“I Impressed Myself With How Confident I Felt”: Reflections on a Computer Science Assessment for K-8 Teachers

Hannah E. Chipman
Rhodes College
Memphis, Tennessee
chihe-20@rhodes.edu

Fernando J. Rodríguez
University of Florida
Gainesville, Florida
fjrodriguez@ufl.edu

Kristy Elizabeth Boyer
University of Florida
Gainesville, Florida
keboyer@ufl.edu

ABSTRACT

The computer science education community has made great strides in promoting diversity and inclusion in computing fields and bringing K-12 CS learning opportunities to broader groups of learners. However, one area that has not been investigated fully is how assessments can influence a learner’s confidence and attitudes towards CS. Stereotype threat and test modality have been shown to affect the performance of CS test takers. This experience report examines how novice CS learners respond to a CS assessment by investigating an increasingly important group of novice CS learners: K-8 classroom teachers. We conducted focus groups with elementary and middle school teachers as part of a week-long CS professional development workshop. The focus groups were held after teachers completed pre- and post-assessments. The assessment instrument featured multiple-choice and short-answer questions with block-based programming snippets. Many teachers reported a positive disposition towards learning CS after completing the pre-assessment, which they attributed to having a growth mindset. Themes related to their confidence involved the difficulty and format of the assessment, with comments about difficulty reducing after the post-assessment. When asked about their thoughts on the assessment from the perspective of their students, they provided suggestions with particular attention to its format. These findings provide insight for CS assessment design and implementation, as well as support further research on the impact of assessments on CS learners.

CCS CONCEPTS

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

KEYWORDS

Professional development; K-8 teachers; assessment

ACM Reference Format:
Hannah E. Chipman, Fernando J. Rodríguez, and Kristy Elizabeth Boyer. 2019. “I Impressed Myself With How Confident I Felt”: Reflections on a Computer Science Assessment for K-8 Teachers. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE ’19), February 27-March 2, 2019, Minneapolis, MN, USA.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3287324.3287478

SIGCSE ’19, February 27-March 2, 2019, Minneapolis, MN, USA

ACM ISBN 978-1-4503-5890-3/19/02…$15.00
https://doi.org/10.1145/3287324.3287478

1 INTRODUCTION

The need to recruit and retain diverse students in computer science is as high as ever. Research on computer science outreach and the role of recruitment interventions has focused on promoting a diverse and inclusive workforce [1, 10, 13], which in turn leads to broader perspectives and new innovations. However, there are still low numbers of women and students from underrepresented groups enrolling in computer science programs [5, 6, 9, 20, 21] due to a wide variety of issues. However, the effects that *knowledge assessments* can have on students’ attitudes towards the field have not been fully explored. Studies have shown that test modality and stereotype threat can affect underrepresented groups’ performance on computer science examinations [17, 24]. These factors can make computer science tests a deterrent for these underrepresented groups who already self-assess their science, technology, engineering, and math abilities as lower than that of non-minority males [16].

In order to better understand the relationship between CS assessments and attitudes towards CS, we conducted a study with an increasingly important population of novice CS learners: classroom teachers in K-8. Many teachers in the US are being asked to learn computer science and then integrate it into their classroom activities, but most teachers have no formal computer science background. To prepare them, the CS education community relies primarily on professional development workshops and other activities. Research has found that professional development workshops are effective in increasing teachers’ and students’ confidence in their coding abilities [14].

We conducted this study in the context of a one-week computer science professional development workshop for K-8 teachers. We designed CS assessments to evaluate how well the teachers learned the CS concepts presented to them. Teachers completed the assessments at the start and end of the week-long workshop and participated in focus groups in which we asked them about their experience taking the assessment and their attitudes toward CS. Their focus group responses were analyzed qualitatively through the use of thematic analysis, in which themes were assigned to their responses. We found that teachers expressed high confidence in their ability to learn CS after both the pre- and post-assessments. When they did report low confidence, they attributed it to the assessment’s difficulty and format. The teachers also provided insight into how the assessments could be improved for student test takers, with specific recommendations about the assessments’ presentation and content. We conclude with a set of practical implications for assessment design and implementation in the classroom. These findings support further research on the impact of assessments on CS learners.

2 RELATED WORK

One of the overarching motivations behind our research is broadening participation in computer science, as there is a significantly low number of women and people of color in computer science and related fields. In 2017, only 18.1% of computer science degrees were awarded to women, while around 9.75% of degrees were awarded to either Black or Hispanic people [11]. There are many barriers to computer science for historically underrepresented groups, including lack of exposure, low self-efficacy, stereotypes about computer scientists, and the societal belief that males are better at science, technology, engineering, and math [2, 16]. We hope to contribute to the literature on this topic by studying the perceptions of a diverse group of teachers and their attitudes toward a CS assessment.

A potential barrier for minorities in computer science that has not been fully explored is CS assessments. How assessment items are presented can significantly impact the performance of women and people of color. In a study of the summative assessment for the AP Computer Science Principles course, code questions were presented to students in either text or block form [24]. The researchers found that students performed better on questions that presented code in a block-based language over questions that showed code in a text-based language. This result was significant for students identifying with underrepresented groups in the field of computer science. A possible explanation discussed in the paper is that some students may be intimidated by text-based languages, which would link the modality to social and cultural influences.

Another factor that can influence how well a person does on a test is stereotype threat, which is when a person's performance is influenced by a known stereotype [19]. In a study of undergraduates in a computer science course, students who were asked for demographic information before a test scored lower on the subsequent computer science exam [17]. By informing and evaluating the design of computer science assessments that reduce bias and stereotype threat, we can ultimately reduce their negative effect on students' attitudes and confidence in computer science.

The assessments designed for teachers in this study were based on literature in block-based assessments. Most prior work has focused on assessing the knowledge of children participating in outreach programs. For example, the Computational Thinking Test (CTt) is a multiple-choice exam in which the questions and answers are presented in block form, focusing on loops and code sequencing [22]. The CTt was tested for convergent validity against assessments that used Bebras Tasks and Dr. Scratch and was found to be partially convergent with both. The K-2 assessment questions, which used code blocks that show directions like north, east, south, and west with the goal of reaching an end point, were based on CTt questions and were meant to demonstrate knowledge of program sequencing. Another study used a block-based test featuring multiple-choice, true/false, and open-ended questions covering topics such as loops, variables, and conditionals to assess what concepts students are learning with block-based curricula [12]. This assessment led us to use short-answer questions in both the K-2 and 3rd-8th assessments, such as questions that ask "What will this program say?" and request a short open answer. The teachers that participated in our workshop learned to program using block-based programming languages, with the goal of using these languages to teach their own students. Therefore, our assessment presented code in a block-based format, drawing from the prior work on block-based assessments.

3 TEACHER CS WORKSHOP STUDY

We conducted the present study in the context of a week-long computer science professional development workshop for K-8 teachers with little to no prior experience in computer science and programming. The workshop's goal was to provide teachers with knowledge and hands-on practice in CS concepts and how to incorporate them into their teaching. It consisted of five 7-hour sessions conducted during the span of a week and included introductions to block-based programming languages and opportunities for the teachers to design activities and lesson plans for their K-8 classrooms.

3.1 Participants

A total of 40 K-8 teachers participated in this professional development, recruited from flyers distributed at several schools in our local county in the Southeastern United States. One teacher did not complete any of the surveys or assessments administered during the workshop, so that teacher's data is excluded from this study. We consider the remaining teachers in two cohorts: K-2 and 3rd-8th grade teachers. As discussed below, these teachers experienced different content and were given different versions of the pre- and post-test. There were 18 teachers in the K-2 cohort and 21 in the 3rd-8th cohort. Table 1 describes the demographic information of these cohorts. Teachers had an average of 10.8 years of teaching experience, with the 3rd-8th teachers specializing in particular subjects: math and science (11), language arts (3), social studies (1), and multiple subjects (6).

3.2 Assessment Design

The goal when designing the assessment was to measure what the teachers would be learning during the professional development workshop, so we carefully aligned the assessment with the workshop's curricular plan. We designed one assessment for each grade cohort based on the programming language each cohort would be using. The two assessments were designed to be similar in question type and difficulty[1]. The workshop activities were designed around block-based languages that the teachers could then directly use in their courses to teach their own students. The workshop content for the K-2 cohort used resources from Code.org's Course A [7], which featured block languages with minimal text and focused on activities that involved moving objects within 2-dimensional environments. The assessment questions drew from examples in the Code.org Course A curriculum and featured the same block images from these materials in the questions and answer choices. The 3rd-8th cohort used Snap*!* in their workshop coding activities [15], and activities were inspired by the Beauty and Joy of Computing curriculum [23]. Assessment questions for these teachers featured Snap*!* code images in their question and answer choices. To assess learning, the same assessment was administered as a pre-assessment and post-assessment.

[1]Both assessment instruments can be found at the following link: https://cise.ufl.edu/research/learndialogue/teacher-workshop/assessments/

Table 1: Teacher demographic information by cohort

Demographic Information	K-2	3rd-8th	Total
Female	17	19	36
Male	1	2	3
Black/African American	3	7	10
Hispanic/Latino	2	2	4
White/Caucasian	13	12	25
Total	18	21	39

In order to better fit into the workshop's schedule, the assessment items consisted of mostly multiple-choice questions with five options, as recommended by prior research on assessment design [4]. Approximately half of the questions were short-answer questions, which were included to get a more accurate assessment of their knowledge, reducing the effects of guessing the right answer on the multiple-choice questions. The K-2 assessment had 14 questions (7 short-answer questions), and the 3rd-8th assessment had 17 questions (7 short-answer questions). Both assessments covered topics on code sequencing, conditionals, loops, and variables. The K-2 assessment focused more on programming tasks related to directional movement in a 2-dimensional space, while the 3rd-8th assessment featured questions on conditions for displaying text or numeric output and conceptual questions (see Figure 1).

We aimed to word questions in a gender-neutral way and avoided questions that could be interpreted as referring to a particular gender, as recommended by prior work on stereotype threat in assessments [17]. We also attempted to avoid questions that required knowledge of more than simple math procedures and science, so as not to elicit negative stereotypes related to science, technology, engineering, and math abilities. Instead, questions were created with the teachers and their students in mind, focusing on what they would find useful, such as the grade calculator question in Figure 1.

3.3 Procedure

The assessments were administered once at the beginning and once more at the end of the workshop. In order to understand the assessment's impact on teachers' understanding of CS concepts, as well as their thoughts on computing and related materials, we conducted focus groups with the teachers after each assessment. The teachers were introduced to their interviewers and split into eight focus groups of three to five participants each, with each group only consisting of teachers from the same grade cohort. The following were the questions asked during the focus groups following the pre-assessment:

- *You just completed a pre-assessment, which we did not expect you to know the answers to. How did you feel about that assessment, in terms of difficulty, content, or format?*
- *How confident do you feel in your ability to succeed in this workshop?*
- *What final thoughts and suggestions do you have regarding this assessment?*

The following were the questions asked during the focus groups following the post-assessment:

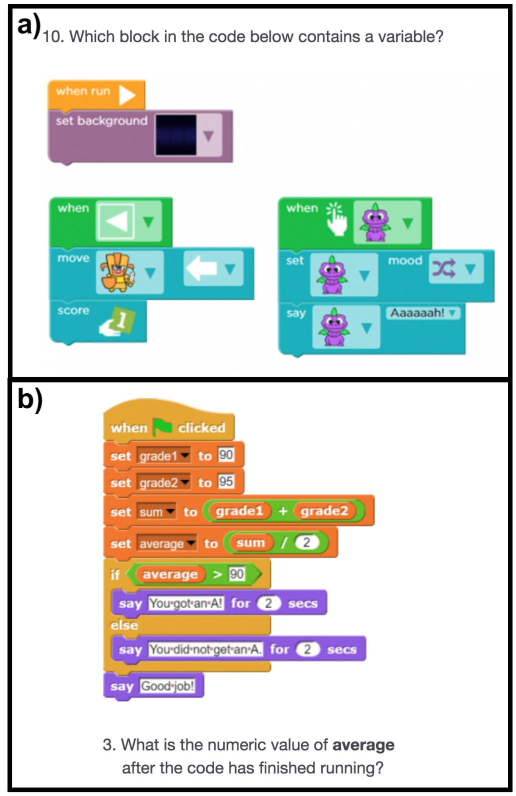

Figure 1: Example assessment items from a) the K-2 assessment and b) the 3rd-8th assessment

- *What are your thoughts regarding the assessment's representation of the topics covered in the workshop?*
- *After taking the assessment the second time, how are you feeling about your computer science abilities?*
- *Now that you are familiar with CS, what recommendations do you have about the assessment?*
- *Research has shown that some assessments have implicit bias. How do you think we could reduce stereotype threat and bias with this assessment?*
- *What final thoughts do you have regarding this assessment?*

3.4 Thematic Analysis

The focus groups were video or audio recorded with permission from the participants, and the recordings were transcribed for further analysis. We used the transcripts to identify emerging themes in the teachers' responses for both focus groups following the pre- and post-assessments. Themes were derived from the most commonly used words in teachers' responses. Each response was assigned a single broad theme based on its content, which was either a new theme or a pre-existing theme that had been assigned to a previous response. After assigning themes to each response, we reviewed the responses once more to assign a subtheme to each response based on the details provided. One focus group was excluded from this analysis because the post-assessment recordings for that focus group were inadvertently not collected.

4 RESULTS

4.1 Assessment Scores

Of the 40 participants, 34 completed both the pre-assessment and post-assessment for their respective grade cohort (14 from the K-2 cohort, 20 from the 3rd-8th cohort). Each item in the assessments was assigned one point value, resulting in a total score of 14 for the K-2 assessment and 17 for the 3rd-8th assessment. The K-2 assessment scores were not normally distributed while the 3rd-8th assessment scores were, according to Shapiro-Wilk tests. The average K-2 pre-assessment score was 60.4% (SD = 19%), and this score increased to an average of 79.1% (SD = 10.6%) for the K-2 post-assessment, showing a significant increase in score from pre- to post-assessment (Wilcoxon signed rank test: $p = 0.0047$; effect size: Cohen's $d = 1.21$). For the 3rd-8th cohort, the average pre-assessment score was 39.4% (SD = 20.1%), increasing to an average of 71.8% (SD = 12.5%) in the post-assessment, also showing a significant increase in score from pre- to post-assessment (t-test: $p < 0.0001$; effect size: Cohen's $d = 1.93$).

4.2 Focus Group Themes Following Pre-Assessment

Table 2 shows a summary of the main themes and subthemes identified from the focus groups following the pre-assessment. The most common theme was **confidence**, mentioned 34 times. Responses in this theme related to questions about how confident the teachers were feeling about their ability to succeed in the professional development workshop. We identified four subthemes for the responses classified under this confidence theme: *confident in ability to succeed* (10 times), *not confident in ability to succeed* (7 times), *neutral confidence in ability to succeed* (3 times), and *positive disposition to learning* (17 times). This last subtheme is distinct in that it refers to responses in which a teacher expressed low confidence in the CS subject matter but a willingness to learn and improve her knowledge, as exemplified by the following quote:

> *"I think you can only go up. At least for me I'm starting with absolutely no knowledge, and so I know that by the end of the week I'm going to have this, all this knowledge. So right now, I'm just sitting here waiting to absorb it and get hands-on practice with it. So right now, I do not have confidence in my ability, but I have confidence in that I'm going to learn this week and be able to do it by the end."* (3rd-8th cohort, 53% on pre-assessment)

The next theme, mentioned 26 times, was that of **question difficulty**. Responses in this theme mentioned the level of effort involved in answering the assessment questions, as well as strategies for completing the assessment. Most teachers mentioned *guessing and using logic* (17 times) as their main strategy, *"I felt like I could guess on some of them through logic, but I wasn't very confident in the answers"* (K-2 cohort, 62% on pre-assessment). Other teachers simply mentioned the *challenging difficulty* (6 times) of the questions, with a few specifically referencing the use of *CS-specific terminology* (3 times) as a source of frustration.

The final theme we identified from the focus group responses related to the assessment's **format** and presentation, and it was mentioned 20 times. Some teachers praised the use of *block-based program images* (9 times) as part of the questions, which helped them more easily visualize the questions. A few teachers indicated a preference towards *fitting all questions into a single window* (5 times) to avoid unnecessary scrolling. Others commented on the *look and feel* (3 times) of the assessment, specifically referencing how selected responses were perceived as being incorrect because they would be highlighted in red, *"We both agree the background color when you click the choice should not be in 'wrong' red"* (K-2 cohort, 46% on pre-assessment). Finally, there were a few mentions about the *variety of question types* (2 times), specifically mentioning the benefits of having different question types.

4.3 Focus Group Themes Following Post-Assessment

Table 3 shows the main themes and subthemes identified from the focus groups following the post-assessment, which was held at the end of the workshop. Once again, the main theme of **confidence** emerged from the teachers' responses to the focus group questions, mentioned 43 times. However, the subthemes were slightly different than in the pre-assessment. The subthemes of *confident in ability to succeed* (17 times) and *not confident in ability to succeed* (3 times) returned, but we also found mentions of *understanding CS concepts* (23 times). The following quote illustrates this subtheme:

> *"Like, I understood the language. The first time I was like, 'what is this?', but this time I'm like 'oh OK,' like I know exactly what they're asking and what the code is saying."* (3rd-8th cohort, 82% on post-assessment)

Teachers also commented on the **format** of the assessments, mentioning it 21 times. Most of their responses involved the survey tool that was being used, referencing the *look and feel* (9 times) and the *variety of question types* (9 times), similar to the responses found in the pre-assessment. Additionally, teachers made comments about the *number of questions* (3 times), mostly relating to insufficient time to cover all of the questions, as described here:

> *"I think there could have been less questions. Even at the beginning, 17 is kind of intimidating. I think it could've been like 10 questions. I think we would've been more comfortable with that. I think we were kind of anxious about 17 and it taking 25 minutes."* (3rd-8th cohort, 65% on post-assessment)

During the focus groups following the post-assessment, we asked teachers for their thoughts on **avoiding bias** (16 times) and for additional suggestions. Some teachers felt that the assessment had *no biased language* (8 times), with mentions on how it does not use gendered pronouns or cultural references. Regarding suggestions, teachers mentioned how the assessment required *English reading and math skills* (5 times) to be completed properly, how *seating arrangements* (2 times) can influence a test taker's confidence if they are sitting next to a highly skilled student, and how students with *limited access to computers* (1 time) could find it more difficult to practice concepts at home.

The last emerging theme from the focus groups following the post-assessment was the **content** of the assessment, mentioned 15 times. In particular, teachers mentioned how the assessment properly covered the topics presented during the workshop, *"I feel like we actually learned what we were assessed on"* (K-2 cohort).

Table 2: Focus group themes following pre-assessment

Theme	Subtheme	Example
Confidence	Positive disposition to learning	*"I have confidence in that I'm going to learn this week and be able to do it by the end."*
	Confident	*"Still confident. I just feel like we'll be able to pick it up."*
	Not confident	*"It made me a little more cautious."*
	Neutral confidence	*"I don't know whether or not I would not want to stop once I got working on a program..."*
Difficulty	Guessing and using logic	*"Yeah, I was guessing and trying to figure out what it all meant."*
	Challenging difficulty	*"It was very difficult, I don't know. When I start reading through it I understand what I don't know and what I need help in."*
	CS-specific terminology	*"Also, some of the language that they use, you have to know that to get through it."*
Format	Block-based program images	*"I agree with the pictures, that definitely helped. Without that, I'd be like 'I don't know."'*
	Fitting questions to window	*"Sometimes it was so long that you couldn't see the box... so you had to scroll."*
	Look and feel	*"I don't think it should be red because it's like 'oh that's wrong."'*
	Variety of question types	*"I thought the format was good 'cause it had like multiple choice and typing answers."*

Table 3: Focus group themes following post-assessment

Theme	Subtheme	Example
Confidence	Understanding concepts	*"I understand. The first time it was kind of like a second language."*
	Confident	*"Yes, I was more confident in my answer choices and I wasn't just randomly guessing."*
	Not confident	*"... it would take a lot more for me to really be able to master it."*
Format	Look and feel	*"The red background wasn't a surprise this time."*
	Variety of question types	*"I thought the assessment was good in that it wasn't all multiple choice."*
	Assessment length	*"I think we were kind of anxious about 17 and it taking 25 minutes."*
Avoiding Bias	No biased language	*"It maintained the language of computer science, it didn't say he/she."*
	Reading and math skills	*"I mean obviously you have to have a pretty good grasp of English."*
	Seating arrangements	*"Me sitting next to a male technology savvy, even though I knew him really well, was really intimidating."*
	Limited computer access	*"A lot of students don't have computers so it's gonna be kind of hard for them to do things at home."*
Content	Coverage of topics	*"I think it was really good, like it covered everything we learned in our group."*

5 DISCUSSION

The most common theme across both focus groups following the pre-and post-assessments was confidence. While some teachers expressed low confidence after the pre-assessment, many expressed a positive disposition towards learning more about computer science, regardless of how they felt they performed on the assessment. To that effect, one teacher said, *"I feel like I can learn. I feel like we all have a growth mindset."* (K-2 cohort, 62% on pre-assessment). People with a growth mindset are more likely to view obstacles as learning opportunities rather than as a reflection of their intellectual ability [18]. In a study of undergraduates in an introductory computer science course, researchers found that fostering students' growth mindset led to better performance on assessments compared to students who were not introduced to the concept of a growth mindset [8]. Having this growth mindset may have been the reason for the teachers' self-reported confidence after both the pre-assessment and the post-assessment. The significant improvement in post-assessment score, as well as the reduction of low confidence responses from the focus groups following the post-assessment, echoes prior research on the benefits of having a growth mindset. The increased confidence of the teachers after the post-assessment is exemplified in the following quote:

> *"I was just, I impressed myself with how confident I felt going in taking it and how I could definitively quickly 'yes' this, 'yes' that instead of hemming and hawing about it."* (K-2 cohort, 77% on post-assessment)

When asked to elaborate on the reasons for their lack of confidence in their performance on the pre-assessment, teachers referenced the difficulty of the assessment questions. Some teachers resorted to guessing, *"I felt like I could guess on some of them, through logic"* (K-2 cohort, 62% on pre-assessment), and others mentioned the use of CS-specific terminology as a source of frustration, *"I think there was a lot of terminology that was questionable"* (K-2 cohort, 46% on pre-assessment). They also commented on the format of the assessment and how its presentation could be potentially frustrating to test takers. In particular, teachers expressed a preference towards keeping all relevant question information in a single screen of the assessment, *"... if they have to flip back and forth between the reading passage and the question, they're too conscious about having to find where they were again"* (K-2, 50% on pre-assessment).

Subthemes relating to the assessment's difficulty were not prominent in the focus groups following the post-assessment. Instead, teachers focused more on the format of the assessment and how to improve the test-taking experience overall. This change in theme can be attributed to the professional development workshop's goals of introducing the teachers to CS concepts and encouraging them to reflect on how to integrate CS into their own classrooms. The teachers' newfound familiarity with the CS concepts reduced the difficulty of the assessment, turning their attention instead to the assessment's presentation.

Some positive teacher opinions on the pre-assessment included the use of block-based program images in the question text and the variety of question types. The use of block-based program code in assessment questions has been shown to help improve scores over questions that use traditional code, with the effect being larger on female students and students from underrepresented minorities in CS [24]. One teacher commented on how the block-based program images helped her deduce the answers to the questions, *"...without pictures I wouldn't have been able to figure it out"* (K-2 cohort, 77% on pre-assessment).

In the focus groups following the post-assessment, teachers were asked to identify potential sources of bias within the assessment to get their perspective on ways the assessment could be unintentionally affecting the confidence or performance of test takers. One teacher discussed how her sitting next to a male teacher who was very knowledgeable in the subject intimidated her and decreased her self-esteem, citing stereotypes related to her gender and ability to succeed in STEM. The following quote presents this concern:

> *"Like in our group, there were four of us from the same school, and me sitting next to a male technology savvy, even though I knew him really well, was really intimidating for the whole week."* (3rd-8th cohort, 59% on post-assessment)

In response, another teacher suggested splitting the teachers into cohorts based on skill level. This suggestion is similar to results from prior research on pair programming, which found that students perform better on computer science exams when they work with partners of similar skill level [3].

5.1 Practical Implications

The results presented here highlight some practical considerations for assessing computer science learning:

- When making a digital assessment, make sure each question fits onto one screen so the learners do not need to scroll.
- Tell learners that a pre-test will have vocabulary they are not familiar with, and that is okay. Ask them to do their best.
- Ensure the number of questions is appropriate for the allotted time, taking into account the unfamiliar vocabulary and tendency of many learners to spend time reasoning through items they do not already know how to answer. Good estimates can be obtained by timing volunteers completing your assessment and adjusting the question number accordingly.
- For learners who may not already hold a strong growth mindset, explain the idea of a growth mindset before and/or after assessments and reinforce it throughout your course.
- Test the functionality of online software for administering surveys if this software is to be used for assessments. Its default color choices for user selections (sometimes red) may appear to suggest an incorrect answer.
- Avoid using gendered language in test items.
- Avoid using unnecessarily complex mathematical or scientific content in test items.
- Carefully consider seating arrangements to mitigate the risk that a student who works more slowly or has less prior experience will feel intimidated by a student who works more quickly or has more prior experience. Set classroom norms around what to do when some students finish early.

5.2 Limitations

Our findings have several limitations. We did not formally validate our assessments, but rather they were defined based on our prior experiences in computer science education research and topics covered in the workshop. Additionally, teachers were asked their thoughts on the assessments from their perspectives as K-8 teachers. These opinions may not directly correspond with those of K-8 students, since the teachers may already have stronger preconceived notions about computer science and assessments in general.

6 CONCLUSION

Despite the great strides the community has made in outreach and recruitment, research on how assessments can affect underrepresented groups in computer science is limited. With this study, we explored how our learners, K-8 teachers with very little prior computing experience, responded to an assessment designed as part of a professional development workshop. The assessment featured multiple-choice and short-answer questions, accompanied by block-based program images. The teachers shared their reflections on the assessment, as well as recommendations for improving its design, in focus groups conducted after both the pre- and post-assessment. We identified emerging themes and subthemes, and teachers' responses were classified based on these themes. We found that teachers expressed high levels of confidence after taking both the pre- and post-assessments, which some attributed to their growth mindset. Recommendations for the assessment included accommodations for test takers with low reading comprehension and math skills, limiting the number of questions shown at once, and paying closer attention to the look and feel of the assessment tool.

There are several directions to consider for future work. It is important to conduct similar studies in a classroom where the assessment score factors into the course grade. One teacher remarked they were not being graded during the workshop, so the pressure to perform well on the assessment was much lower, which could have influenced the confidence of the test takers. It is also important for long-term lines of research to continue to investigate assessments in the context of young learners who are still deciding on their career paths. Findings from this research can be used to inform further research on how to improve the test-taking experience for computer science learners.

ACKNOWLEDGMENTS

We would like to thank the LearnDialogue Group for their helpful input. This material is based upon work supported by the National Science Foundation under grants DRL-1640141 and DRL-1721160. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Lecia J. Barker, Charlie McDowell, and Kimberly Kalahar. 2009. Exploring Factors that Influence Computer Science Introductory Course Students to Persist in the Major. In *ACM SIGCSE Bulletin*, Vol. 41. ACM, 153–157.

[2] Skyler J. Bock, Lindsay J. Taylor, Zachary E. Phillips, and Wenying Sun. 2013. Women and Minorities in Computer Science Majors: Results on Barriers from Interviews and a Survey. *Issues in Information Systems* 14, 1 (2013), 143–152.

[3] Grant Braught, John MacCormick, and Tim Wahls. 2010. The Benefits of Pairing by Ability. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*. ACM, 249–253.

[4] Philip Sheridan Buffum, Eleni V Lobene, Megan Hardy Frankosky, Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester. 2015. A Practical Guide to Developing and Validating Computer Science Knowledge Assessments with Application to Middle School. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, 622–627.

[5] Sapna Cheryan, Victoria C. Plaut, Caitlin Handron, and Lauren Hudson. 2013. The Stereotypical Computer Scientist: Gendered Media Representations as a Barrier to Inclusion for Women. *Sex Roles* 69, 1-2 (2013), 58–71.

[6] Sapna Cheryan, Sianna A. Ziegler, Amanda Montoya, and Lily Jiang. 2017. Why are some STEM fields more gender balanced than others? *Psychological Bulletin* 143, 1 (2017), 1–35.

[7] Code.org. 2018. Course A. https://studio.code.org/s/coursea-2018.

[8] Quintin Cutts, Emily Cutts, Stephen Draper, Patrick O'Donnell, and Peter Saffrey. 2010. Manipulating Mindset to Positively Influence Introductory Programming Performance. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*. ACM, 431–435.

[9] Wendy DuBow, Joanna Weidler-Lewis, and Alexis Kaminsky. 2016. Multiple Factors Converge to Influence Women's Persistence in Computing: A Qualitative Analysis of Persisters and Nonpersisters. In *2016 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. 1–7.

[10] Katrina Falkner, Claudia Szabo, Dee Michell, Anna Szorenyi, and Shantel Thyer. 2015. Gender Gap in Academia: Perceptions of Female Computer Science Academics. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 111–116.

[11] National Center for Science and Engineering Statistics (US) (NCSES). 2017. Women, minorities, and persons with disabilities in science and engineering. (2017).

[12] Shuchi Grover and Satabdi Basu. 2017. Measuring Student Learning in Introductory Block-Based Programming: Examining Misconceptions of Loops, Variables, and Boolean Logic. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. ACM, 267–272.

[13] Margaret Hamilton, Andrew Luxton-Reilly, Naomi Augar, Vanea Chiprianov, Eveling Castro Gutierrez, Elizabeth Vidal Duarte, Helen H. Hu, Shoba Ittyipe, Janice L. Pearce, Michael Oudshoorn, and Emma Wong. 2016. Gender Equity in Computing: International Faculty Perceptions and Current Practices. In *Proceedings of the 2016 ITiCSE Working Group Reports*. ACM, 81–102.

[14] Karla Hamlen, Nigamanth Sridhar, Lisa Bievenue, Debbie K. Jackson, and Anil Lalwani. 2018. Effects of Teacher Training in a Computer Science Principles Curriculum on Teacher and Student Skills, Confidence, and Beliefs. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, 741–746.

[15] Brian Harvey and Jens Mönig. 2010. Bringing "No Ceiling" to Scratch: Can One Language Serve Kids and Computer Scientists?. In *Constructionism*. 1–10.

[16] Catherine Hill, Christianne Corbett, and Andresse St. Rose. 2010. *Why so Few? Women in Science, Technology, Engineering, and Mathematics*. American Association of University Women.

[17] Amruth N. Kumar. 2012. A Study of Stereotype Threat in Computer Science. In *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education*. ACM, 273–278.

[18] Laurie Murphy and Lynda Thomas. 2008. Dangers of a Fixed Mindset: Implications of Self-Theories Research for Computer Science Education. In *ACM SIGCSE Bulletin*, Vol. 40. ACM, 271–275.

[19] Hannah-Hanh D. Nguyen and Ann Marie Ryan. 2008. Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence. *Journal of Applied Psychology* 93, 6 (2008), 1314.

[20] Bureau of Labor Statistics. 2018. Employment Projections: Employment by detailed occupation, 2016 and projected 2026.

[21] Bureau of Labor Statistics. 2018. Labor Force Statistics from the Current Population Survey: Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity.

[22] Marcos Román-González, Jesús Moreno-León, and Gregorio Robles. 2017. Complementary Tools for Computational Thinking Assessment. In *Proceedings of International Conference on Computational Thinking Education (CTE 2017), S. C Kong, J Sheldon, and K. Y Li (Eds.). The Education University of Hong Kong*. 154–159.

[23] Lawrence Snyder, Tiffany Barnes, Dan Garcia, Jody Paul, and Beth Simon. 2012. The First Five Computer Science Principles Pilots: Summary and Comparisons. *Inroads* 3, 2 (2012), 54–57.

[24] David Weintrop, Heller Killen, and Baker Franke. 2018. Blocks or Text? How Programming Language Modality Makes a Difference in Assessing Underrepresented Populations. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS 2018)*. 328–335.