Athens Journal of Sciences 2023, 10: 1-18
https://doi.org/10.30958/ajs.X-Y-Z

Assessment of Computational Thinking: A Study of Preservice Teachers' Knowledge and Beliefs

By Qing Li, Scot W. McNary± & Tracy Boyd°*

Aiming to bridge the gap in the field of computational thinking and computing education, this study examines preservice teachers' knowledge of and self-efficacy related to CT assessment. It also develops a reliable instrument to understand the construct of teacher assessment of CT. Specifically, the research questions focus on investigating the extent of preservice teachers' knowledge and level of self-efficacy of CT assessment, and identifying the difference between demographic variables on teacher knowledge and self-efficacy related to CT assessment. Adopting a cross-sectional survey design, the participants were 182 preservice teachers. Both descriptive statistics and Exploratory Factor Analysis were used to analyze the data. The results show that the preservice teachers know little about how to use CT assessment to help students, and they believe they know even less about using specific assessment techniques to accomplish assessment.

Keywords: *computational thinking, assessment, teacher knowledge, teacher self-efficacy*

Introduction

This study aims to bridge the gap identified in the field of computational thinking (CT) and computing education by investigating preservice teacher (hereafter teachers) knowledge of and self-efficacy related to CT assessment. CT has been recognized as a new "basic skill" that all, not just a few, k-12 students need to master. A national push on teaching CT for all learners has started, as evidenced by a growing number of states adopting policies to support CT education as well as the increased funding for rigorous CT teacher preparation and professional development (https://code.org/advocacy/state-facts/MD.pdf).

A challenge to implement this policy, however, is the lack of qualified teachers, as illustrated by various documents including the report from the national organization Code.org (https://code.org/files/2018_state_of_cs.pdf). Preparing in-service teachers and pre-service teachers are essential to meet demand. In response to the shortage of qualified teachers, some institutions have begun to develop CS/CT preparation programs for prospective k-12 teachers.

Yet, a review of relevant literature identifies a significant gap: there is a lack of research on effective assessment related to CT education, especially for preservice education and in-service professional development (Tang et al. 2020).

*Professor, Towson University, USA.
±Associate Professor, Towson University, USA.
°Doctoral Student, Towson University, USA.

Limited, if any, research exists investigating assessment of CT learning in connection to pre-service/in service teacher education. This gap hampers the efforts of developing CT teacher preparation programs because assessment is a critical aspect of education. Assessment not only provides data to inform instructional practices, but also offers stakeholders' needed evidences to justify a range of decisions from funding policy to placing learners in different programs (Fischer 2010, Thomas 2012).

On the other hand, teacher knowledge and beliefs, including their self-efficacy, significantly influence their behaviors and effectiveness (Bray-Clark and Bates 2003). It is well established in the literature that teacher self-efficacy is positively correlated to teaching methodologies and student achievement (Kaya et al. 2020). According to Bender and colleagues (Bender et al. 2016), CS/CT teachers face challenges, including but not limited to, teachers' beliefs about their insufficient qualification and lack of preparation, their lack of opportunities to collaborate with peers; high pressure to remain current with the ever changing technologies. These authors posit that teacher beliefs not only play an important role in addressing such challenges, but also should be carefully considered when developing CS/CT teacher preparation programs. Several studies, including a couple of literature review papers (Kallia 2017, Tang et al. 2020) which have analyzed hundreds of studies published from 2001-2019 focusing on assessment of computing courses, discovered that though some attention has focused on assessment in computing courses, there is a dire need for the advancement of knowledge on assessment in CS/CT teacher preparation.

Similarly, Yadav and colleagues (Yadav et al. 2015) have argued that understanding teachers' experience and beliefs related to CS/CT education is a critical first step towards the development of effective assessment strategies and approaches. The review of the existing literature shows the lack of a psychometrically sound instrument that assesses teachers' knowledge and self-efficacy of assessment connected to CT. This study, therefore, attempts to address the need by enhancing our understanding of computing education assessment focusing on teacher knowledge and beliefs. Another purpose of this work is to help develop a reliable instrument to understand the construct of teacher assessment of CT.

Related Literature

The importance of assessment is unquestionable. As a comprehensive data gathering and evaluation process, assessment provides teachers useful information to improve their teaching. To optimize the development of schooling and learning, assessment practices need to meet the highest standards (Jaipal-Jamani and Angeli 2017).

Teacher assessment knowledge embodies a synthesis of their content and pedagogical knowledge (Millay 2018). Teachers must realize that merely presenting information is not enough to ensure learning, and effective learning is a complex interplay of the teaching process and its outcomes (Bond 1995). When assessing learning, teachers are required to align specific goals, determine the extent to which anticipated outcomes are achieved, and make appropriate decisions accordingly (Gonzales and Callueng 2014). Further, better understanding of tools and strategies for assessment directly enhances teachers' development in general, because effective assessment produces more helpful data to improve teaching practices (Henze et al. 2008).

Merely equipping teachers with the needed knowledge and skills, however, do not guarantee successful assessment practice because their beliefs, including their self-efficacy, also largely affect their behaviors. Self-efficacy theory, originally developed by Bandura (1977), points to the predictive value of teachers' confidence in connection with the success of their performance. That is, whether a teacher would successfully apply certain knowledge or skills is greatly dependent on their beliefs in their competency to implement such skills.

Assessment and CT

Although CT has gained increased attention in recent years and numerous studies have published focusing on various aspects of CT education, less work has focused on assessment of CT. The research on assessments related to CT works to fill the gaps of student assessment. For instance, Relkin et al. (2020) designed "unplugged" CT assessments for students, aged five through nine. Similarly, Polat et al. (2021) researched the need for evaluating secondary school students through a range of perspectives to provide more comprehensive CT assessments. Some used standardized multiple-choice or performance assessment based on the analysis of students' developed coding artifacts (Mouza et al. 2017, Tang et al. 2020). Even though these niche areas beneficial, to effectively incorporate CT into a K–12 curriculum, it is necessary to provide teachers with guidance for how to assess it (Grover and Pea 2013). Additionally, teachers must develop personal CT competency for their professional work to certify achievement of learning objectives (Menon et al. 2019).

In a recent paper, Tang and colleagues (Tang et al. 2020) systematically reviewed 96 studies of CT assessment between 2010-2019. Focused on educational context, assessment constructs and types, and reliability and validity evidences, they found that only 15% of the studies examined teacher education, leading to their conclusion that more studies are needed at this level. They also found that all of these studies focused on examining and assessing teachers' understanding of CT and related beliefs, with virtually no study explored teacher knowledge of or beliefs about CT assessment. Another review study (Wang et al. 2021) focused on integrating CT in STEM education. Their semi-systematic literature review of 55 empirical studies on this topic showed that the assessment of student learning in CT integration into STEM subjects adopted different approaches with various objectives.

For example, Adler and Kim (2018) studied preservice teachers learning of CT via modeling and simulation in a science methods course. Their survey results showed that the preservice teachers learned CT and intended to integrate CT into their future teaching. Another study (Cetin 2016) examined the effect of Scratch-based instruction on preservice teachers' understanding of basic programming concepts and their attitudes toward program. Adopting a mixed method design, the results showed the Scratch-based instruction allowed the development of a meaningful learning environment to help preservice teachers understand basic computing concepts.

A related study (Jaipal-Jamani and Angeli 2017) examined the impact of robotics on preservice teachers. The 21 elementary preservice teachers took a science methods course in which robotics were integrated. The results showed a significant knowledge gain in science concepts and basic CT skills, and increased both the preservice teachers' interest in robotics and their self-efficacy in using robotics for instructional purposes.

Some studies examined the effects of different programming templates on preservice teachers' learning of CT. A study (Pala and Mıhçı Türker 2019) examined the impact of Arduino IDE and C++ programming languages on preservice teachers' knowledge. Using the CT Skills Scale survey, their results showed a knowledge gain in the Arduino IDE group in creativity, algorithmic thinking, critical thinking, but not problem solving or cooperativity factors. In addition, no gain was identified in the C++ group.

Another line of research is to automate the assessment of learner produced artifacts. Based on the framework of Brennan and Resnick (2012), several technology tools were developed, including *Hairball* and *Dr. Scratch* (Boe et al. 2013, Moreno-León and Robles 2015). These software templates enabled automatic evaluation of CT skills through analysis of learner generated programming products. Li and Pustaka (2020) examined the impact of educational game development experience on teachers' pedagogical and content knowledge related to CT. Using Dr. Scratch, they quantitatively analyzed the games created by 80 educators and concluded that game creation allowed teachers to develop an overall proficiency in CT skills.

Teacher Beliefs

Moving from teacher knowledge to teacher beliefs, though limited, some studies have examined preservice teachers' perceptions and others investigated effects of interventions on them, usually through surveys or interviews. These investigations have focused on teacher attitudes toward CT and CT teaching, as well as their confidence in teaching CT.

For instance, an Australian study (Bower and Falkner 2015) examined the CT related perceptions of 44 preservice teachers (33 females, 11 males) who were enrolled in an education course. The results showed that most preservice teachers had misconceptions, ranging from treating CT as general technology use like searching internet, to equating CT integrated teaching to using technology in classrooms. A majority of them lacked confidence in teaching CT, while several were overconfident with associated misunderstanding. They welcomed opportunities to gain content, pedagogical and technological knowledge related to CT teaching.

A Canadian study (Gadanidis et al. 2017) explored the experience of 143 preservice elementary teachers learning CT in a math education course. The blended nine-week, 18-hour course aimed to help teachers gain content and pedagogical knowledge related to mathematics teaching with CT. Through analysis of teacher reflection and online discussions, the case study found that teachers developed some new ideas related to CT and the course helped reduce teacher apprehension towards CT in mathematics teaching and learning.

A review study (Cabrera 2019) examined 24 existing papers and identified that teachers tended to hold preconceptions impacting their understanding of CT. Specifically, they often equate CT to one of the following concepts: 1) technology integration, 2) coding, 3) problem solving, and 4) "thinking like a computer". Additional teachers might believe that CT should not be part of k-12 education for reasons ranged from CT is too difficult to learn, to certain student groups could not manage to acquire such skill, to conflicts with the curriculum, to constraints such as time limitation and instructional structure of schools.

In summary, while CT has gained increasing attention from various groups, studies that focused on assessment in teacher education are still relatively limited. Amongst this limited exploration, even less, if any, work has examined teacher knowledge or beliefs related to assessment of CT. To bridge this gap, this study investigates the knowledge and self-efficacy of preservice teachers connected to CT assessment.

Research Questions

Grounded in constructivism, this study considers assessment beyond the idea of "assessment of learning". Rather, assessment is considered for learning, of learning, and as *learning*. Sound assessment, therefore, requires teachers' deep understanding and a high level of self-efficacy of the specific assessment tools, techniques and strategies. Specifically, this is guided by the following research questions:

1. To what extent do preservice teachers know about using CT assessment?
2. What level of self-efficacy do preservice teachers hold about CT assessment?
3. Are there differences between levels of salient demographic variables (i.e., gender, ethnic identity, program of study) on derived factor scores of teacher knowledge and self-efficacy related to CT assessment?

Methods

Participants

Adopting a cross-sectional survey design, the participants were 182 preservice teachers enrolled in a university in the mid-Atlantic region of the United States. Nearly three-quarters of respondents were females, and two thirds were White.

The average age among those who reported it was 21.6 (SD = 3.5, N = 168). One third of the participants enrolled in the secondary program. Among the participants, about 45% were juniors and close to half were the combination of sophomore and senior students.

There was some variety in concentrations represented although Social Studies (36%), English (31%), Special Education (30%), and Early Childhood (28%) were the most frequently endorsed. About 14% said "Yes" and another 36% said "Maybe" when asked whether they would be interested in teaching computer science after graduation. See Table 1 and Table 2 for a complete description of participants.

Table 1. *Participants' Demographics*

Student demographic variable	**N = 182(%)**
Gender	
Female	135 (75)
Male	41 (23)
Prefer not say/Missing	3 (2)
Race/Ethnicity	
White	105 (66)
Black	22 (14)
Hispanic, mixed, Native American	16 (10)
Asian	10 (6)
Other	5 (3)
Program	
Secondary	55 (33)
Special Education	36 (22)
Early Childhood	31 (19)
Elementary	26 (16)
Middle School	11 (7)
Early Childhood/Special Education	8 (5)
Year	
Freshman	6 (3)
Sophomore	37 (21)
Junior	80 (45)
Senior	47 (27)
Graduate/other	7 (4)
	N/Mean/SD
Age	168/21.59/3.53
Number of prior Math/Computer Science Courses	151/9.55/11.28

Table 2. *Participants' Characteristics*

Student characteristic	Yes (%)	No (%)	Maybe (%)
Interested in teaching CS/CT	8 (14)	39 (51)	21 (36)
Concentration			
-Social Studies	52 (36)	92 (64)	
-English	45 (31)	99 (69)	
-Special Ed	43 (30)	101 (70)	
-ECE	40 (28)	104 (72)	
-Math	19 (13)	125 (87)	
-Science	16 (11)	128 (89)	

Percentage may not add to 100% due to students with double concentrations.

Instrument & Analysis

Adapted from the work of Gonzales and Callueng (2014), a three-step procedure was used to develop this survey. First, based on the existing literature, the initial instrument was created. Second, a group of experts were asked to judge the suitability and accuracy of the questions, as well as the general organization of the instrument. Based on the feedback, we adjusted the instrument, eliminating inappropriate items and reorganizing remaining ones. One suggestion adopted was moving the demographic information section from the beginning to the end because research shows that putting them up front may trigger stereotype threat (Cohoon et al. 2011). Third, a group of five students who represented the targeted sample were asked to pilot the instrument. The data collected allowed us to further modify the instrument. The significant changes on the instrument concerned rewording and additional detail for some items to avoid confusion. This modified instrument became the final survey questionnaire used for the study.

The survey consisted of a total of 33 items. The first part, containing 10 questions, asked teacher knowledge about using CT assessment. The second part, with 16 questions, evaluated teachers' self-efficacy about assessment practices and strategies. The last part, a total of seven questions, asked about teachers' demographic information.

Means, SDs, and frequency of selected responses were computed to address the research questions one and two of this study. To answer the research question three, factor analysis was conducted to determine the structural characteristics of the measure. All analyses were conducted using R [Version 4.0.3; R Core Team (2021)] and the R-package psych [Version 2.0.12; Revelle (2020)].

Results

CT Assessment Knowledge

The first research question focused on teachers' knowledge of CT assessment. This was investigated from two perspectives: 1). Teacher knowledge about using CT assessment and 2). Their knowledge of specific types of CT assessment.

First, pre-service teachers were asked how much they knew about using assessment of CT knowledge for various purposes. The scale used to record responses was: 1 = no knowledge 2 = little knowledge 3 = some knowledge 4 = know well 5 = know very well. On average, survey respondents responded they did not know how to do any of the tasks well as means were below 2.0 ("little knowledge") for all items, and no more than 12% of the sample claimed they knew any task well/very well (see Table 3).

Table 3. *Pre-service Teachers Knowledge about CT Assessment (N = 182)*

How much do you know about using CT assessment to:	Mean	SD	Know well or Very Well *n* (%)
1. guide students to set their goals and monitor their own CT learning progress?	1.98	0.94	9 (5.0)
2. demonstrate to students how to do self-assessment of their CT learning?	1.92	0.95	8 (4.5)
3. determine how students can learn CT on their own in class?	1.86	0.94	11 (6.1)
4. help students develop clear criteria of good CT learning?	1.89	0.96	11 (6.1)
5. set the criteria for students to assess their own performance related to CT in class?	1.86	0.97	13 (7.3)
6. measure extent of CT learning at the end of a lesson or subject?	1.92	1.04	18 (10.2)
7. make final decisions about the level of CT learning that students achieved at the end of a lesson or subject?	1.85	1.01	15 (8.4)
8. help students improve their CT learning process and class performance?	1.99	1.05	17 (9.5)
9. assist students to determine their CT learning strengths and weaknesses in class?	1.97	1.06	20 (11.2)
10. identify better CT learning opportunities for students in class?	1.93	1.09	16 (8.9)

Note: 1 = no knowledge 2 = little knowledge 3 = some knowledge 4 = know well 5 = know very well.

A second set of questions concerned specific types of assessment methods used to assess CT ability and knowledge. The same 1-5 response scale used for CT Assessment Knowledge was used for these items. As indicated in Table 2, the means are quite low—none exceed 2 ("little knowledge") on the five-point response scale. Only a small percentage of teachers (4-8% "know well" or "very well") claimed knowledge of any of the six specific assessment techniques inquired about. What is most notable is that for all six assessment methods, respondents' mean response was less than the CT Assessment knowledge items. That suggests that respondents believe their knowledge about specific techniques is less than their knowledge about general use of assessment for CT. Put another way, although respondents felt they knew little about how to help students assess their own CT learning, they believe they know even less about using specific assessment techniques (e.g., Artifact Based Interviews) to accomplish assessments. This suggests there is room to develop knowledge about all six assessment techniques. See Table 4 for details.

Table 4. *Preservice Teachers Knowledge about Types of Assessment (N = 182)*

How much do you know about:	Mean	SD	Know Well or Very Well *n* (%)
1. using Artifact-Based Interviews to assess students' CT practices?	1.76	0.94	11 (6.1)
2. creating appropriate Artifact-Based Interviews questions in order to assess students' CT practices?	1.75	0.93	10 (5.6)
3. using Design Scenarios to assess students' CT practices?	1.77	0.92	9 (5.1)
4. creating appropriate Design Scenarios in order to assess students' CT practices?	1.77	0.91	10 (5.6)
5. using Learner Documentation to assess students' CT practices?	1.82	1.01	14 (7.8)
6. creating appropriate Learner Documentation in order to assess students' CT practices?	1.77	0.94	8 (4.5)

Note: 1 = no knowledge 2 = little knowledge 3 = some knowledge 4 = know well 5 = know very well.

Self-Efficacy of CT Assessment

The second research question aimed to examine teachers' beliefs, specifically their self-efficacy, about CT assessment. Preservice teachers were asked to judge their own ability to develop and implement CT assessment. The scale use for these items ranged from 1 = Strongly disagree to 5 = Strongly agree. A score of 3 indicated a Neutral response to the item. All but two of the items had means less than 3, suggesting that the average response was to disagree. Items such as "I can develop appropriate CT assessment plans" (M = 2.25 SD = 1.12) and "I can create good rubrics for CT assessment" (M = 2.29 SD = 1.12) were the lowest rated, indicating general disagreement with these claims and therefore relatively lower efficacy for these tasks compared to the others. These items relate to planning for CT assessment and are similar to the findings about setting criteria for students' learning in Table 2 above. The two items for which respondents produced an average exceeding 3 were reverse-scored items ("I do not know how to construct objective tests of CT" and "I am not confident in reporting CT assessment results") and 54% and 39% of the sample agreed with these statements, respectively. These responses are consistent with the remaining items suggesting that many respondents do not feel capable of developing and implementing assessments of students CT learning. One item that was reverse-scored, "I am not good at scoring and marking CT tests and assessment tools" resulted in a mean of less than 3, suggesting a relative strength for at least some respondents. Of all the tasks, respondents felt most efficacious about scoring and marking tests, yet still about 30% of the sample agreed they were not good at this task. See Table 5 for details.

Table 5. *Respondent Efficacy Judgement for Specific CT Assessment Tasks (N = 182)*

Rate your degree of agreement with this statement:	Mean	SD	Agree or strongly agree *n* (%)
1. "I can write good learning outcomes of CT."	2.36	1.09	27 (15.2)
2. "I do not know how to construct objective tests of CT."*	3.38	1.36	96 (53.9)
3. "I can define tasks for performance assessment (i.e. assess students by asking them to perform tasks) of CT."	2.38	1.12	30 (16.8)
4. "I can choose the most appropriate item type (e.g. multiple choice, true/false) for a CT test."	2.66	1.26	54 (30.2)
5. "I can ask essay questions for CT assessment."	2.35	1.17	35 (19.7)
6.:"I can create good rubrics for CT assessment."	2.29	1.12	29 (16.3)
7. "I can develop appropriate CT assessment plans."	2.25	1.12	28 (15.6)
8. "I am not good at scoring and marking CT tests and assessment tools."*	2.90	1.29	55 (30.7)
9. "I can link learning outcomes with CT assessment processes."	2.31	1.09	27 (15.1)
10. "I am not confident in reporting CT assessment results."*	3.18	1.35	70 (39.1)

Note: 1=Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree. * indicates a reverse-scored item.

Demographic Variables, Knowledge & Self-Efficacy

The third research question examined possible differences between demographic variables on teacher knowledge and self-efficacy related to CT assessment. The 26 items across the three proposed scales were submitted to an Exploratory Factor Analysis (EFA). Preliminary measures of factorability suggested item intercorrelations could be factored; the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) was 0.93 for the correlation matrix (0.70 is conventionally considered adequate) and MSAs for items ranged between 0.68-0.97; Bartlett's test of Sphericity was statistically significant, X^2 (325) = 5668.90; $p < 0.001$, implying the correlation matrix was not an identity matrix and therefore presumably containing at least one factor. The scree plot (Figure 1), parallel analysis (Figure 2), and Velicer's MAP suggested that four factors were reasonable to extract, and all four had eigenvalues that exceeded 1.0, the Kaiser criterion. Collectively, these four factors explained 69% of the total variance. The first three factors explained 63%. That is, the factor analysis produce 3 factors that fit the data well.

Figure 1. *Scree Plot*

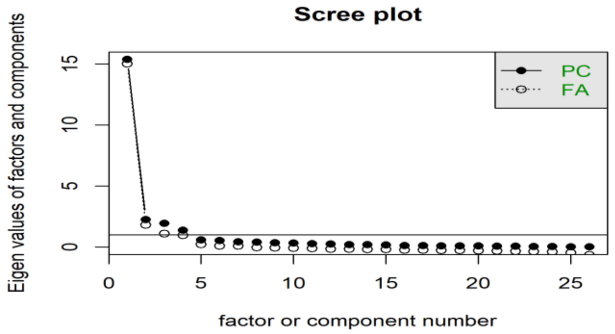
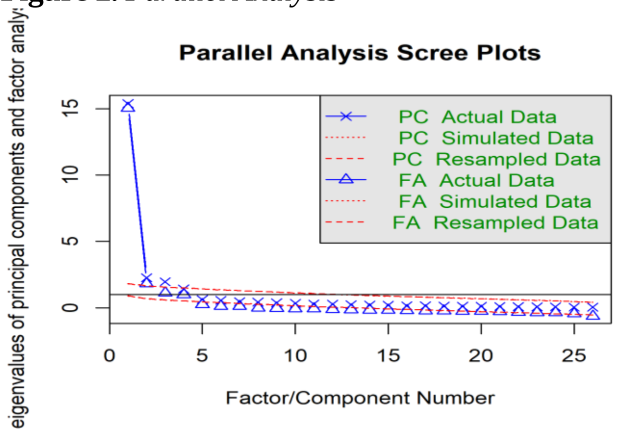


Figure 2. *Parallel Analysis*



The four factors were rotated obliquely via Oblimin rotation which allows factors to correlate among themselves. Loadings for the four factors on each variable are shown in Table 6 and a path diagram of the measurement model is shown in Figure 3.

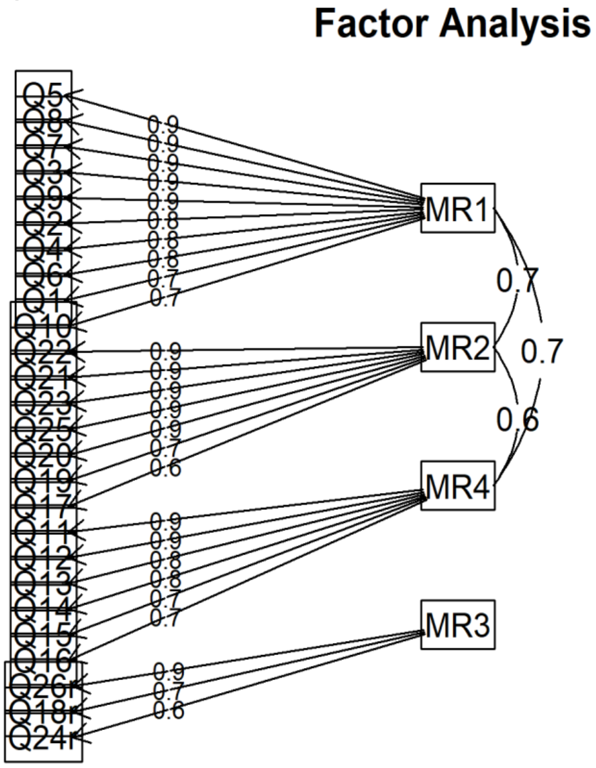
Table 6. *Factor Loadings for Four Factors*

Item	Assessment Knowledge	Assessment Technique Knowledge	Assessment Self-Efficacy	Reverse-scored Assessment Self-Efficacy items	communality
Q1[1]	0.74		0.16		0.69
Q2[1]	0.86				0.71
Q3[1]	0.87	-0.16	0.14		0.75
Q4[1]	0.83				0.76
Q5[1]	0.91				0.85
Q6[2]	0.79				0.81
Q7[2]	0.88				0.79
Q8[3]	0.90				0.84
Q9[3]	0.87				0.85
Q10[3]	0.68		0.21		0.82
Q17	0.24	0.65			0.76
Q19	0.20	0.68			0.77
Q20		0.88			0.82
Q21		0.92			0.82
Q22		0.92			0.79
Q23		0.90			0.77
Q25		0.90			0.68
Q11	-0.11		0.91		0.50

Q12			0.93		0.65
Q13			0.84		0.78
Q14	0.15		0.76		0.84
Q15	0.20		0.69		0.84
Q16	0.19		0.67		0.88
Q18r				0.69	0.44
Q24r				0.67	0.80
Q26r				0.90	0.80
Eigenvalue	7.94	5.48	4.62	1.80	
Cronbach's alpha	0.97	0.96	0.88		
Factor correlations					
Assessment Technique Knowledge	0.65				
Assessment Self-Efficacy	0.73	0.59			
Reverse-Scored Assessment Self-Efficacy	0.21	0.17	0.21		

Note: superscripts[1,2,3] refer to factors in Gonzales and Callueng (2014).

Figure 3. *Measurement Model*



Communalities were high (> 0.50) for all but items Q11 and Q18, suggesting common factors explain more than half the items variance for almost all items. The factor loadings show that the conceptual grouping of the items was very similar to the empirical grouping of them. Where cross-loadings occurred there were small (≤ 0.20; see italicized loadings in Table 5), and other loadings on factors exceeded 0.60. There were three exceptions to the otherwise tidy factor structure, and that was the clustering of the three reverse-scored Assessment Efficacy items. These items did not correlate very highly with the other factors and seem to be related mainly because of the form of the questions "I do not know…", "I am not good at….", "I am not confident…". For this reason, these items were not used to create the Assessment Efficacy factor and only the seven items that remained in that scale were used.

Factor scores were computed so that comparisons of sub-groups of respondents could be conducted. Factor score means for specific subgroups are shown in Table 7. Sample sizes vary considerably between subgroups reducing power for many comparisons, and significance tests are of questionable value for comparisons among groups in a study whose purpose was not to conduct group contrasts. Instead, group comparisons would be more useful once conducted on a new sample designed to represent demographic balance in the desired population. However, it may be useful to note several patterns of subgroup differences.

Table 7. *Factor Scores by Student Group*

		Assessment Knowledge			Assessment Technique Knowledge			Assessment Efficacy		
Group1	n	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Gender										
Male	41	1.95	0.90	2.00	1.87	0.98	1.50	2.46	1.10	2.57
Female	135	1.89	0.89	1.80	1.73	0.81	1.50	2.35	1.00	2.43
Prefer not to say/Other	3	2.10	1.04	1.60	2.06	1.18	1.83	2.24	0.54	2.00
Program										
Early childhood	31	2.07	1.04	1.70	1.83	0.93	1.67	2.42	0.89	2.43
Elementary	26	1.90	0.86	2.00	1.69	0.68	1.58	2.54	0.88	2.57
Middle school	11	2.49	0.76	2.40	2.48	0.92	2.17	3.09	0.86	3.43
Secondary	55	1.76	0.93	1.30	1.75	0.92	1.33	2.35	1.16	2.43
Special education	36	1.91	0.77	2.00	1.66	0.77	1.33	2.13	0.93	2.07
Early Childhood/Special Education	8	1.39	0.55	1.10	1.31	0.46	1.00	1.77	1.15	1.07
Stage										
Freshman	6	1.98	1.02	1.65	1.94	1.00	1.75	2.55	1.05	2.43
Sophomore	37	1.61	0.76	1.20	1.53	0.63	1.17	2.32	0.96	2.43
Junior	80	2.20	0.92	2.10	1.96	0.93	1.71	2.68	0.96	3.00
Senior	47	1.74	0.80	1.60	1.67	0.83	1.25	1.97	1.02	1.71
Other	1	1.00	NA	1.00	1.00	NA	1.00	1.86	NA	1.86
Graduate	6	1.48	0.85	1.20	1.58	0.88	1.33	1.79	1.18	1.29
Ethnicity										
Asian	10	1.66	0.86	1.20	1.52	0.78	1.00	2.34	1.18	2.21
Black	22	2.09	1.08	1.90	1.83	0.89	1.50	2.49	1.07	2.43
Hispanic	9	2.00	1.05	2.00	1.83	0.95	1.33	2.21	0.78	2.29
Mixed	5	1.76	1.11	1.20	1.67	0.92	1.00	2.29	1.23	2.57
Native American	2	2.05	1.34	2.05	2.33	1.18	2.33	3.64	0.51	3.64
Other	5	2.32	1.43	2.00	2.37	1.47	2.00	2.69	1.54	3.71
White	105	1.85	0.81	1.80	1.70	0.82	1.33	2.34	1.00	2.43
Interested in Teaching Math/Comp Sci?										
Yes	8	2.38	1.20	2.30	2.50	1.33	2.92	2.45	1.20	2.93
No	30	1.80	0.80	1.65	1.92	0.88	1.92	2.40	1.16	2.64
Maybe	21	1.77	0.90	1.30	1.74	0.89	1.50	2.58	1.20	2.86
Overall	182	1.92	0.89	2.00	1.76	0.85	1.50	2.37	1.01	2.43

It appears that means for males exceeded females for all three factors, and middle school program respondents exceeded all other respondents for all three factors. That is, on average, males, compared to females, reported higher level of Assessment Knowledge, Assessment Technique knowledge, and self-efficacy.

Similarly those enrolled in middle school programs, compared to those in other programs, believed they had higher levels of knowledge and efficacy. Junior respondents' mean scores exceeded other respondents' mean scores for the Assessment Knowledge and Assessment Technique Knowledge factors, and all but Freshman respondents (n = 6) for the Assessment Efficacy factor.

Non-white respondents reported more knowledge and efficacy than White and Asian respondents, e.g., Black > Hispanic > White > Asian for the Assessment Knowledge factor; Black and Hispanic respondents' mean > White > Asian respondent's means for the Assessment Technique Knowledge factor; Black respondent means > White = Asian > Hispanic respondent means for the Assessment Efficacy factor.

Respondents who acknowledged they are interested in teaching Math/ Computer science classes reported higher means than those not interested, or who said they may be for the Assessment Knowledge and Assessment Technique Knowledge factors. Respondents who only maybe interested in teaching Math/ Computer Science reported larger Assessment Efficacy factor means than others. The sample size for respondents answering this question was fewer than half (n = 59) of all responses so it is difficult to interpret this pattern.

Table 8 shows correlations between the factor scores and respondent age and number of college credits in math or computer science coursework. Correlations are small to moderate in size, and statistically significant for the math/computer science coursework variable and Assessment Knowledge, $r(149) = 0.21$, $p < 0.05$, and Assessment Efficacy, $r(149) = 0.19$, $p < 0.05$. It shows that the higher number of math and or computer science credit hours, the more likely the participants believed they had more assessment knowledge and high levels of self-efficacy of CT assessment.

Table 8. *Correlations of Age and Hours in Computer Classes with Factor Scores*

	N	**Assessment Knowledge**	**Assessment Technique Knowledge**	**Assessment Efficacy**
Age	168	0.11	0.12	0.08
Math/Comp Sci. credit hours	151	0.21*	0.15	0.19*

Note: *p < 0.05.

Discussion and Conclusions

Despite the growing interest in CT education, studies on assessment of CT, especially related to teacher preparations is scarce. The current study addresses this gap through the examination of teacher knowledge and self-efficacy related to CT assessment. Outcomes from this research provide valuable information to deepen our understanding of effective ways to assess computing education. This can also inform teacher education in designing appropriate CS/CT education programs as well as useful assessment tools.

Shedding lights on the field of assessment in relation to CS/CT education, this study have several results worthy of further discussion. The most significant finding perhaps is that a vast majority of the teachers have limited knowledge and low self-efficacy related to CT/CS assessment. On average, the teachers do not know how to do any of the CT assessment tasks, only one in ten teachers claim they know any of the assessment tasks well. With respect to specific types of and self-efficacy related to CT assessment, the results are similar. In other words, teachers know little about how to use CT assessment to help students, and they believe they know even less about using specific assessment techniques to accomplish assessment. This is astonishing because it demonstrates how our preservice teachers are not prepared to integrate CT and computing into their practice.

Various researchers (Barr and Stephenson 2011, Li 2021) have argued that since CT is a problem solving skill used in all disciplines, CT should be integrated in all subjects rather than just in stand-alone CS courses. Thus, most, if not all, teachers should be prepared to integrate CT into their subject areas (Barr and Stephenson 2011, Yeni et al. 2021). Yet our results show that a vast majority of our preservice teachers are not equipped with the needed knowledge and skills to do so. It is shocking, but perhaps not surprising. Assessment has long been treated as a less pressing topic to teach in teacher education, usually because of the tight schedule of teacher preparation programs. Similarly, CT/CS training is usually not part of preservice teacher programs owning largely to its resent introduction. Preservice teachers' lack of knowledge and low level of self-efficacy in CT assessment suggests unless changes in teacher preparation of CS/CT instruction are made, they will struggle to make instructional decisions that align with best practices.

Another important contribution of this study comes from the results of the exploratory factor analysis. As discussed earlier, there is a dire need of reliable instruments to measure the construct of teacher assessment of CT. Our results show that the tool we developed is a psychometrically sound instrument to understand teacher knowledge and self-efficacy of assessment connected to CT which addresses the gap identified in the existing literature. The ten Assessment Knowledge items were adapted from Gonzales and Callueng (2014). In their EFA, the ten items were spread across three factors, unlike the current findings in which all ten Assessment Knowledge items cohered on one factor. However, the five items on the Assessment Knowledge factor share with Gonzales and Callueng's (2014) Assessment as learning factor have similarly high factor loadings and high communalities. There are at least four reasons for why there is diverge in factor loading among the remaining five items: the first reason could be the adaptation of the items from a general assessment context to the Computational Thinking assessment context. Secondly, the samples differ in age and background: Gonzales and Calleung's sample consisted of practicing teachers whereas the current sample is completely pre-service teachers. Thirdly, cultural differences between teacher preparation, classroom experience, and educational systems between Filipino teachers and US teachers may result in different experiences. Fourth, each survey appeared with different items for different purposes and the context of the survey may invoke different response styles and knowledge schema. Some combination of these reasons may also apply, thereby producing different item intercorrelations between the samples.

This study has practical implications as well. As argued by Yadav and colleagues (Yadav et al. 2017), teacher education programs provide an opportune time to help teachers develop competencies to embed CT in their future classrooms. In addition, previous work (Leonard et al. 2018) has demonstrated that carefully designed instructional practice can increase teachers' self-efficacy beliefs in relation to CT knowledge. Given the importance of assessment in teaching and learning of any subject, CS/CT included, our results suggest that it is imperative that preservice teachers are offered opportunities to learn the needed knowledge and skills related to CT assessment. Recognizing how busy preservice teacher programs are, we recommend that assessment to be integrated into the existing coursework rather than standalone assessment courses. For instance, the methods courses and the existing educational technology courses in teacher education programs are a natural fit for exposing CT assessment knowledge to teachers.

Like any educational research, this study has its own limitations. First, the sample is preservice teachers who have very limited classroom experiences. Future research are recommended to examine other groups including in service teachers. Secondly, the data is collected from students enrolled a university located in an urban city. Generation of the results to other populations need to take cautious.

Although CT education has gained some traction in educational research, the exploration of CT assessment related to teachers are extremely limited. Since assessment plays such a critical role in teaching and learning, it is essential that researchers pay close attention to this area. Until we have a deep understanding of CT assessment and teacher education, we can start designing and developing programs that best prepare our teachers to teach CT.

References

Adler RF, Kim H (2018) Enhancing future K-8 teachers' computational thinking skills through modeling and simulations. *Education and Information Technologies* 23(4): 1501–1514.

Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84(2): 191–215.

Barr V, Stephenson C (2011) Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? *ACM Inroads* 2(1): 48–54.

Bender E, Schaper N, Caspersen ME, Margaritis M, Hubwieser P (2016) Identifying and formulating teachers' beliefs and motivational orientations for computer science teacher education. *Studies in Higher Education* 41(11): 1958–1973.

Boe B, Hill C, Len M, Dreschler G, Conrad P, Franklin D (2013) Hairball: Lint-inspired static analysis of scratch projects. Paper presented at the proceedings of the *44th ACM Technical Symposium on Computer Science Education*.

Bond LA (1995) *Critical issue: rethinking assessment and its role in supporting educational reform.* Oaks Brooks, IL: North Central Regional Education Laboratory.

Bower M, Falkner K (2015) Computational thinking, the notional machine, pre-service teachers, and research opportunities. Paper presented at the proceedings of the *17th Australasian Computing Education Conference (ACE 2015*).

Bray-Clark N, Bates R (2003) Self-efficacy beliefs and teacher effectiveness: Implications for professional development. *Professional Educator* 26(1): 13–22.

Brennan K, Resnick M (2012) New frameworks for studying and assessing the development of computational thinking. Paper presented at the proceedings of the *2012 Annual Meeting of the American Educational Research Association*. Vancouver, Canada.

Cabrera L (2019) Teacher preconceptions of computational thinking: a systematic literature review. *Journal of Technology and Teacher Education* 27(3): 305–333.

Cetin I (2016) Preservice teachers' introduction to computing: exploring utilization of scratch. *Journal of Educational Computing Research* 54(7): 997–1021.

Cohoon J, Cohoon JM, Soffa ML (2011) Focusing high school teachers on attracting diverse students to computer science and engineering. Paper presented at the *2011 Frontiers in Education Conference (FIE)*.

Fischer TB (2010) *The theory and practice of strategic environmental assessment: towards a more systematic approach*: Routledge.

Gadanidis G, Cendros R, Floyd L, Namukasa I (2017) Computational thinking in mathematics teacher education. *Contemporary Issues in Technology and Teacher Education* 17(4): 458–477.

Gonzales RD, Callueng CM (2014) Classroom assessment practices of Filipino teachers: measurement and impact of professional development. *Essentials on Counseling and Education: A Festchrift of Prof. Rose Marie Salazar-Clemena. Manila: Asian Psychological Services and Assessment (APSA)*, 220–242.

Henze I, van Driel JH, Verloop N (2008) Development of experienced science teachers' pedagogical content knowledge of models of the solar system and the universe. *International Journal of Science Education* 30(10): 1321–1342.

Jaipal-Jamani K, Angeli C (2017) Effect of robotics on elementary preservice teachers' self-efficacy, science learning, and computational thinking. *Journal of Science Education and Technology* 26(2): 175–192.

Kallia M (2017) *Assessment in computer science courses: a literature review*. Available at: https://royalsociety.org/~/media/policy/projects/computing-education/assessment-literature-review.pdf.

Kaya E, Newley A, Yesilyurt E, Deniz H (2020) Measuring computational thinking teaching efficacy beliefs of preservice elementary teachers. *Journal of College Science Teaching* 49(6).

Leonard J, Mitchell M, Barnes-Johnson J, Unertl A, Outka-Hill J, Robinson R, et al. (2018) Preparing teachers to engage rural students in computational thinking through robotics, game design, and culturally responsive teaching. *Journal of Teacher Education* 69(4): 386–407.

Li Q (2021) Computational thinking and teacher education: an expert interview study. *Human Behavior and Emerging Technologies* 3(2): 324–338.

Li Q, Pustaka A (2020) Computational Thinking and teacher education: can digital game building help? *Athens Journal of Sciences* 7(2): 89–114.

Menon D, Romero M, Viéville T (2019) Computational thinking development and assessment through tabletop escape games. *International Journal of Serious Games* 6(4): 3–18.

Millay L (2018) *Teachers' professional knowledge and formative assessment practices: an empirical study from middle school earth science instruction in the context of an educational improved community*. Master's Thesis Dissertation. University of Maine.

Moreno-León J, Robles G (2015) Analyze your Scratch projects with Dr. Scratch and assess your computational thinking skills. Paper presented at the *Scratch Conference*.

Mouza C, Yang H, Pan Y-C, Ozden SY, Pollock L (2017) Resetting educational technology coursework for pre-service teachers: a computational thinking approach to the development of technological pedagogical content knowledge (TPACK). *Australasian Journal of Educational Technology* 33(3): 61–76.

Pala FK, Mıhçı Türker P (2019) The effects of different programming trainings on the computational thinking skills. *Interactive Learning Environments* 29(7): 1090–1100.

Polat E, Hopcan S, Kucuk S, Sisman B (2021) A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology* 52(5): 1965–1980.

Relkin E, de Ruiter L, Bers MU (2020) TechCheck: development and validation of an unplugged assessment of computational thinking in early childhood education. *Journal of Science Education and Technology* 29(4): 482–498.

Tang X, Yin Y, Lin Q, Hadad R, Zhai X (2020) Assessing computational thinking: a systematic review of empirical studies. *Computers & Education* 148: 103798.

Thomas M (2012) Teachers' beliefs about classroom assessment and their selection of classroom assessment strategies. *Journal of Research and Reflections in Education* 6(2): 104–115.

Wang C, Shen J, Chao J (2021) Integrating computational thinking in STEM education: a literature review. *International Journal of Science and Mathematics Education* 20(Nov): 1949–1972.

Yadav A, Gretter S, Good J, McLean T (2017) Computational thinking in teacher education. In *Emerging Research, Practice, and Policy on Computational Thinking*, 205–220. Springer.

Yadav A, Burkhart D, Moix D, Snow E, Bandaru P, Clayborn L (2015) *Sowing the seeds: a landscape study on assessment in secondary computer science education*. Available at: https://www.csteachers.org/documents/en-us/78193875-4ed9-42ea-8224-841d13da1f74/1.

Yeni S, Grgurina N, Hermans F, Tolboom J, Barendsen E (2021) Exploring teachers' PCK for computational thinking in context. Paper presented at the *16th Workshop in Primary and Secondary Computing Education*.