers and students who miss an item choose the same distractor? Table 1 shows that students and teachers selected the same distractor for four of the five items. They differ on item 13: students were more likely to choose response option B while teachers were more likely to choose response option C.

Indeed, CS education has the unique challenge of training teachers in both CK and PCK through in-service PD. The fact that many teachers are making the same mistakes as students on the content knowledge assessment suggests that their content knowledge is still emerging (and often just a small step ahead of their students), and that PD for CS has a dual burden of addressing CK for teachers and also equipping them with the PCK to do the same for their students.

3.2 Part Two: Selecting the Most Popular Distractor

One aspect of PCK is understanding that students are going to misunderstand concepts, and one difference between a novice and experienced teacher is that the more experienced teacher can generally identify popular misconceptions. In order to measure this, we asked teachers to identify the distractor that students are most likely to select. Table 1 shows the response distribution of 473 students. Table 2 shows that first year teachers were less likely to select the most popular distractor compared to teachers with more CSP experience. For example, item 16 shows that two of the six first year CSP teachers selected the most popular distractor while seven of 14 teachers with a year or more of experience selected the most popular distractor.

Table 2. Number and Percentage of Teachers Correctly Identifying the Most Popular Distractor

Item	First Year CSP (n=6)	1+ Years CSP (n=14)
13	3 (50%)	10 (71%)
15	0 (0%)	5 (36%)
16	2 (33%)	7 (50%)
17	2 (33%)	4 (29%)
18	2 (33%)	7 (50%)

Overall, teachers with one or more years of experience had difficulty consistently identifying the most common misconception ($M = 47.2\%$) and first year teachers scored equivalent to guessing ($M = 30\%$).

3.3 Part Three: Identifying the Misconception

We asked teachers to explain the student thought processes that would lead a student to select the most popular distractor. We then compared the teachers' responses to an expert response. Table 3 shows the number and percentage of teachers who broadly identified the same student thought processes as the CS PCK expert. This table shows that the first year CSP teachers had greater difficulty answering this item and that teachers with one

or more years of experience performed better but have room for growth. For example, item 16 is a difficult item (only 38.3% of 473 students answered it correctly); only one of six first year teachers correctly identified the student thought processes leading them to the most popular distractor whereas more than half (8 of 14) of the experienced teachers accurately identified the students' thought process.

Table 3. Number and Percentage of Teachers Correctly Identifying the Students' Misconceptions

Item	First Year CSP (n=6)	1+ Years CSP (n=14)
13	2 (33%)	6 (43%)
15	1 (17%)	10 (71%)
16	1 (17%)	8 (57%)
17	1 (17%)	6 (43%)
18	0 (0%)	5 (36%)

3.4 Part Four: Describing an Appropriate Instructional Response

The final portion of the PCK assessment asks teachers to describe how they would respond to a student or a group of students who selected the most popular distractor. We answer this portion of the item in two steps: first we look at the responses from all teachers, and second, we asked whether there were any teachers who did not correctly identify the students' misconception but correctly identified an appropriate response. Table 4 shows the number and percentage of new and experienced teachers who answered this item commensurate with a CS PCK expert. Most first year CSP teachers do not adequately identify an appropriate instructional response whereas between one-third and one-half of experienced teachers do.

Table 4. Number and Percentage of Teachers Correctly Identifying the Appropriate Instructional Response

Item	First Year CSP (n=6)	1+ Years CSP (n=14)
13	1 (17%)	5 (36%)
15	0 (0%)	8 (57%)
16	1 (17%)	8 (57%)
17	0 (0%)	8 (57%)
18	0 (0%)	4 (29%)

We acknowledge the subjective nature of this approach; importantly, many of the incorrect responses were either "I don't know" (e.g. "Honestly, I am not sure how to best explain this one.") or focused on test-taking strategies rather than instruction (e.g. "Practice breaking down the question carefully every time"). The envelope of correct responses is broad and focuses on instructional practice that directly addresses the misconception. An example of a correct response to one of the items is: "I would pull together the students who missed the problem and have them trace the code on a piece of paper and chart the value of result at each iteration of the loop. I would then demo it on the board."

The team further noticed that incorrect responses tended to focus on “explaining” while correct responses focused on “doing.” For example, some teachers proposed explanations such as “I would tell them to look at the random...,” “I will review loops,” and “remind them that variables can change values.” These responses did not adequately describe instructional interventions and were not awarded any points.

Correct responses primarily involved doing: “I would ask them what prints with each one and how they are different....” Two other teachers offered:

I would pull those students [the ones answering incorrectly] into a group and code it with them. I would put in the code from answer B and run it. Then I would ask ‘how many times did the #1 get rolled?’ I would demo and lead them through the correct solution.

I would have them draw a table of values. A column for factor, a column for counter and ask them to fill it out for each of the 12 iterations.

These responses indicated an intervention, an activity where students engaged and interacted with the coding concept or code itself in order to address the students’ misunderstanding.

We also sought to understand whether any teachers who did not correctly identify the misconception may have correctly identified the appropriate instructional response. We found that this occurred ten times and primarily among experienced teachers. This phenomenon occurred one time for item 13, four times for item 16, three times for item 17, and once for item 18. Among first year teachers, only one who missed the misconception correctly identified the instructional response (Item 16). This may reveal a potential area of further study: how often do teachers provide the correct instructional response without fully understanding the students’ misconception? Finally, we acknowledge the limitations of matching teacher responses to an expert response and note that a number of more robust approaches would yield more accurate results; specifically, researchers might develop a rubric depicting acceptable and unacceptable qualities of responses [25].

4 RESULTS: MULTI-LEVEL MODELING

We present the multi-level modeling results in three parts. We first describe the percentage of the variance in the outcome variables that is explained by classroom placement. Second, we describe how the student-level variables affect student attitudes, content knowledge, and intent to persist (Table 1, Model 1). Third, we add in classroom-level predictors to determine the role of PCK and other classroom-level predictors above and beyond the student-level predictors (Table 1, Model 2).

4.1 Percentage of Each Outcome Predicted by Classroom Variables

The first step is determining what percentage of the outcome at the student-level is explained by the grouping structure of the data by calculating the Intraclass Correlation Coefficient (ICC). These analyses essentially run a simple regression with only the grouping

variable as a predictor of each of the three outcomes. The ICC values indicate 20% of the variation in student attitudes, 32% of the variation in student content knowledge, and 13% of the variation in students’ intention to persist are explained by classroom membership. These findings indicate that a significant portion of the variation in student outcomes is explained by the students’ classroom. The following analyses examine which student- and classroom-level variables contribute to the variation in student outcomes.

4.2 Student- and Classroom-Level Predictors of Student Outcomes

Table 5 shows the results of three MLM analyses, one each for student attitudes, student content knowledge, and students’ intention to persist. Each analysis contains two models: model 1 uses student-level variables to predict the outcome variable, and model 2 adds classroom-level variables to the student-level variables.

Table 5. MLM Results

Student Attitudes (Post)	Model 1	Model 2
<i>Student-level variables</i>		
Authenticity	.291 (.04)***	.233 (.04)**
Attitudes (Pre)	.586 (.03)***	.646 (.04)***
<i>Classroom-level variables</i>		
PCK		.145 (.07)*
Self-efficacy		-.027 (.04)
Content Knowledge (Pre)		.008 (.07)
Implementation		.101 (.04)*
Intent to Persist (Post)	Model 1	Model 2
<i>Student-level variables</i>		
Authenticity	.170 (.04)***	.126 (.03)**
Attitudes (Pre)	.559 (.04)***	.626 (.03)***
<i>Classroom-level variables</i>		
PCK		.039 (.10)
Self-efficacy		-.050 (.06)
Content Knowledge (Pre)		.053 (.010)
Implementation		.086 (.06)
Content Knowledge (Post)	Model 1	Model 2
<i>Student-level variables</i>		
Authenticity	.051 (.05)	.021 (.06)
Attitudes (Pre)	.086 (.05)	.082 (.06)
CKA (Pre)	.529 (.05)***	.515 (.07)***
<i>Classroom-level variables</i>		
PCK		.382 (.16)*
Self-efficacy		-.022 (.10)
Content Knowledge (Pre)		-.260 (.16)
Implementation		-.006 (.10)

Note. Values shown are the Standardized Beta (standardized error) coefficients which are effect sizes in regression and MLM models. * indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$.

The first MLMs in Table 1 examines student attitudes. Model 1 shows that authenticity (student perceptions that the learning environment is authentic) and attitudes toward computing upon entering the program both significantly predict post student

attitudes. Model 2 adds classroom-level variables and shows that PCK and classroom implementation both predict student attitudes above and beyond attitudes prior to the program and perceived authenticity of the environment.

The second set of MLMs in Table 1 examine students' intent to persist. Model 1 shows that both authenticity and attitudes towards computing upon entering the program significantly predict students' intent to persist; however, none of the classroom-level variables in Model 2 were significantly related to intent to persist above and beyond the student-level variables.

The third set of MLMs in Table 1 examine student content knowledge. Model 1 shows that neither authenticity nor student attitudes upon entry to the program predicted content knowledge, but their pretest assessment of content knowledge significantly predicted their posttest assessment of content knowledge. Furthermore, in adding the classroom-level variables in Model 2, PCK was a significant predictor of students' posttest content knowledge above and beyond their pretest content knowledge.

5 Limitations

We present this approach to assessing professional learning as exploratory and acknowledge numerous limitations to be addressed as studies involving more teachers and classrooms become commonplace. First, multi-level models are strengthened by increasing the number of clusters: for this study, that means increasing the number of teachers (level 2). Second, the content knowledge instrument could be strengthened through panel reviews by experts to improve face and construct validity of the items and distractors and to account for culture bias. Sadler et al. [36] recommend calculating distractor strength, that the most popular distractor should account for at least 50% of the overall distractor responses. Two of our items were just short of this threshold, and we recommend administering the content knowledge assessment to a representative sample of students beforehand and select PCK items from the group of items with distractor strengths greater than 50%. We also recommend designing distractors that produce one distractor that meets the 50% distractor strength threshold. Third, the analysis revealed that one item (Item 15) did not contain what many believed to be the best distractor and that the assessment would be improved by identifying misconceptions represented by each distractor are the most common misconceptions. Fourth, we acknowledge that teaching experience is correlated with PCK and that an important next step is to determine whether PCK continues to grow with experience and the extent to which these variables are correlated.

6 Discussion

This paper analyzes the EarSketch theory of change and confirms findings from prior work [29] that an authentic learning environment and student attitudes contribute to students' intentions to persist in future computing endeavors. This analysis also reveals two important classroom-level connections to students' content knowledge: Teacher pedagogical content

knowledge (PCK) significantly correlates with students' content knowledge performance and teachers' CS content knowledge does not. Relatedly, a closer analysis of teachers' PCK shows that many teachers struggle to correctly identify student misconceptions even though they may have possessed the same misconceptions as students.

While certainly with limitations, this paper offers a broad method to provide feedback on professional learning. It asks teachers to take the same content knowledge assessment as students, also asks teachers to identify their students' strongest misconceptions, compares student misconceptions based on actual student performance to teachers understanding of student misconceptions, asks teachers to describe the misconception, and finally asks teachers to reflect on instructional strategies to best address the most pressing student needs. We envision follow-up support during the academic year that reinforces content knowledge while focusing on students' understanding of CS concepts.

Given the shortage of CS teachers, school administrators have developed a variety of strategies for preparing CS educators. Our findings suggest that PCK is an important factor in student content knowledge growth and that experienced teachers tend to outperform novices in PCK. Professional development activities focused on identifying and addressing student misconceptions by "doing" rather than "explaining" may help accelerate the growth of pedagogical skills in educators. As with any discipline, a teacher's pedagogical ability to present a subject in a way that makes it comprehensible to others may be as important as their content knowledge.

As an emerging field in K-12 education, most CS teachers receive in-service training on the job instead of pre-service training before they enter the profession. This places a significant burden on CS professional development programs, which often must focus on basic content knowledge for teachers new to the field as well as on the PCK necessary to be effective in the classroom. In the uphill climb to equip teachers with necessary content knowledge across short periods of PD, some programs (including EarSketch PD) have struggled to find time to focus on PCK. In some cases, CS education leaders have sought to hire experienced computing professionals without experience in teaching to fill instructional needs, in the hopes that these individuals would already possess the requisite content knowledge.

The findings presented in this initial study suggest the importance of teacher PCK on student content knowledge outcomes, even more so than teacher content knowledge. We hope that these findings will encourage CS professional development providers to increasingly focus on PCK in their training and will encourage CS education researchers to further explore the relationship between PCK and student outcomes.

ACKNOWLEDGMENTS

The EarSketch project receives funding from the National Science Foundation (1417835 and 1649671). EarSketch is available online at earsketch.gatech.edu.

REFERENCES

- [1] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0747-5978\(91\)90020-T](https://doi.org/10.1016/0747-5978(91)90020-T)
- [2] Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, 52(1), 154–168. <https://doi.org/10.1016/j.compedu.2008.07.006>
- [3] Ben-Bassat Levy, R., & Ben-Ari, M. (2018). The Evaluation of Robotics Activities for Facilitating STEM Learning. In W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh, & D. Obdržálek (Eds.), *Robotics in Education* (Vol. 630, pp. 132–137). https://doi.org/10.1007/978-3-319-62875-2_12
- [4] Bybee, R., Taylor, J., Gardner, A., Van Scotter, P., Powell, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E Instructional Model: Origins and Effectiveness*. Colorado Springs, CO: BSCS. Retrieved from https://bscs.org/sites/default/files/_legacy/BSCS_5E_Instructional_Model-Full_Report.pdf
- [5] Carroll, J. 1998. Minimalism Beyond the Nurnberg Funnel. MIT Press.
- [6] Century, J., Rudnick, M., & Freeman, C. (2010). A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge. *American Journal of Evaluation*, 31(2), 199–218. <https://doi.org/10.1177/1098214010366173>
- [7] College Board (2016). AP Computer Science Principles Curriculum Framework Accessed March 3, 2016 from <https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-computer-science-principles-course-and-exam-description>.
- [8] College Board. (2017). AP Computer Science Principles: Course and Exam Description. New York, NY: College Board.
- [9] diSessa, A. A. (2018). A Friendly Introduction to “Knowledge in Pieces”: Modeling Types of Knowledge and Their Roles in Learning. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited Lectures from the 13th International Congress on Mathematical Education* (pp. 65–84). Springer International Publishing.
- [10] Engelman, S., Magerko, B., McKlin, T., Miller, M., Edwards, D., & Freeman, J. (2017). Creativity in Authentic STEAM Education with EarSketch (pp. 183–188). ACM Press. <https://doi.org/10.1145/3017680.3017763>
- [11] Finley, F. N. (1986). Evaluating Instructing: The Complementary Use of Clinical Interviews. *Journal of Research in Science Teaching*, 23, 635.
- [12] Freyberg, P. and Osborne, R. 1985, in Learning in Science: The Implication of Children's Science, ed. R. J. Osborne and P. Freyberg, Auckland, Heineman, 166.
- [13] Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: effective use of theories of change and logic models* (1st ed.). San Francisco, CA: Jossey-Bass.
- [14] Gale, J., Grossman, S., Ryan, M., Newton, S., Usselman, M. (2018). Exploring curriculum implementation using self-report enactment checklists. Paper presented at National Association of Research in Science Teaching conference. (Atlanta, GA).
- [15] Guzdial, M., & Morrison, B. (2016). Growing computer science education into a STEM education discipline. *Communications of the ACM*, 59(11), 31–33. <https://doi.org/10.1145/3000612>
- [16] Haladyna. (2012). Developing and Validating Multiple-choice Test Items (3rd ed.). Routledge. <https://doi.org/10.4324/978020382594>
- [17] Halloun, I. A. and Hestenes, D. (1985). The Initial Knowledge State of College Physics Students. *American Journal of Physics*, 53, 1043.
- [18] Hammer, D. (1996). Misconceptions or P-Prims: How May Alternative Perspectives of Cognitive Structure Influence Instructional Perceptions and Intentions. *Journal of the Learning Sciences*, 5(2), 97–127. https://doi.org/10.1207/s15327809jls0502_1
- [19] Harel, I. and Papert, S. 1991. *Constructionism*. Ablex Publishing Corporation, New York, NY.
- [20] Howard, G.S. (1980). Response-shift bias: a problem in evaluating interventions with Pre/Post self reports. *Evaluation Review*, 4, 93–106.
- [21] Howard, G. S., Ralph, K. M. Gulanick, N. A., Maxwell, S. E., Nance, D., & Gerber, S. L. (1979). Internal invalidity in pretest-posttest self-report evaluations and the re-evaluation of retrospective pretests. *Applied Psychological Measurements*, 3, 1–23.
- [22] Hox, J. J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Erlbaum.
- [23] Hubbard, A. K. (2017). Learning to Teach Computer Science: Qualitative Insights into Secondary Teachers' Pedagogical Content Knowledge. Ph.D Dissertation. Northwestern University.
- [24] Hubwieser, P., Magenhein, J., Mühling, A., & Ruf, A. (2013). Towards a conceptualization of pedagogical content knowledge for computer science. Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research – ICER '13, 1. <https://doi.org/10.1145/2493394.2493395>
- [25] Huntley, M. A. (2012). Using concerns-based adoption model theory to develop tools to examine variations in mathematics textbook implementation. In D. J. Heck, K. B. Chval, I. R. Weiss, & S. W. Ziebarth (Eds.), *Approaches to studying the enacted mathematics curriculum*. Charlotte, NC: Information Age Publishing.
- [26] Hee-Sun Lee & Nancy Butler (2003) Making authentic science accessible to students, *International Journal of Science Education*, 25:8, 923–948, DOI: 10.1080/09500690305023
- [27] Knezek, G. and Christensen, R. 1996. Validating the Computer Attitude Questionnaire (CAQ). Presented at Annual Meeting of the Southwest Education Research Association (New Orleans, Louisiana, January 26, 1996). AMSER96.
- [28] Magerko, B., Freeman, J., McKlin, T., Reilly, M., Livingston, E., McCoid, S., & Crews-Brown, A. (2016). EarSketch: A STEAM-Based Approach for Underrepresented Populations in High School Computer Science Education. *ACM Transactions on Computing Education*, 16(4), 1–25. <https://doi.org/10.1145/2886418>
- [29] McKlin, T., Magerko, B., Lee, T., Wanzer, D., Edwards, D., & Freeman, J. (2018). Authenticity and Personal Creativity: How EarSketch Affects Student Persistence. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education - SIGCSE '18* (pp. 987–992). Baltimore, Maryland, USA: ACM Press. <https://doi.org/10.1145/3159450.3159523>
- [30] McKlin, T., Wanzer, D., Lee, T., Magerko, B., Edwards, D., Grossman, S., Freeman, J. (2019). Implementing EarSketch: Connecting Classroom Implementation to Student Outcomes. In *Proceedings of 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*, February 27–Mar. 2, 2019, Minneapolis, MN, USA. ACM, NY, NY, USA, 7 pages. <https://doi.org/10.1145/3287324.3287379>
- [31] Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, 24(3), 315–340. <https://doi.org/10.1177/109821400302400303>
- [32] Newton, Sunni H.; Gale, Jessica; Alemdar, Meltem; and Wind, Stefanie. (2018) "Assessment of Practice-Focused Middle School Science Modules". Interdisciplinary STEM Teaching & Learning Conference. (Savannah, GA) <https://digitalcommons.georgiasouthern.edu/stem/2018/2018/37>
- [33] Nunnally, J. C. (1964). *Educational Measurement and Evaluation*, New York: McGraw-Hill.
- [34] Piaget, J. and Inhelder, B. 1967, The Child's Conception of Space, New York: W.W. Norton.
- [35] Pratt, C., McGuigan, W. & Katzev, A. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation*, 21,341-349.
- [36] Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2009). The Astronomy and Space Science Concept Inventory: Development and Validation of Assessment Instruments Aligned with the K-12 National Science Standards. *Astronomy Education Review*, 8(1). <https://doi.org/10.3847/AER2009024>
- [37] Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, 50(5), 1020–1049. <https://doi.org/10.3102/0002831213477680>
- [38] Saldaña, J. (2013). The coding manual for qualitative researchers. Thousand Oaks, CA: Sage.
- [39] Shaffer, D.W., and Resnick, M. 1999. Thick Authenticity: New media and authentic learning. *Journal of Interact. Learn. Res.* 10, 2, 195-215.
- [40] Shinas, V. H., Yilmaz-Ozden, S., Mouza, C., Karchmer-Klein, R., & Glutting, J. J. (2013). Examining Domains of Technological Pedagogical Content Knowledge Using Factor Analysis. *Journal of Research on Technology in Education*, 45(4), 339–360. <https://doi.org/10.1080/15391523.2013.10782609>
- [41] Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- [42] Smith III, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences*, 3(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1
- [43] Turkle, S. (Ed.). (2008). *Falling for science: objects in mind*. Cambridge, Mass: MIT Press.
- [44] Wanzer, D., McKlin, T., Edwards, D., Freeman, J., & Magerko, B. (2019). Assessing the Attitudes Towards Computing Scale: A Survey Validation Study. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education - SIGCSE '19*, 859–865. <https://doi.org/10.1145/3287324.3287369>
- [45] Wiebe, E., Williams, L., Yang, K., and Miller, C. 2003. Computer science attitude survey. *Comput. Scie.* 14.
- [46] Yadav, A., & Berges, M. (2019). Computer Science Pedagogical Content Knowledge: Characterizing Teacher Performance. *ACM Transactions on Computing Education*, 19(3), 1–24. <https://doi.org/10.1145/3303770>
- [47] Yoon, Yoon, S., Evans, M. G., & Strobel, J. (2014). Validation of the Teaching Engineering Self-Efficacy Scale for K-12 Teachers: A Structural Equation Modeling Approach: Validation of the Teaching Engineering Self-Efficacy Scale. *Journal of Engineering Education*, 103(3), 463–485. <https://doi.org/10.1002/jee.20049>