Measuring Student Performance and Readiness in Algorithms and Data Structures Courses

PAULA GOOLKASIAN, Psychological Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA
KALPATHI SUBRAMANIAN and ERIK SAULE, Computer Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA
JAMIE PAYTON, Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA

Studying the factors that can lead to student success and persistence has been the subject of a large body of CS education research. Our earlier work on tools to engage students in Algorithms and Data Structures courses demonstrated uneven benefits, especially among underrepresented students. Using a national pool of students from 13 colleges/universities in Algorithms and Data Structures courses, we used pre and post online surveys to assess the psychometric characteristics of a revised measure of prerequisite proficiency together with 4 others (confidence in computing, precollege computing experiences, skill with computer applications and course engagement) to test for gender and ethnic differences and to investigate the factors that would be useful in predicting final course grade. We were able to confirm differences in prerequisite proficiency; however, differences were a characteristic of additional factors such as precollege experiences, self-rated skill levels and confidence in computing. Regression analysis showed these differences can be mitigated by directing efforts at multiple components, rather than just prerequisite proficiency. Measures taken in the post-survey at the end of the semester showed a reduced number of demographic differences, when compared to the pre-survey. Prerequisite proficiency and performance scores were found to be significant predictors of course grades.

CCS Concepts: • **Theory of computation** → **Sorting and searching**; • **Social and professional topics** → **Computer science education**; **Race and ethnicity**; **Gender**;

Additional Key Words and Phrases: Preparation gap, algorithms, data structures, precollege experiences, pre-survey, post-survey, gender, ethnicity

ACM Reference format:
Paula Goolkasian, Kalpathi Subramanian, Erik Saule, and Jamie Payton. 2026. Measuring Student Performance and Readiness in Algorithms and Data Structures Courses. *ACM Trans. Comput. Educ.* 26, 4, Article 63 (May 2026), 19 pages.
https://doi.org/10.1145/3803796

This work was supported by grants from the National Science Foundation (Grant Nos. 1726809, 2142381).
Authors' Contact Information: Paula Goolkasian, Psychological Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA; e-mail: pagoolka@charlotte.edu; Kalpathi Subramanian (corresponding author), Computer Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA; e-mail: krs@charlotte.edu; Erik Saule, Computer Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA; e-mail: esaule@charlotte.edu; Jamie Payton, Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA; e-mail: jamie.payton@njit.edu.

ACM 1946-6226/2026/5-ART63
https://doi.org/10.1145/3803796

1 Introduction

As detailed in a systematic literature review by Hellas et al. [12], CS education research has focused on studying factors that have led to or predict student success. Both academic and non-academic factors have been studied and analyzed. Broadly, they relate to course activities, academic performance, learning features, and assessment. Wilson et al. [31] looked at 12 such factors to predict student success, and found that comfort level and math background had positive influence on success, while attribution to luck and game playing had a negative influence. Lishinski et al. studied self-efficacy, interest and belongingness in a CS1 course [17]. They analyzed data throughout the course period to understand students' "momentary experiences" and compared underrepresented students with their peers. Underrepresented students' self-efficacy correlated with final grades, but interest was negatively correlated with grades. Hooshangi et al. [13] performed a large study in a CS2 course to understand student performance and persistence. They confirmed earlier results on the strong correlation between CS1 and CS2 grades, while gender and race showed no significance. A number of researchers have looked at similar factors targeted towards course readiness, such as student experiences with computing [21], precollege experiences [1], course engagement [11, 26], and confidence/comfort with computing [30].

An additional approach to understanding student success in CS, particularly with regard to underrepresented students, can also be found by looking at the work of researchers on *Intersectionality.* This is an analytical approach defined by Cole [9] which considers multiple categories of social group membership simultaneously to extract meaning and consequences. Rather than build complex designs and large samples, Cole [9] proposed asking three questions (who is included in the category, the role of inequality, and what are their similarities) and she argues that the answers may lead to a conceptual shift in understanding that social group. Ireland et al. [14] reviewed literature on Black women and girls in STEM education using 60 sources. They studied a number of different themes, including (1) identity, (2) STEM interest, confidence, persistence, (3) achievement, ability, and (4) support systems. They provide both research and practical implications of their study and detail opportunities for intersectionality to advance educational research. Lunn et al. [18] used a multi-institutional database to understand participation (enrollment and graduation rates) of marginalized groups in computer science. They looked at participation trends over time in order to explain shifts based on historical events and activities. They provide guidelines for participation, such as peer support and mentoring opportunities, culture specific support, and supporting returning students for upskill/reskill goals. Smith [27] looked at the under-representation of women in STEM and points to Cheng's approach of considering each person in a continuum [7], ranging from congressive (interdependence) to ingressive (independent), rather than a simple gender binary. However, this brings challenges in terms of requiring new instruments to assess a student's place in this continuum, as well as developing learning outcomes appropriate to this classification. Rankin et al. [25] explored intersectionality through the lived experiences of Black women at different stages (undergrad, grad, early career professionals) to understand their persistence in computing. Initial results indicated how gendered racism can negatively impact Black women recruitment and retention in computing.

In a related series of studies, researchers focused on prerequisite proficiency for students in Algorithms and Data Structures courses [22, 23]. They worked on developing a concept inventory, which could be used as the basis of a pretest to measure how much students learned from the required series of first year courses in CS, primarily CS1 and CS2. Many instructors had noticed that in spite of successfully passing prerequisite CS courses, there was a considerable difference in course readiness among their students. Parker et al. [22] developed and validated a language free assessment with multiple choice items in pseudocode. Kraus-Levey et al. [16] used it as a pretest to investigate gender and ethnic differences in prerequisite course proficiency. Their findings showed small differences for women but sizeable differences for Non-Asian minorities. Cheng et al. [8, 10] revised the concepts to properly assess prerequisite concepts and developed a language free 14-item measure, a beginning step to develop a standardized assessment of what students learn in introductory CS courses.

For good reason, a significant amount of work has focused on early programming courses such as CS1, as it is the entry point to the major; however, sophomore level core courses such as Algorithms and Data Structures that constitute the theoretical aspects of computer science have also had a history of high attrition and low persistence. Secondly, few studies have gone beyond a single institution, course or semester. An exception to this is the replication study by Zingaro et al. [32] that spanned five institutions to understand the relationship between achievement goals vs grades and interest. They found that mastery-based goals are strongly correlated with grades and interest in CS courses. Normative and appearance goals were not correlated with grades and interest, in contrast to earlier work with psychology courses. However, the review of six mastery-based models by Szabo et al. [29] details some caution in their implementation, calling for additional resources (unlimited time, student support), that need to be taken into account.

Hellas et al. [12] in their review highlight several weaknesses in prior research for predicting student success. They encourage sharing datasets to help replication of published work, the use of multiple contexts (across multiple institutions, multiple semesters) for more robust studies, use of well-studied metrics, measuring reliability/validity of the metrics used to measure success, which helps with data quality and repeatability.

Our purpose in this article is to build on previous research efforts in studying the factors that play a role in student preparedness for Algorithms and Data Structures courses. Our *specific aims are to mitigate some of the weaknesses in past research identified by Hellas et al. [12] and to develop a better understanding of gender and ethnic differences by adopting the analytical approach of intersectionality toward social groups.* This approach was motivated from our previous efforts [2, 5, 6, 19, 20, 28] with developing tools to engage students in Algorithms and Data Structures courses. When testing student responses to the tools, we found that they were not perceived by all CS students equally and that underrepresented groups derived less benefits, despite spending more time completing the assignments. For example, in comparison to White and Asian men, there were many more negative comments from women and Black/Hispanic men. Both groups expressed less interest in the assignments, which were perceived as confusing and needing more documentation. Given the large variations in the sizes of the ethnic/gender groups and the very limited numbers of Non-Asian minorities participating in the surveys, the findings needed to be replicated with larger group sizes before definitive conclusions can be made. The results of this pilot study are available in the supplementary material associated with this manuscript.

Our specific research questions are whether some of the preliminary demographic differences that we observed in our previous sample of Algorithms and Data Structures students may be due to a preparation gap and whether that gap is related to grades. We extend previous research in two significant ways. First, we have a national (USA) sample of students who use the BRIDGES infrastructure for homework programming assignments in Algorithms and Data Structures courses from a variety of institutions (community colleges, 4-year colleges, and universities) rather than a single institution [8, 16, 22]. Collecting student responses anonymously with online Qualtrics surveys across an academic year, we obtained a diverse and representative sample of CS students that were sufficient to compare male and female responses as well as ethnic differences among White, Asian, Black, and Hispanic groups.

Secondly, we adopted a comprehensive approach to understand which of the factors identified in past research may underlie the preparation gap and whether some of the factors may show gender and ethnic differences. In addition to measuring prerequisite proficiency with a pretest used by other researchers [8, 22], our surveys included separate subcomponents that were highlighted by researchers interested in predicting success in CS from observing freshman CS students. Our choice of subcomponents were based on past researchers' success with demonstrating their importance in predicting student success in CS together with our ability to measure them with acceptable psychometric characteristics. We measured precollege experiences [1], self-rated skill with computer applications, a confidence with computing scale, and a course engagement scale [11]. The last two components were included to address some of the psychological and educational factors associated with course performance. With each of the subcomponents that we tested, we assessed its psychometric characteristics and once determined to be suitable, use the subcomponent scores to investigate demographic differences and also to investigate whether any of the component scores would be useful in predicting the students' final course grade in Algorithms and Data Structures courses.

The work we present here is consistent with earlier work [8, 15, 16], focusing on sophomore students enrolled in Algorithms and Data Structures courses with a wide range of skills and capabilities, in spite of passing the prerequisite courses (such as CS1, CS2). We used pre and post-surveys during the first and last two weeks of a semester to measure all of the subcomponent tests together with final course grades.

2 Methods

2.1 Participants

During the 2022–2023 academic year, instructors who were teaching Algorithms and Data Structures courses and using the BRIDGES infrastructure for homework assignments, were asked to recruit their students as volunteers for the online survey. Seven instructors volunteered for the Fall '22 and 10 for the Spring '23 semesters. Prior to their participation, each of the instructors and students signed an informed consent that described the purpose of the study and asked for permission to have final course grades sent to the project team at the end of the semester. Student grades and responses to the survey were identified through an anonymous code assigned by the instructor. Instructors were provided only summaries of their class responses and were not aware of any individual student's survey scores. The research study was conducted with the approval of UNC Charlotte Institutional Review Board. Table 1 lists each of the institutions involved in the data collection and the number of students who participated. Across the two semesters, data were collected online from 17 different Algorithms and Data Structures courses, taught by 14 different instructors.

2.2 Measures

The survey was divided into pre and post sections and as much as possible was developed from existing scales. The pre-survey included four components (precollege experience in computer science, self-rated skill with computer applications, confidence with computing, and prerequisite proficiency), while the post-survey repeated the confidence with computing scale and added a course engagement scale. The order of the components and the items within them remained consistent for all of the students across both Fall '22 and Spring '23 semesters.

2.2.1 Pre-Survey. Table 2 presents the items for three of the subcomponent measures : the precollege experience, skill with computer applications, and the confidence with computing scale.

Pre College Experience to CS. Alvarado et al. [1] studied grade level differences in CS students from a large university who had varied types of precollege experiences. Grade differences were found to persist throughout the CS curriculum for those with and without certain types of experiences.

Table 1. Number of Responses from Participating Institutions

Institution	Type (Size)	N	%
Wheaton College	4-year (<5,000)	26	2.0
Virginia Commonwealth U	University (>10,000)	330	24.8
U of Florida	University (>10,000)	507	38.1
Clovis College	Comm. Coll. (<1,000)	46	3.5
Eastern Nazarene College	4-year (<5,000)	8	0.6
U of Wisconsin-Parkside	4-year (<5,000))	12	0.9
Calvin U	University (<5,000)	83	6.2
UNC Charlotte	University (>10,000)	210	15.3
Hanover College	4-year(<5,000)	3	0.2
Valdosta State U	4-year (>10,000)	20	1.5
Walla Walla U	4-year (<5,000)	37	2.8
College of Staten Island	4-year (>10,000)	32	2.4
Birmingham Southern Coll.	4-year (<5,000)	22	1.7

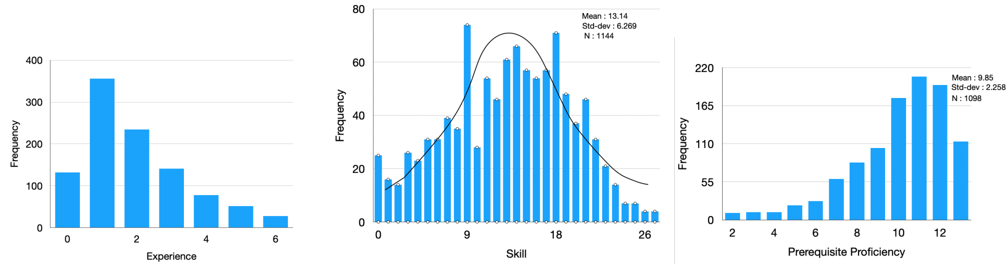
Table 2. Pre-Survey Used in the Study

Category	Choices
Pre-College Experiences (select all that apply)	Took CS in middle, high school Took 1 or more CS courses at previous college/univ Self taught 1 or more prog. lang. Participated in a CS club Learned html Participated in a summer CS exper. No experience
Skill Level (*Computer Applications*) (Rating: Basic, Intermediate, Advanced)	Word processing applications Spreadsheet applications Databases applications Presentation applications Multimedia applications Web design applications Web search engines Communication applications Computer games
Confidence with Computing Scale (5 pt Likert) (Strong Disagree-Strong Agree)	I enjoy using computers* I feel comfortable using computers* I think that computers are difficult to use I am willing to learn more about computers* I feel threatened when learning how to program I believe that it is important to learn how to program I know many advanced features in the software that I use I am comfortable learning new software or using new computers

The confidence with computing scale was repeated with the post-survey. The three items that were included in the final confidence score are identified with an asterisk. Prerequisite proficiency test is available on request from the authors.

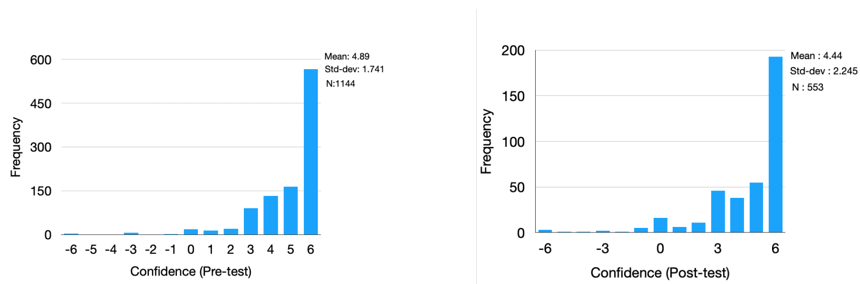
Based on their work, we listed their precollege experiences and ask our participants to select all that applied before the students' first CS course at their present college/university. This item was scored by summing the number of experiences selected by each student. Scores ranged from 0 (no experience) to 6.

Skill with Computer Applications. Wilson et al. [31] identified previous computer experiences as predictive of success in CS together with comfort level and math background. To measure



(a) Pre-College Experience (b) Skill with Computers (c) Prerequisite Proficency

Fig. 1. Distribution of pre-survey components.



(a) Confidence in Computing (Pre-Test) (b) Confidence in Computing (Post-Test)

Fig. 2. Distribution of confidence in computing.

computing skill, we asked students to use a 4-point scale (none (0), basic (1), intermediate (2), advanced (3)) while working with each of the 9 applications outlined in Table 2 . For each item, 0–3 points were given depending upon the level of the skill and adding each of the item scores together to create a total score, which ranged from 0 to 27.

Confidence with Computing. Eight items from a variety of confidence scales were adapted to assess students' confidence with computing. Students were asked to indicate on a 5-point Likert scale the extent to which they disagreed/agreed with each statement. Items were scored from −2 (strongly disagree) to +2 (strongly agree) to indicate the degree of agreement. Two of the items (item 3 and 5) were reverse scored. Item scores were summed to a total score.

Prerequisite Proficiency. We based our prerequisite proficiency test on the research of Cheng et al. [8] and Denzel et al. [10]. With some minor modifications, we used their 14-item multiple-choice proficiency test written in pseudocode to measure student's knowledge from introductory CS courses. Total score was calculated by summing the number of correct responses.

Figure 1 shows the distribution of the pre-survey components, while Figure 2 shows the distribution of confidence with computing for the pre- and post-tests.

2.2.2 Post-Survey. Confidence with Computing. The 8-item scale used in the pretest was repeated to measure changes in students' confidence with computing as a result of the semester long course.

Course Engagement. We used the Course Engagement Scale developed by Handelsman et al. [11] as our measure of student engagement in the Algorithms and Data Structures course. It is a 24-item scale with four subscales that measure study skills, emotional engagement, participation/interaction engagement, and performance engagement. The students responded on a 5-point scale that varied from 1 (Does not Describe Me) to 5 (Describes Me Extremely Well). It is a well-used measure with strong psychometric characteristics and it has recently been found by Shappie et al. [26] to predict the final course grades of a sample of African American students.

2.3 Procedure

Just prior to the beginning of each semester, instructors who signed an informed consent to participate in the BRIDGES program were sent a recruiting script to solicit volunteers from the students who were enrolled in their Algorithms and Data Structures course. Included was a Qualtrics link, which they were asked to share with their students, to provide informed consent to participate in the BRIDGES project and to have their grades sent to the project evaluator at the end of the semester. Student responses to all Qualtrics surveys were deidentified. Each student was provided a unique code by the instructor and that code was used to agree or disagree with the option of having their grades included in the study. A list of codes from the students who agreed in every class was returned to the instructor by the program evaluator and only those students' final grades were sent to the project team and included in the study.

Both the pre and the post-surveys were administered online on Qualtrics and was self-paced. Time to complete the pre-survey ranged from 10 to 15 minutes and 5 to 10 minutes for the post-survey. Following the receipt of the informed consent, a link to the pre-survey was sent to the instructor who circulated it to their students in a supervised setting sometime during the first week of the semester. Toward the end of the semester using a similar procedure, a link to the post-survey was sent so that students could respond in a supervised setting sometime during the last two weeks of the semester.

Students signed on to both of the surveys using an arbitrary code assigned by their instructor. They were assured that their individual responses would be deidentified and known only to the project evaluator. Instructors were sent a summary of the class responses to the surveys.

3 Results

Survey data were collected nationwide from Algorithms and Data Structures courses. Table 3 shows the gender and ethnic distribution in each of the data collection efforts conducted during the 2022–2023 academic year. Over a thousand students participated in the pre-survey and the post-survey data included 551 students. There were fewer participants in the post-survey because of withdrawals during the semester, class absences, as well as some student reluctance to participate in repeated surveys. Chi Square tests were used to compare the frequencies of the White, Black, Asian, and Hispanic students within each of the Men and Women categories. Comparisons of the Other ethnic group and the Non-Binary gender category were not tested because of the very low or absent frequency counts. None of the comparisons within either the Men or Women student groups showed any significant pre/post differences in percent of representation. Chi Square values varied from 0.000 to 1.175 with resulting p values between 1.0 and 0.356. Figure 3 plots the percent of pre vs post representation and shows the overlap in the ethnic distribution within each of the gender categories.

3.1 Psychometric Characteristics of Measures

3.1.1 Pre-Survey. Means (SD) and intercorrelations for the four components in the pre-survey are presented in Table 4. Although the component scores are significantly correlated with each other, the values were low, ranging from .18 to .35, indicating that only 3–12 percent of the variance between any 2 of the measures were accounted for by each of the paired relationships. Since none

Table 3. Distribution of Survey Data by Gender and Ethnic Group

	Men		Women		Non Binary		Total	
Pre-Survey								
	N	%	N	%	N	%	N	%
White	332	43.7	103	33.8	9	40.9	444	40.9
Black	75	9.9	38	12.5	1	4.5	114	10.5
Asian	223	29.4	109	35.7	8	36.4	340	31.3
Hispanic	97	12.8	43	14.1	0	0.0	140	12.9
Other	32	4.2	12	3.9	4	18.2	48	4.4
Total	759	100	305	100	22	100	1086	100
Post-Survey								
White	172	45.7	61	35.9	2	40.0	235	42.6
Black	55	14.6	31	18.2	1	20.0	87	15.8
Asian	108	28.7	53	31.2	1	20.0	162	29.4
Hispanic	27	7.2	17	10.0	1	20.0	45	8.2
Other	14	3.7	8	4.7	0	0.0	22	4.0
Total	376	100	170	100	5	100	551	100

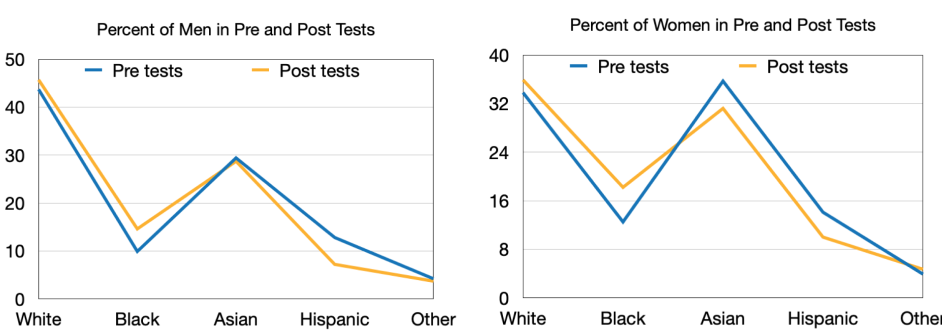


Fig. 3. Percent of men and women participating in pre/post tests.

Table 4. Means, SD, and Intercorrelations for Pre-Survey

Subscore	Mean	SD	N	1	2	3
1. Skill	13.14	6.27	1144			
2. Confidence	4.89	1.74	1144	0.29^{**}		
3. Experience	1.97	1.49	1144	0.35^{**}	0.18^{**}	
4. Proficiency	9.85	2.26	1098	0.21^{**}	0.21^{**}	0.23^{**}

$^{**}p < 0.001$ (two-tailed).

of the measures were strongly associated with each other, it suggests that the four component scores were measuring different constructs each of which represented by a unique concept.

Precollege CS Experience. On average, the students reported two experiences with computer science before enrolling in their college or university. Figure 1(a) shows the skewed distribution of the total number of precollege experiences, where the mode is one and 7% of the respondents reported none.

The most frequent experience was taking a CS course at a college/university (31%) or high school/middle school (22%). Only 12% reported learning HTML and 12% had self-taught one or more programming languages. Participating in a CS club yielded a 10% response and participating in a summer CS camp was the least reported experience (6%).

Skill with Computer Applications. Figure 1(b) shows that when the students rated their skill level with a variety of computer applications, the total score across the nine applications was normally distributed with a majority of students indicating a basic to intermediate level of skill. With our data, the scale was found to have a high degree of internal consistency (Cronbach's $\alpha = 0.90$) (McDonald's $\omega = .90$). A factor analysis of the nine items showed one major factor accounting for 56% of the variance and a second factor with an additional 11%. All nine items loaded to the major factor (loadings ranged from 0.58–0.82) and only one item (database applications) had a higher loading to the second factor (0.65).

Confidence with Computing. With our data, the 8-item scale was found to have an acceptable internal reliability (Cronbach's $\alpha = 0.73$); but item-total statistics together with a factor analysis of the scale showed that when several of the items were dropped and only three of the items retained (Items 1, 2, 4), the scale rose to a stronger level of internal consistency (Cronbach's $\alpha = 0.82$) (McDonald's $\omega = .83$). For this reason, we retained for our analysis the 3-item confidence with computing scale with total scores that ranged from −6 to +6.

The distribution of the three-item confidence scores is presented in Figure 2(a). The data are skewed with a majority of the responses at ceiling, indicating strong confidence in computing. A factor analysis of the 3-item scale showed one factor accounting for 74% of the variance with all three items loading to that factor. Factor loading ranged from 0.82 to 0.90.

Prerequisite Proficiency. We did an item analysis and found that all of the items correlated significantly with the total-score with varied item difficulty (ranging from 0.33 to 0.97). Reliability was measured with a Kudor Richarson-20 score of 0.68. Item-total statistics showed, however, that when item 1 was dropped, the reliability of the scale rose to an acceptable level of .70. Item 1 was a difficult item (.33) that showed the lowest item-total correlation (0.28), and a high item variance. To maximize the reliability of our scale, we dropped the first item and retained the remaining 13 items for our data analysis. Total scores on the scale ranged from 0 to 13. The distribution of the proficiency scores is in Figure 1(c). Scores are skewed toward the high end of the distribution with 9% of the scores below 7.

3.1.2 Post-Survey.

Confidence with Computing. When the confidence in computing scale was again taken during the last two weeks of the semester, reliability of the scale was found to be stronger in the post-survey when compared to the pre; but a similar pattern was found, such that measures of internal consistency increased when the 8-item scale (Cronbach's $\alpha = 0.81$, McDonald's $\omega = 0.81$) was reduced to a 3-item scale (Cronbach's $\alpha = 0.91$, McDonald's $\omega = 0.91$). To maximize the reliability of our measurements, we retained for our analysis the 3-item confidence in computing scale with scores that varied from −6 to 6. As shown in Figure 2, the same skewed distribution of scores was found and the pre and post confidence scores were strongly correlated, $r = 0.84$, $p < .001$. A comparison of the confidence scores in the pre- and post-survey found a small but significant drop in computing confidence, $t = 2.761$, $p = .006$, Cohen's $d = 2.135$. Mean scores for the pre and post confidence in computing scale are as follows: 4.82 ($SD = 1.88$) and 4.54 ($SD = 2.11$). The drop in confidence could have happened because the students at the end of the semester had a more realistic view of what

Table 5. Means, Correlations for Course Engagement Subscales

Subscore	Mean	SD	1	2	3
1. Study Skills	30.74	6.985			
2. Emotional	16.41	4.197	0.66**		
3. Participation	17.36	5.148	0.66**	0.64**	
4. Performance	10.34	2.704	0.43**	0.53**	0.49**

**p < 0.001 (two-tailed).

the course was all about at the end of the semester. While in the beginning, their expectations were probably influenced by their experiences from prerequisite courses.

Course Engagement. With our data, the reliability of the four subscales using Cronbach's α ranged from 0.79 to 0.86. Sub scores were created by summing the item scores within each of the subscales. The 9-item study skills factor has scores ranging from 9 to 45. Emotional engagement has 5 items with scores ranging from 5 to 25, while participation includes 6 items and a range of scores from 6 to 30. Performance engagement has only 3 items with scores ranging from 3 to 15. Table 5 presents the means and strong intercorrelations among the 4 subscales of the course engagement scale.

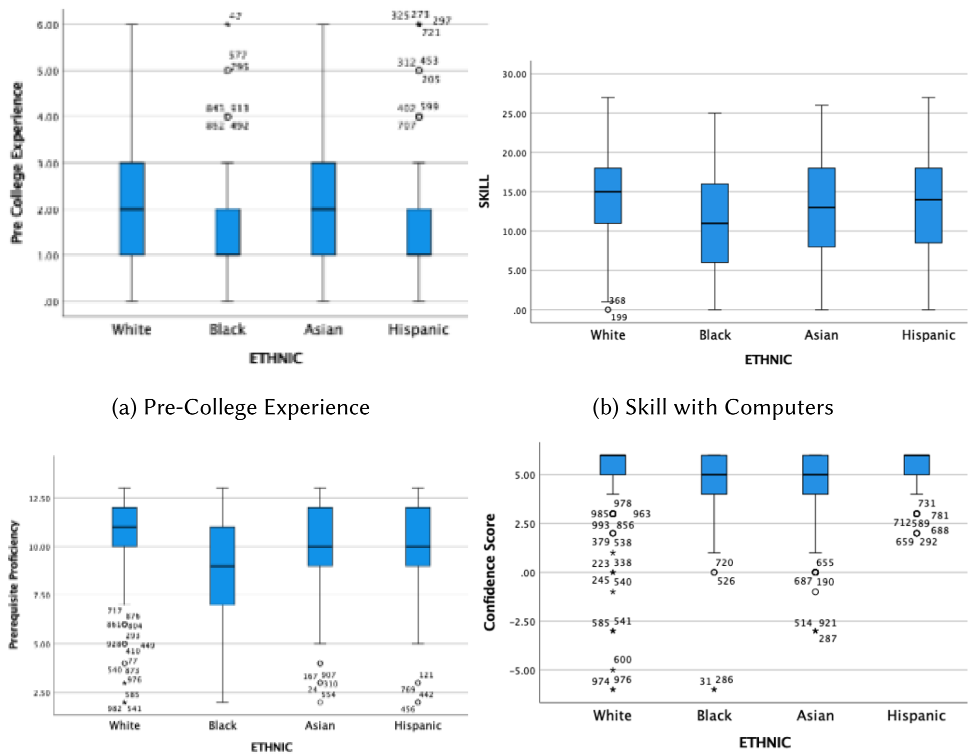
3.2 Analysis of Demographic Differences

A general linear model with type III sum of squares was used to test for gender and ethnic differences in all of the component measures collected in the pre- and post-surveys. The small number of responses in the "Nonbinary" category were not included in the gender analysis; and similarly, the small number of responses using the "Other" designation were not included in the analysis of ethnic effects. A multivariate analysis was used for the four components measures in the pre-survey and for the four sub scores of the engagement scale. Follow-up univariate analyses were conducted on the significant main effects with additional *post hoc* comparisons using Bonferroni corrected p values, to locate differences following main effects of ethnic group. All statistical tests were conducted with SPSS version 28 and a significance level of 0.05 was used for the analyses.

3.2.1 Pre-Survey Components. The analysis of the pre-survey components resulted in a significant multivariate effect of gender, (Wilks' Lambda = 0.968, $p < 0.001$, ${\eta_p}^2 = .032$) and ethnic group, (Wilks' Lambda = 0.919, $p < 0.001$, ${\eta_p}^2 = .028$) but no interaction effect ($p = 0.115$).

Follow-up univariate tests showed that Men (5.03) had higher confidence when compared to Women (4.65) ($F = 8.36$, $p = 0.004$, ${\eta_p}^2 = .024$) and better scores on the proficiency test ($F = 6.39$, $p = 0.012$, ${\eta_p}^2 = .024$). Means for the Men and Women students are 10.20 and 9.95, respectively. There were no gender differences in precollege CS experience ($p = 0.057$) or in self-rated skill with computer applications ($p = 0.819$).

Univariate tests of the main effect of ethnic group showed significant effects for all four component scores; Pre college experience, $F = 4.94$, $p < 0.002$, ${\eta_p}^2 = .014$; Skill, $F = 5.63$, $p < 0.001$, ${\eta_p}^2 = .016$; Confidence, $F = 9.15$, $p < 0.001$, ${\eta_p}^2 = .044$; and Proficiency Score, $F = 16.68$, $p < 0.001$, ${\eta_p}^2 = .047$. Figure 4 shows the mean differences across the four ethnic groups for each of the measures in the pre-survey. *Post hoc* comparisons ($p < 0.05$) showed that Black participants were found to differ from White participants on all four measures and from Asian participants on both precollege experiences and proficiency scores. Hispanic participants were also found to have fewer precollege experiences than Asian participants but their confidence scores were slightly higher. Asian



(a) Pre-College Experience

(b) Skill with Computers

(c) Prerequisite Proficiency

(d) Confidence Level

Fig. 4. Gender and ethnic differences.

participants had lower confidence in computing scores in comparison to White participants but otherwise, showed a similar profile as the White participants.

Given the pattern of gender and ethnic differences in prerequisite proficiency scores, we did an additional regression analysis to determine whether covarying differences in precollege experiences and confidence in computing would eliminate the gender and ethnic differences that were shown in the above analysis. The results of that analysis are in Table 6. When the covariates were used and the effects of precollege experiences and confidence in computing were accounted for, the gender differences in proficiency were no longer significant; but not ethnic group differences. The adjusted means for the White, Black, Asian, and Hispanic groups are, respectively, 10.66, 9.03, 10.39 and 9.93.

3.2.2 Post-Survey Components. A gender effect was also significant in the multivariate analysis of the sub scores for the Course Engagement Scale (gender effect, Wilks Lambda = 0.957, $p < 0.001$, ${\eta_p}^2 = .043$; but there was no ethnic effect, Wilks Lambda = 0.973, $p = 0.306$, ${\eta_p}^2 = .009$) or a gender by ethnic interaction ($p = 0.817$).

Follow-up univariate tests showed that Women had higher study skills (M = 32, SD = 6.90) than Men (M = 30, SD = 6.87), ($F = 12.10$, $p < 0.001$, ${\eta_p}^2 = 0.023$) but did not differ significantly from Men on any of the other three engagement sub scores ($p = 0.689$ for emotional; $p = 0.207$ for participation, $p = 0.318$ for performance). An analysis on the confidence in computing scale, repeated in the post-survey, did not show any effect of gender ($F = 3.59$, $p = 0.06$) or ethnic group ($F = 1.99$, $p = 0.115$).

Table 6. Regression Analysis on Prerequisite Proficiency Scores with Precollege Experience and Confidence in Computing Used as Covariates

Source	SS	df	MS	F	Sig	η_p^2	Power
Experience	130.224	1	130.224	25.453	<0.001	0.01	0.978
Confidence	159.439	1	159.439	31.163	<0.001	0.034	1.00
Gender	16.319	1	16.31	3.19	0.074	0.002	0.278
Ethnic	203.014	3	67.67	13.226	<0.001	0.037	1.0
Gend. × Ethn.	12.391	3	4.13	0.807	0.406	0.003	0.266
Error	5157.285	1008	5.116				
Total	115581.00	1018					
Total Correct.	5935.396	1017					

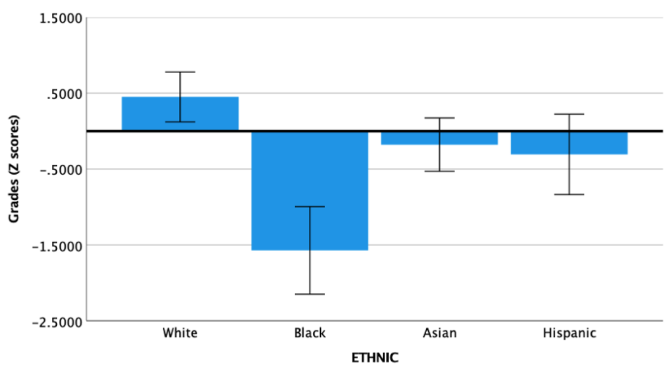


Fig. 5. Distribution of grades.

3.2.3 Grades. At the end of the semester, final course grades were sent by the instructors for the students who provided informed consent. Instructors varied in the format that they used to send the grades, so all letter grades were transformed into a numerical format using a scale.[1] The numerical grades for each of the classes were then transformed into Z scores by putting each grade in relationship to the mean and standard deviation for the class.[2] The analyses were then performed on both the numerical grades and the Z scores. Since both analyses resulted in the same effects, we are presenting the statistical results from the Z score analyses. The skewed distribution of those grades is presented in Figure 5. Z scores were used in the analysis rather than numerical grades because there were varying class sizes, instructors, and types of institutions. Anyone of those factors could have different influences on how grades were assigned.

The analysis on the Z scores did not show gender differences ($F < 1$); but there was a significant effect of ethnic group, $F = 12.17$, $p < 0.001$, ${\eta_p}^2 = 0.05$. *Post hoc* tests showed that Black participants' grades were significantly lower than the other groups. Means (SDs) on a 100-point scale for the White, Black, Asian, and Hispanic groups are respectively; 87 (11.40), 76 (19.7), 85 (11.8), and 83.7 (13.5).

[1]Grade conversion scale: $A+ = 100, A = 96, A- = 92, B+ = 89, B = 86, B- = 82, C+ = 79, C = 76, C- = 72, D+ = 69, D = 66, D- = 62, F = 50$.
[2]For each class section, a mean M and SD was computed and the Z score was calculated as $Score_z = (Score_{numeric} - M)/\text{SD}$.

Table 7. Regression Analysis with Z Scores as Final Grades

Model	Sum of Sq.	df	Mean Sq.	F	Sig.
Regression	72.85	9	8.094	10.222	<0.001[a]
Residual	209.048	264	0.792		
Total	281.897	273			

Dependent variable: Z score.
[a]Predictors: (Constant), Post confidence, Proficiency, Participation, Experience, Skill, Pre confidence, Performance, Skills, Emotional.
R = .508, R^2 = .258.

	B	Std Error	Beta	t	Sig
Experience	.020	.043	.027	.456	.649
Skill	−.009	.010	−.057	−.940	.348
Confidence1	−.015	.039	−.024	−.385	.701
Proficiency	.113	.025	.258	4.548	<.001
Skills	.006	.011	.042	.558	.577
Emotional	.009	.020	.036	.449	.654
Participation	−.002	.014	−.011	−.151	.880
Performance	.137	.026	.367	5.278	<.001
Confidence2	.012	.032	.023	.368	.713

Dependent Variable: Z score.

Confidence 1: pre-survey measure.

Confidence 2: post-survey measure.

In order to determine whether any of the subcomponents of the pre and post surveys could be used to predict final grades in Algorithms and Data Structures, a regression analysis was conducted on the Z scores. Results of that analysis are summarized in Table 7. The two measures that predicted final grades were the prerequisite proficiency score (B = .113) and performance scale from the Course Engagement Scale (B = .137). In general, all of the component scores were significantly correlated with the final grades except for self-rated skill with computer applications and the confidence in computing score in the pre-survey. The fact that the confidence in computing score in the pre-survey was not related to grades while the confidence scores in the post-survey was, reinforces the argument that the pretest confidence was based on something other than experiences in the course itself (most likely a carryover from prerequisite courses), while in the post-test the confidence score most likely reflected the course experience itself since it was found to correlate with course grade.

4 Discussion

4.1 Gender and Ethnic Effects

Adopting a comprehensive approach to studying the preparation gap among a diverse sample of students enrolled in Algorithms and Data Structures courses during the 2022–2023 academic year was successful in identifying ethnic and gender differences. Consistent with the past literature [16], there were differences in prerequisite proficiency. On average, Women scored lower than Men, and Black participants had lower scores than White, Asian, and Hispanic participants. However, the group differences were not limited to proficiency scores, but rather a characteristic of all the components of the pre-survey administered during the beginning of the semester. Women had fewer precollege experiences with computing than Men; and among the ethnic groups, White and Asian participants had more experiences than Hispanic participants, but Black participants had the fewest number of computing experiences. Confidence in computing was also found to vary among the groups. Women showed lower confidence than Men, and among the ethnic groups Black and Asian participants showed lower confidence than Hispanic and White participants. Additionally, there were differences among the groups when students were asked to self-rate their skill with a variety of computer applications. There was an ethnic difference with Black participants showing lower scores in comparison to the other groups.

Interestingly, however, when precollege experiences and confidence with computing scores were used as covariates in a regression analysis on the prerequisite proficiency scores the group differences were reduced as evidenced by the lowered effect sizes in comparison to the analysis without the covariates. As shown in Table 6, the gender difference was no longer significant, but there was a small lingering effect of ethnic group. These data suggest that efforts to address the group differences in a CS preparation gap may be useful if they were directed at multiple components involved in the pre-survey rather than just a sole focus on prerequisite proficiency [8, 16]. For example, providing enhanced precollege and/or early college computing experiences to Women and Non-Asian minorities may reduce some of the demographic differences in the preparation gap found with computer science majors. A focus on prerequisite proficiency in the absence of the other components measured in this research may not provide the most effective way of reducing or eliminating demographic group differences.

It is also important to note that the measures taken at the end of the semester in the post-survey did not show as many demographic differences as the pre-survey. Only a gender difference in one of the four components of the Course Engagement Scale was significant, Women had higher study skills than Men. No other ethnic or gender difference was significant in the component scores of the Course Engagement Survey. In addition, the confidence in computing scale when repeated during the last two weeks of the semester did not show any ethnic or gender differences. What this suggests is that underrepresented group members who stick with the CS course in Algorithms and Data Structures until the last two weeks of the semester have comparable confidence and are just as engaged in their coursework with comparable study skills as majority group members.

The students who participated in the surveys were allowed to self identify their social groups. Those who chose the nontraditional categories of Binary (for gender) and Other (for ethnic group) could not be summarized together with the other groups because they were too few in number and their data were much more variable in comparison to the more traditional groups. The variability was most likely due to the fact that in the Other category, the data were split between those who listed multiple categories (such as White/Black, or Asian/Black); while others choose to list more specific but often overlooked ethnic groups such as Middle Eastern or Native American.

Studying multiple categories of social group membership as suggested by researchers in intersectionality was helpful in identifying different profiles for several of the underrepresented minorities in CS. Participation rates as shown in Table 3 and Figure 3 in this sample were similar to the CS enrollment trends reported by Lunn et al. [18]. There is no doubt that White and Asian men have much higher participation rates in this survey and in CS in general than Black and Hispanic men. The Women showed similar trends except that Asian Women had the highest participation rate followed closely by White Women. It is also noteworthy that among the White participants, Women represent only 23 percent of the sample of all White participants; however the percentage of Women within the participant populations of under-represented minorities is larger (34 percent for Black, 33 percent for Asian, and 30 percent for Hispanic participants).

4.2 Course Grades

An analysis of the final course grades when corrected through the use of Z scores for differences in college and instructor standards showed an ethnic difference but no effect of gender. Although all scores were skewed toward the high end, average scores for the Black participants were lower than the others.

All of the component scores of the pre- and post-surveys were significantly correlated with the final grades except for two of the pre-survey components: self-rated skill with computer applications and confidence in computing. When all of the component scores were entered into a regression, they explained 26% of the variance in final course grades. In addition, the prerequisite proficiency score and the performance subscale score on the Course Engagement Survey were found to be significant predictors of course grades.

4.3 Psychometric Characteristics of Measures

The data collected in this research provide some additional information about the reliability and validity of the pre- and post-survey component measures. Although most of the items in the measures were based on previous research on preparation gaps in computer science, we were able to build on those efforts to develop them into survey instruments to test their usefulness in predicting grades and identifying gender and ethnic differences.

Dropping the first item of the prerequisite proficiency test developed by Cheng and his colleagues at Boston College [8, 10] resulted in a 13-item test using pseudocode that was found to have acceptable internal consistency and could be used as a predictor of final course grade. In addition, content validity of the prerequisite proficiency score was demonstrated by its correlation with both grades and the performance subscale of the Course Engagement Survey. The prerequisite proficiency test shows promise as an effective measure of prerequisite proficiency and would benefit from continued research to enhance its reliability and validity. Adding more items written in pseudocode and additional data collection efforts testing its predictive use are warranted, however.

When the confidence in computing scale was reduced to a 3-item scale it showed strong reliability in both the pre- and post-survey. It identified some important demographic differences when used in the beginning of the semester and its use as a covariate was successful in reducing the gender difference found with prerequisite proficiency scores. When used for a second time in the post-survey, it was significantly correlated to the final course grade; but the confidence score was lower. As indicated in the results section, this is most likely due to the fact that a confidence in computing score taken during the first two weeks of the semester was likely a result of the students' comfort with the CS prerequisite courses; while the post-test confidence score was a more realistic assessment of comfort/confidence in computing in the course itself.

Similarly, developing the precollege CS experience item introduced by Alvarado et al. [1] into a component score in the pre-survey proved to be a very effective way of identifying demographic differences among CS students and its use as a covariate helped to reduce the demographic groups effect on proficiency scores.

The only component of the pre-survey that did not add any unique information to our understanding of the preparation gap was the self-rated skill with computer applications. The measure had a high degree of reliability, but it was not found to correlate significantly with the final course grade.

The four sub scores of the Course Engagement Scale developed by Handelsman et al. [11] were useful in showing that the demographic differences that were evident in the beginning of the semester, could not be attributed to group differences in course engagement. Other than the higher score on study skills shown by Women, there were no other gender or ethic differences. Also, interesting is the fact that the performance sub score of the Course Engagement Scale was consistent with previous literature [26] as a significant predictor of final course grade.

4.4 Related Work

As stated earlier, the goal of this work is to build on previous efforts, confirm some of the earlier results and take a closer look at gender and ethnic differences by conducting studies in Algorithms and Data Structures courses. Similar studies in earlier work have focused largely in CS1 [3, 24] and a fewer in CS2 [13]. These studies looked at some of the factors we also investigated here and were able to confirm those results. For instance, Wilson et al. [31] found undergrad students' comfort level predicted midterm grades. Our results are consistent with that and extended that finding to final course grades. In addition, we showed that confidence (comfort level) with computing when used as a covariate decreased both the gender and ethnic differences in performance on prerequisite proficiency.

Similar to Zingaro et al. [32] we also used a multi-institutional sample in our studies, but also added final grades and additional features. We addressed several of the weaknesses called out by Hellas et al. [12], such as the use of online surveys, securing informed consent of both instructors and students, provide data on reliability and validity of the measures used in the studies. We also replicated our results across two semesters, contributing to the robustness of our data collection efforts.

Consistent with the past literature [4, 31] we reinforced the importance of precollege experiences and confidence (comfort) in computing as relevant factors in predicting prerequisite proficiency and final course grades. However, our results did not find that self-rated skill with nine computer applications was an important factor, even though it assessed skill with the applications and was found to have high reliability and strong factor loading to a single factor.

5 Limitations

Although we were able to draw a diverse sample of CS students from a wide variety of institutions located in varied areas of the US and had sufficient samples to study gender effects (in Men and Women) and ethnic differences (among White, AA/Black, Asian, and Hispanic students in the US), we did not have sufficient numbers to explore all of the ethnic by gender subgroups. In particular, Black and Hispanic women were underrepresented in our sample. So, our findings are limited to drawing conclusions about the main effects of gender and ethnic groups but not their interaction. The main effects identified the demographic differences that were prevalent but did not provide any additional information as to the causal agents of those differences.

Also, the fact that the sample was smaller for the post-test relative to the pre-test added attrition as a limiting factor to our findings, particularly as it relates to the regression with grades. The reduced sample in the post-test was a result of a number of factors that were difficult to quantify because of the number of institutions involved in the testing (i.e., students who dropped the class, absences during post-survey testing, and individual student decisions not to participate in the post test). However, it is important to point out that when we compared the ethnic/gender distribution of the Men and Women samples there were no significant pre/post test differences in the social groups.

Also, although we collected samples from 13 different US institutions of varying sizes and types, a majority of the sample came from the three largest institutions. However, it should also be noted that this is where we obtained the majority of underrepresented student samples.

6 Conclusions

In this article, we presented a comprehensive study to get a better understanding of gender and ethnic differences on a national sample of students from 13 institutions across two semesters. We paid careful attention to ensure the reliability and validity of the study metrics as well as ethical considerations in performing the study. Based on the past literature and the results of this study, we conclude that requiring students to take prerequisite courses such as CS1 and CS2 may not be sufficient to remedy the demographic differences in computing experiences that students have prior to entering the major. Differences in prerequisite proficiency may be influenced by a student's precollege experiences with computing. However, our results may also suggest that simply providing access to prior experiences is not the only concern when aiming to reduce ethnic and gender differences in preparation gaps; we find that the impact of prior preparatory learning experiences on students' confidence when used as a covariate decreased both the gender and ethnic differences in performance. These prior preparatory learning experiences may affect their computing confidence and subsequently, their desire to continue in the major. Securing a diverse workforce in computing may require the availability of computing experiences that build knowledge and confidence earlier rather than later in the educational process.

Acknowledgment

We would like to thank the reviewers for their thoughtful comments that helped improve the manuscript.

References

[1] Christine Alvarado, Gustavo Umbelino, and Mia Minnes. 2018. The persistent effect of pre-college computing experience on college CS course grades. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. ACM, New York, NY, 876–881.

[2] Allie Beckman, Matthew Mcquaigue, Alec Goncharow, David Burlinson, Kalpathi Subramanian, Erik Saule, and Jamie Payton. 2020. Engaging early programming students with modern assignments using BRIDGES. In *Proceedings of the CCSC CP. The Consortium for Computing Sciences in Colleges (CCSC)*, 74–83.

[3] Susan Bergin and Ronan Reilly. 2005. Programming: Factors that influence success. *SIGCSE Bulletin* 37, 1 (2005), 411–415. DOI: https://doi.org/10.1145/1047124.1047480

[4] Susan Bergin and Ronan Reilly. 2005. Programming: Factors that influence success. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '05)*. ACM, New York, NY, 411–415. DOI: https://doi.org/10.1145/1047344.1047480

[5] David Burlinson, Matthew Mcquaigue, Alec Goncharow, Kalpathi Subramanian, Erik Saule, Jamie Payton, and Paula Goolkasian. 2023. BRIDGES: Real world data, assignments and visualizations to engage and motivate CS majors. *Education and Information Technologies* 29, 15 (2023), 1–27.

[6] David Burlinson, Mihai Mehedint, Chris Grafer, Kalpathi Subramanian, Jamie Payton, Paula Goolkasian, Michael Youngblood, and Robert Kosara. 2016. BRIDGES: A system to enable creation of engaging data structures assignments with real-world data and visualizations. In *Proceedings of ACM SIGCSE '16*. ACM, New York, NY, 18–23.

[7] Eugenia Cheng. 2020. *X + Y: A Mathematician's Manifesto for Rethinking Gender*. Hachette.

[8] Qingwan Cheng, Angela Tao, Huangliang Chen, and Maira Marques Samary. 2022. Design an assessment for an introductory computer science course: A systematic literature review. In *Proceedings of the 2022 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–8.

[9] Elizabeth R. Cole. 2009. Intersectionality and research in psychology. *The American Psychologist* 64, 3 (2009), 170.

[10] Brennan Denzel and Maíra Marques Samary. 2023. A general concept inventory for introductory computer Science—A work in progress. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE '23)*. ACM, New York, NY, 1228.

[11] Mitchell M. Handelsman, William L. Briggs, Nora Sullivan, and Annette Towler. 2005. A measure of college student course engagement. *The Journal of Educational Research* 98, 3 (2005), 184–192.

[12] Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Predicting academic performance: A systematic literature review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '18 Companion)*. ACM, New York, NY, 175–199.

[13] Sara Hooshangi, Margaret Ellis, and Stephen H. Edwards. 2022. Factors influencing student performance and persistence in CS2. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education (SIGCSE '22)*, Vol. 1. ACM, New York, NY, 286–292.

[14] Danyelle T. Ireland, Kimberley Edelin Freeman, Cynthia E. Winston-Proctor, Kendra D. DeLaine, Stacey McDonald Lowe, and Kamilah M. Woodson. 2018. (Un)hidden figures: A synthesis of research examining the intersectional experiences of black women and girls in STEM education. *Review of Research in Education* 42, 1 (2018), 226–254. DOI: https://doi.org/10.3102/0091732X18759072

[15] Sophia Krause-Levy, Sander Valstar, Leo Porter, and William G. Griswold. 2020. Exploring the link between prerequisites and performance in advanced data structures. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. ACM, New York, NY, 386–392.

[16] Sophia Krause-Levy, Sander Valstar, Leo Porter, and William G. Griswold. 2022. A demographic analysis on prerequisite preparation in an advanced data structures course. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education (SIGCSE '22)*, Vol. 1. ACM, New York, NY, 661–667.

[17] Alex Lishinski, Sarah Narvaiz, and Joshua M. Rosenberg. 2022. Self-efficacy, interest, and belongingness—URM students' momentary experiences in CS1. In *Proceedings of the 2022 ACM Conference on International Computing Education Research (ICER '22)*, Vol. 1. ACM, New York, NY, 44–60.

[18] Stephanie Lunn, Leila Zahedi, Monique Ross, and Matthew Ohland. 2021. Exploration of intersectionality and computer science demographics: Understanding the historical context of shifts in participation. *ACM Transactions on Computer Education* 21, 2 (2021), 30. DOI: https://doi.org/10.1145/3445985

[19] Matthew McQuaigue, Mack Larson, Philip Smith, Sydney Melech, Kalpathi Subramanian, and Erik Saule. 2024. Engaging CS1 students with audio themed assignments. *Journal of Computing Sciences in Colleges* 39, 8 (2024), 158–172.

[20] Matthew Mcquaigue, Jay Strahler, Kalpathi Subramanian, Erik Saule, and Jamie Payton. 2022. Location based assignments in early CS courses using BRIDGES engages students. *Journal of Computing Sciences in Colleges* 38, 5 (2022), 107–116.

[21] Carolina Milesi, Lara Perez-Felkner, Kevin Brown, and Barbara Schneider. 2017. Engagement, persistence, and gender in computer science: Results of a smartphone ESM study. *Frontiers in Psychology* 8 (2017), 602.

[22] Miranda C. Parker, Mark Guzdial, and Shelly Engleman. 2016. Replication, validation, and use of a language independent CS1 knowledge assessment. In *Proceedings of the 2016 ACM Conference on International Computing Education Research (ICER '16)*. ACM, New York, NY, 93–101.

[23] Leo Porter, Daniel Zingaro, Soohyun Nam Liao, Cynthia Taylor, Kevin C. Webb, Cynthia Lee, and Michael Clancy. 2019. BDSI: A validated concept inventory for basic data structures. In *Proceedings of the 2019 ACM Conference on International Computing Education Research (ICER '19)*. ACM, New York, NY, 111–119.

[24] Keith Quille and Susan Bergin. 2018. Programming: Predicting student success early in CS1. A re-validation and replication study. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '18)*. ACM, New York, NY, 15–20.

[25] Yolanda A. Rankin and Jakita O. Thomas. 2020. The intersectional experiences of black women in computing. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. ACM, New York, NY, 199–205.

[26] Alexander T. Shappie and Scott M. Debb. 2019. African American student achievement and the historically black university: The role of student engagement. *Current Psychology* 38, 6 (2019), 1649–1661.

[27] Julie M. Smith. 2021. Beyond the gender binary in computing education research. In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER '21)*. ACM, New York, NY, 444–445. DOI: https://doi.org/10.1145/3446871.3469794

[28] Jason Strahler, Matthew Mcquaigue, Alec Goncharow, David Burlinson, Kalpathi Subramanian, Erik Saule, and Jamie Payton. 2020. Real-world assignments at scale to reinforce the importance of algorithms and complexity. In *Proceedings of CCSC North East*. Journal of Computer Sciences in Colleges, Evansville, IN, 166–175.

[29] Claudia Szabo, Miranda C. Parker, Michelle Friend, Johan Jeuring, Tobias Kohn, Lauri Malmi, and Judithe Sheard. 2025. Models of mastery learning for computing education. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSETS 2025)*. ACM, New York, NY, 1092–1098.

[30] Arne Weigold and Ingrid K. Weigold. 2021. Measuring confidence engaging in computer activities at different skill levels: Development and validation of the brief inventory of technology self-efficacy (BITS). *Computers & Education* 169 (2021), 104210.

[31] Brenda Cantwell Wilson and Sharon Shrock. 2001. Contributing to success in an introductory computer science course: A study of twelve factors. In *Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education (SIGCSE '01)*. ACM, New York, NY, USA, 5.

[32] Daniel Zingaro, Michelle Craig, Leo Porter, Brett A. Becker, Yingjun Cao, Phill Conrad, Diana Cukierman, Arto Hellas, Dastyni Loksa, and Neena Thota. 2018. Achievement goals in CS1: Replication and extension. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. ACM, New York, NY, 687–692.

Received 18 June 2025; revised 25 December 2025; accepted 5 March 2026