Computers in the Schools

Interdisciplinary Journal of Practice, Theory, and Applied Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/wcis20

An Exploration of Factors Impacting Middle School Students' Attitudes Toward Computer Programming

Mete Akcaoglu, Joshua M. Rosenberg, Charles B. Hodges & Jonathan C. Hilpert

COMPUTERS IN THE SCHOOLS
https://doi.org/10.1080/07380569.2021.1882209

An Exploration of Factors Impacting Middle School Students' Attitudes Toward Computer Programming

Mete Akcaoglu[a], Joshua M. Rosenberg[b], Charles B. Hodges[a], and Jonathan C. Hilpert[c]

[a]Instructional Technology, Department of Leadership, Technology, and Human Development, Georgia Southern University, Statesboro, Georgia, USA; [b]STEM Education, Department of Theory and Practice in Teacher Education, The University of Tennessee, Knoxville, Knoxville, Tennessee, USA; [c]Department of Educational Psychology and Higher Education, University of Nevada Las Vegas, Las Vegas, Nevada, USA

ABSTRACT
Computer programming is becoming an essential skill for young students regardless of their education or career goals. Therefore, for students to develop and for educators and researchers to accurately measure self-efficacy in and value for programming is important. Although student motivation in subject matter can be measured using self-report surveys, these types of instruments are prone to error due to inherent biases. In this quasi-experimental and cross-sectional study, we examined whether using a computer programming pretest before taking a perception survey (i.e., providing examples of the key concept in the survey beforehand) had an impact on students' self-reported self-efficacy and utility value, thus helping collect more accurate data. Results showed no significant difference on self-efficacy and value scores between those who received a pretest and those who did not. In further analysis, however, it was found that pretest performance was positively correlated with self-efficacy and value. In addition, boys reported significantly higher self-efficacy and value than girls, confirming gender disparity outlined in previous research. An exploratory, but important, finding was that there was an interaction between gender and test performance for the self-efficacy and utility value. While boys' who had high pretest scores also reported higher levels of self-efficacy and utility value, it was the opposite for girls.

KEYWORDS
Computer programming; gender; interest; motivation; self-efficacy; utility value

Understanding the language of computers and being able to apply computational skills to problems and topics across the curriculum are essential skills for today's younger generation to be successful in their future careers (Ydav et al., 2016). Although there are widespread efforts (e.g., CS4All) to support young students to develop competence in computer science (CS),

CONTACT Mete Akcaoglu makcaoglu@georgiasouthern.edu Instructional Technology, Department of Leadership, Technology, and Human Development, Georgia Southern University, 275 College of Education Drive, Room 3113, Statesboro, GA 304585 630, USA

students' career choices are likely impacted by several factors above and beyond knowledge and competence, such as motivational, contextual, and social factors (Tran, 2018). For example, the role of well-developed utility value is well known for its effect on continued engagement with school subjects and is a useful framework for understanding how students develop optimal motivation for subject matter (Eccles & Wigfield, 2002; Hecht et al., 2019). It also is known that self-efficacy, or an individual's confidence in their ability to accomplish a task (Bandura, 1977), is a predictor of effort and persistence in a task (Eccles & Wigfield, 2002). To develop the next generation of students equipped with knowledge and motivation to tackle computing tasks, it is important to accurately measure and understand students' motivation toward computer science and attend to problems that negatively impact it.

A potential difficulty for attempts to measure young students' motivation through self-report surveys is that children may not accurately report perceptions or abilities to perform tasks with which they lack experience or are not fully familiar (Nicholls et al., 1985; Wigfield et al., 1998). When coupled with the common issues innate in self-report surveys (e.g., social desirability), research efforts do not yield accurate data sensitive to perception changes (cf. Akcaoglu, 2013). Inaccuracies can potentially negatively impact assessment or prediction efforts. To collect more accurate data, one possible—but unexplored—solution is to use a pretest to provide concrete examples for the complex concepts measured in the survey (e.g., completing a pretest on computer programming before completing a motivation survey on computer programming). The underlying purpose of providing the examples is that students' perceptions may become more contextualized and their perceptions of their ability may be more closely aligned with their performance, leading to more accurate self-reports. Therefore, in this study, our main goal was to investigate the potential impact of a pretest on students' self-reported motivation (i.e., self-efficacy and utility value) for computer programming.

In addition to potential difficulties regarding measuring student motivation for computer programming, there are also known, substantive problems regarding student motivation in the domain of computer science. Motivation constructs are impacted by personal, social, and environmental factors (e.g., Eccles & Wigfield, 2002). Particularly, in computer science, there is a well-established disparity between boys and girls, where boys choose computing courses and degrees more than girls (Cheryan et al., 2009). Males and females also differentially report self-efficacy toward STEM disciplines more broadly (Cheryan et al., 2017; Hill et al., 2010; Lischinski et al., 2016). The reasons for the disparity in perceived competence are likely social, as opposed to some inherent difference between the sexes. However, it is not clear if the provision of concrete examples before taking a self-efficacy assessment can more closely align girls' perceptions of competence with their actual ability.

It is crucial for researchers, as well as those interested in increasing student interest in computer science careers, to be able to accurately assess student motivation (i.e., utility value and self-efficacy) toward computer programming, and the factors, like gender, that impact these perceptions. To address these challenges, we developed a study to test whether providing concrete examples of computer programming tasks before assessing students' perceived value and self-efficacy would impact the accuracy of their self-reports, particularly with regard to reducing gender disparities in self-efficacy for computer programming reported in the literature.

Background

Motivation and programming

Motivation is a theoretical construct used to predict the initiation, direction, intensity, persistence, and quality of human actions (Brophy, 2010; Maehr & Meyer, 1997). Motivational constructs can impact an individual's choice, persistence, and performance in tasks (Wigfield & Eccles, 2000). Motivation is an important but relatively under-explored construct in the teaching and learning of computer science (Lishinski & Yadav, 2019), and measuring or understanding changes in motivation can help us make predictions about future behavior or actions.

This study was centered on two key motivational constructs due to their unique and well-established roles in predicting future preference and performance: utility value and self-efficacy. *Utility value* can be defined as the perceived usefulness of a task (Wigfield & Eccles, 2000). Compared to the intrinsic aspects of value (i.e., enjoyment), utility value is extrinsic in nature (i.e., the desire for future gain). For example, utility value for computer programming would be seeing the relevance of it for people's future lives or careers. One primary role of utility value in the psychology of learning is its role in the development of interest for particular topics or academic domains (Eccles & Wigfield, 2002; Hidi & Renninger, 2006). As individuals develop more utility value for a particular academic topic, their interest in related tasks also increases; therefore, utility value is a key motivational construct for future engagement and continued pursuit in the face of challenges.

Self-efficacy can be defined as peoples' judgment of how well they can perform on an upcoming task (Bandura, 1977). Self-efficacy's importance in academic achievement in diverse educational contexts has been well documented by educational researchers, and in technology-rich contexts specifically (Hodges, 2018). Self-efficacy has also been shown to be related to an individual's choice of college classes and careers (i.e., Lent et al., 1991), and it has been hypothesized that the development of self-efficacy in CS can be a predictor for students' success in their future careers (Lawanto et al., 2017).

Utility value and self-efficacy may be able to help us predict students' future behaviors in CS education and, perhaps, occupations. In the context of this study, understanding students' value and self-efficacy, and what impacts these perceptions, can help us make determinations about the current state of the environment for computer science education. It also can inform the development of interventions to overcome the potential barriers. Furthermore, the application of these specific constructs within the domain of computer science education appears to be unique in the research literature. A recent report (Margulieux et al., 2019) indicated that only 25% of the quantitative studies in the field of computer science education included a validated perception survey and only a few measured interest and utility value. Thus, the approach may benefit research and educators in further extending and refining the knowledge in the field.

Accuracy in self-report measures

Given that measuring motivational outcomes through a self-report survey is a central goal of this research, it also is essential to understand how these perceptions are not always accurate representations of motivation and ability. For example, it is possible for participants to over-report motivation for socially desirable concepts (Rasinski et al., 2005). Further, in a recent report, it was found that teenagers tended to overestimate and over-report their abilities (Jerrim et al., 2019). This effect is more pronounced for males than females (Gutierrez & Price, 2017), which is argued to be due to carelessness in answering the survey or trying to conform to the socially desirable outcomes. Beyond social desirability, participants might not be familiar with the concepts discussed in the survey or misunderstand them, leading to reporting inaccuracies (Blad, 2016; Muis et al., 2014). Considering self-report surveys are widely used instruments to measure motivation and might be fraught with inherent problems, it was our goal to determine whether the problems can be overcome by providing students with a context for the motivation items through a pretest.

To collect more accurate self-report data, in this study a pretest was implemented before a survey. The goal for the pretest was to activate prior knowledge and provide an anchor so that students could make a determination of their motivation for the task. We expected there to be a positive correlation between students' performance on the test and their perceptions, especially in self-efficacy (e.g., Lishinski et al., 2016; Stajkovic et al., 2018).

The mechanism by which providing a pretest before a motivation survey could improve the accuracy of responses can be explained by the concept of priming. Priming in survey research happens when individuals are presented with information before responding to a survey. Research indicates that priming can impact individuals' responses because of the salient information presented during priming (Vitale et al., 2008). In other words, through priming mental representations are activated that can alter a person's response, for example, by creating bias or improving accuracy through clarifying the questions (Rasinski et al., 2005; Tulving & Schacter, 1990). In a recent meta-analysis, priming effects were found to be more pronounced if the participants value (or are primed to value) the outcomes of the measured behavior (Weingarten et al., 2016). Researchers (e.g., Rasinski et al., 2005) have also indicated that more accurate responses, even those socially undesirable, can be achieved through survey design and wording. Within the context of computer science motivation research in general, and this study specifically, a pretest may increase the accuracy of students' responses, providing an improved measure of their motivation and prediction of future outcomes.

Gender and computer programming

There is a well-evidenced disparity between boys and girls in terms of their motivations and career aspirations toward STEM disciplines in general, and specifically in CS (Cheryan et al., 2017; Google Inc. & Gallup Inc, 2016, 2017; Lishinski & Yadav, 2019). According to recent reports by Google and Gallup, women are less likely to work in CS fields; they are encouraged to do so less than men by both their parents and teachers; they are less interested in CS-related fields or engagement in them in the future; and they report less confidence than men, especially as they get older. The difference was present even when girls were as competent as, or even more competent than, boys.

In their extensive work identifying the reasons for the gender gap in some STEM (engineering and physics) and CS fields, Cheryan and her colleagues (2016) found that there were three reasons for the observed disparity between women and men: (a) issues related to stereotypes in those fields (i.e., masculine cultural signals), (b) lack of early exposure, and (c) differences in self-efficacy. Stereotypes surrounding CS, particularly in the classroom, can be unwelcoming to some female students and manipulating these to be non-stereotypical helped increase girls' interest to enroll in an introductory CS course (Master et al., 2016). Positive early experience can also help counteract some of the impacts of the stereotypes. However, it should be noted that in STEM fields such as mathematics, where everyone gets early exposure, there are still disparities between genders (Master et al., 2016), and the stereotype effect emerges as the more prominent factor. These differences are more pronounced as students get older and their perceptions solidify. However, even at early ages, children are aware of group belonging, and they tend to perform better and persist longer if they feel they are part of a group (Master et al., 2017). Although the development of stereotypes is a systematic, cultural phenomenon, providing examples of programming tasks before a survey could help mitigate the gender disparity by impacting the accuracy of self-reported self-efficacy and utility value for girls in our study.

The present study

In this study, we had three overarching research questions. As a part of our first research question, first, we aimed to investigate the impact of the pretest on students' self-efficacy and utility value perceptions, and whether their pretest performance impacted their self-reported perceptions toward CS was necessary. The specific research questions follow:

1a. Is there any difference in utility value toward computer programming between students who took the pretest and those who did not?
1b. Is there any difference in self-efficacy toward computer programming between students who took the pretest and those who did not?
1c. Does pretest performance predict students' self-efficacy toward computer programming?
1d. Does pretest performance predict students' utility value toward computer programming?

Second, we wanted to know whether the previously observed gender-gap existed in our study context. More specifically, we aimed to investigate gender difference in terms of the students self-efficacy and utility value toward CS:

2a. Is there any difference in utility value toward computer programming between girls and boys, after controlling for taking the pretest?
2b. Is there any difference in self-efficacy toward computer programming between girls and boys, after controlling for taking the pretest?

Finally, we aimed to investigate the gender differences by taking into account the students' pretest performances. To investigate this issue, our research questions follow:

3a. Does pretest performance predict self-efficacy toward computer programming differently for girls and boys?
3b. Does pretest performance predict utility value toward computer programming differently for girls and boys?

Methods

Participants and context

The participants for this study were recruited from a middle school in the southeastern United States. The school is designated as a rural fringe, Title 1 school by the U.S. Department of Education. There were 657 students (296 Black, 28 Hispanic, 301 White, 11 Asian, and 21 Other) attending the school, with 347 students eligible for free lunch. Data were collected from a subsample of ($N = 216$) students across sixth, seventh, and eighth grades. There were 121 boys and 92 girls (three students did not report gender). Gender distribution (Table 1) was not significantly different between students in the control and experimental conditions, χ^2 (1, $N = 213$) = 0.030, $p = .863$ (see Table 1).

Instruments

To measure the two dependent variables (i.e., utility value and self-efficacy), we used a survey composed of scales that have previously produced valid evidence in populations of middle school students. The two scales were moderately correlated, indicating that they were unique enough to measure distinct constructs, Pearson's $r = 0.62$, $p < .001$.

To measure utility value, we used an adapted and content-validated version (Akcaoglu, 2013) of the utility value scale developed by Hulleman et al. (2017) (e.g., "I can apply computer programming to the real world."). The 6-item scale had high internal consistency, $a = .912$. To measure self-efficacy, similarly, we used an adapted and content-validated version (Akcaoglu, 2013) of the Academic Efficacy items from Midgley et al. (1996) *Patterns of Adaptive Learning Scales (PALS)* (e.g., "I am certain I can master the computer programming skills."). The self-efficacy scale also had high internal consistency, $a = .929$. In addition, we collected demographic information about students' gender.

We adapted a 6-item test of computer programming (three multiple choice and three short answer items) based on Design by Numbers (DBN), a simple computational design language based on computer programming principles (Maeda, 2001). The DBN programming language was used previously in OECD's Programme for International Student

Table 1. Gender distribution by condition.

	Gender		
Condition	Female	Male	Total
Control	43	58	101
Pretest	49	63	112
Total	92	121	213

Assessment (PISA) to measure students' problem solving (OECD, 2003). The test has been specifically used to measure system analysis and design and has been used in other contexts with middle school students (Akcaoglu, 2014). Having been created following basic programming concepts, in DBN the students were asked to do two different tasks: (a) In multiple-choice questions, students were required to investigate a graphical output and find the appropriate code that produces it (e.g., Figure 1); and (b) in short answer questions, students were asked to write the code that would produce the displayed output graphic. Since DBN included key CS elements and was validated before, it was a suitable proxy for a CS pretest. Students' pretests were evaluated using an answer key. We scored the number of correct answers for each student to calculate their pretest performance score.

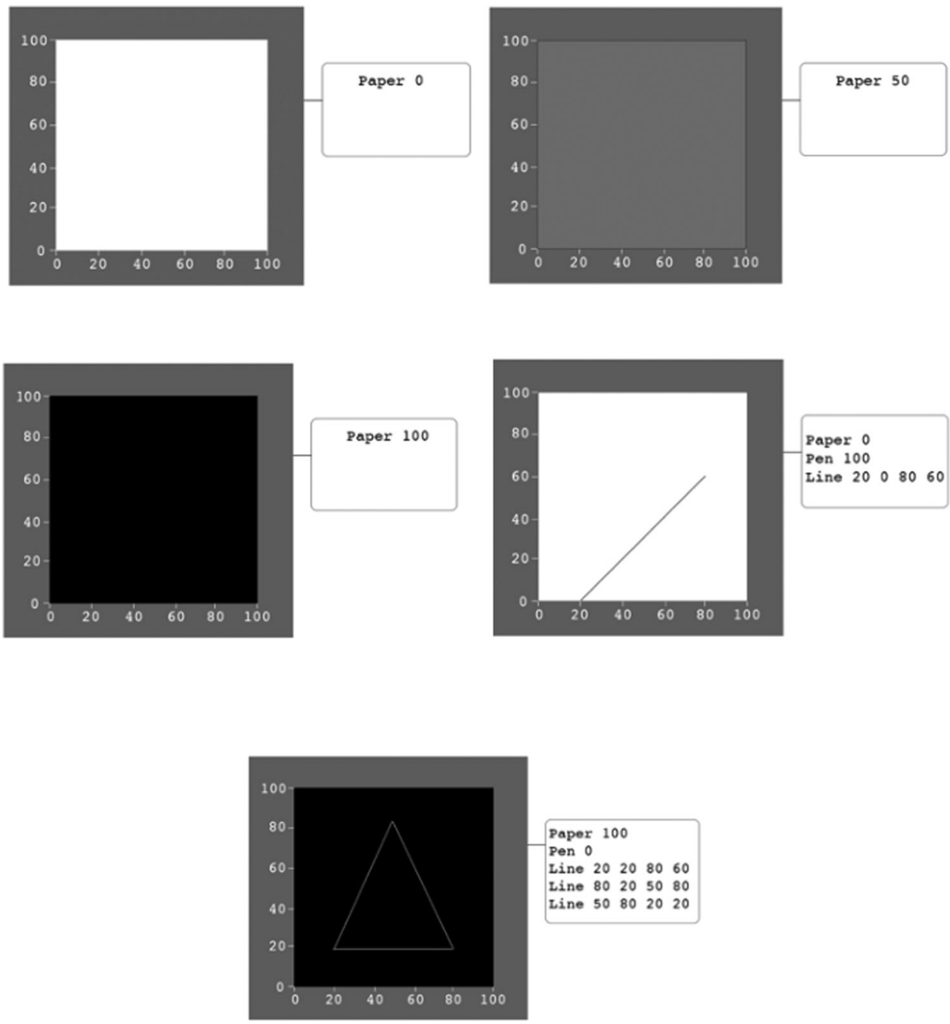

Figure 1. Sample Computer Science Test items based on Design by Numbers (Maeda, 2001).

Procedures

Because we were interested in understanding the impact of the pretest, we employed a quasi-experimental design where we randomly assigned students into two conditions at the beginning of their STEM lab: one with a pretest of computer programming preceding the survey, and one without the pretest. Students in each condition completed their survey in the allocated class time, supervised by their regular classroom teacher. Classroom activities resumed following the completion of the research activity.

Data analysis

We used the JASP statistical software (JASP Team, 2019) to analyze the research data. To analyze the data related to the first research question (i.e., the impact of pretest on the dependent variables) we conducted an independent samples t test. To answer the follow-up research question (i.e., the impact of test performance on the dependent variables), we ran a linear regression with students' pretest scores as an independent variable. Next, to examine the differences between genders, after controlling for testing condition, we conducted two analyses of covariance (ANCOVAs), using gender as the fixed factor, and the pretest (whether students took the pretest or not) as the covariate. Finally, to analyze data related to the third research question (i.e., gender x pretest performance interaction), we conducted an analysis of covariance (ANCOVA) with gender as the fixed factor and students' pretest performance as the covariate, and we investigated the interaction between the two variables.

Results

Results for RQ1: Impact of pretesting on self-efficacy and utility value

To determine if there were any differences in the students' self-reports of utility value and self-efficacy toward computer programming as a result of the pretest, we ran two independent samples t tests. The results suggested that there was not a statistically significant effect of taking the pretest on students' (RQ1a) self-efficacy, $t(214) = 0.66$, $p = .948$, or (RQ1b) utility value, $t(214) = 1.168$, $p = .244$. The effect sizes for the differences were small, Cohen's $d = .009$ and .159, respectively. In other words, taking the pretest of computer programming did not significantly impact how students responded to the motivation items. The means are reported in Table 2.

Next, we investigated whether the students' test performance predicted their self-efficacy or utility value. To examine this, we ran bivariate correlations between the dependent variables and the students' test performance.

Table 2. Means, number of students, and standard deviations in pretest and control groups.

	Group	N	Mean
Self-efficacy	Control	103	3.769
	Pretest	113	3.752
Utility value	Control	103	3.137
	Pretest	113	3.412

The correlation between self-efficacy and students' test performance was significant, $r = 0.196$, $p = .042$. Correlation between utility value and performance was also significant, $r = 0.253$, $p = .007$. The results indicated that increased pretest performance was associated with increased self-efficacy and utility value.

Results for RQ2: Gender and pretest

To examine gender differences in students' self-efficacy and utility value toward computer programming, we conducted two ANCOVAs with gender as the independent variable, pretest as the covariate to account for pretest vs. control conditions, and motivation constructs as dependent variables (RQ2a; self-efficacy, and RQ2b: utility value).

Our results indicated that the there was a significant effect of gender on self-efficacy after controlling for the pretest, $F(1, 210) = 10.573$, $p < .001$. The effect size for the difference was moderate, Cohen's $d = .45$. The covariate, pretest, was not significantly related to self-efficacy, $F(1, 210) = 0.011$, $p = 0.917$.

Our second ANCOVA results indicated that there was not a significant effect of gender on utility value after controlling for the pretest, $F(1, 210) = 2.919$, $p = .089$. There was, however, a small effect size, $d = .27$. The covariate, pretest, was not significantly related to utility value, $F(1, 210) = 1.733$, $p = 0.189$. Means by gender are shown in Table 3.

As seen in Table 3, for each motivation construct, after controlling for students' experimental group, boys reported higher utility value and self-efficacy, although for utility value the difference was not statistically significant.

Results for RQ3: Interaction between test performance and gender

In order to answer the third research question, we conducted two ANCOVAs to investigate the interaction between gender and students' pretest performance on self-efficacy (RQ3a) and utility value (RQ3b).

The interaction between gender and test performance was significant for self-efficacy (RQ3a), $F(1, 108) = 6.512$, $p = .012$. We observed that the mean for girls was significantly lower ($M = 3.332$, $SD = 1.917$) than boys ($M = 4.117$, $SD = 1.959$), with a moderate effect size, $d = .40$. The results

Table 3. Descriptive statistics for the outcome variables split by gender.

	Group	*N*	Mean	*SD*
Self-efficacy	Female	92	3.297	1.688
	Male	121	4.122	1.934
Value	Female	92	3.060	1.575
	Male	121	3.463	1.829

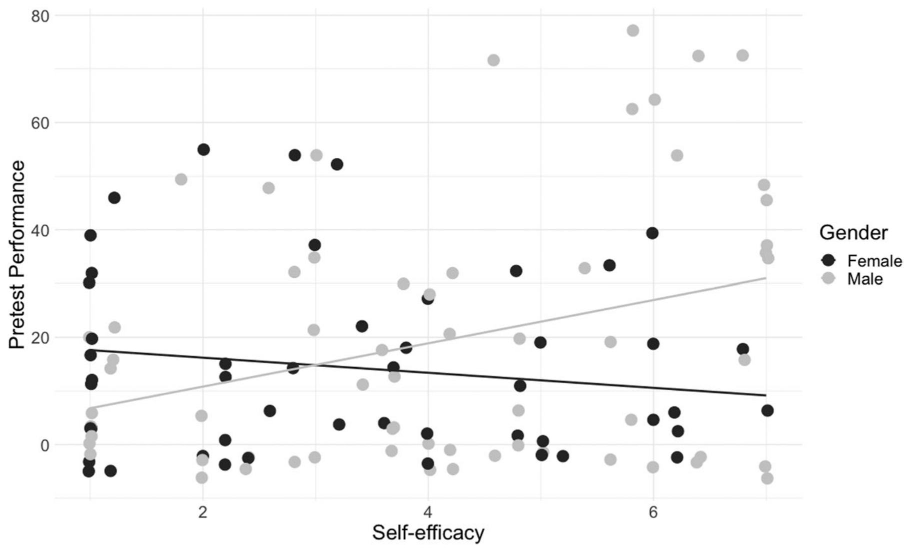

Figure 2. Interaction between gender and test performance for self-efficacy.

indicated that while boys who scored higher in the pretest reported higher self-efficacy, the girls who scored high reported lower self-efficacy (See Figure 2).

Next, we conducted a similar analysis for utility value (RQ3b). The interaction for utility value was also significant, $F(1, 108) = 6.682$, $p = .011$. Again, boys reported higher utility value ($M = 3.673$, $SD = 1.99$) than girls ($M = 3.127$, $SD = 1.721$). The effect size was small ($d = .27$). The results indicated that boys who did well on the pretest also reported higher utility value, while girls who did well on the pretest reported lower utility value (See Figure 3).

Discussion

Key findings

Our aim was to understand whether taking a pretest before a self-report survey had an influence on students' perceptions of utility value and self-efficacy for computer programming. The results showed that taking a pretest was not a significant predictor of students' perceptions. Taking the pretest did not impact students' self-perceptions of utility value and self-efficacy toward computer programming.

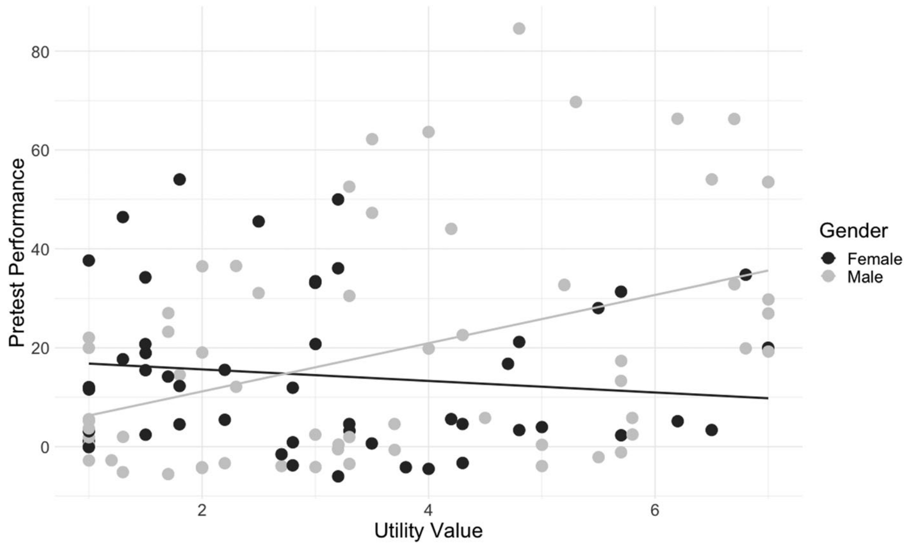

Figure 3. Interaction between gender and test performance for utility value.

Among those who took the test, however, there was an effect of performance on the students' perceptions. The students who performed better on the test reported higher utility value and self-efficacy in general. This effect of test performance can be explained by previous literature on utility value (e.g., Hulleman & Harackiewicz, 2009), where students reported higher utility value after reflecting on usefulness of a task for their future lives (Hulleman et al., 2017). In a similar vein, the pretest might have worked as a *relevance* task, where students had a chance to actually see what computer programming meant, and thus helped the test-takers to perceive more utility value toward computer programming. Priming literature also suggests that during priming new learning can occur and impact the outcomes (Lenz, 2009). It is therefore possible that students' performance might be an indication of things being "learned" about computer programming, leading to the observed differences.

In addition to investigating the impact of a pretest and test performance, we also explored influence of gender on students' perceptions toward computer programming. Our results confirmed the findings from previous research (e.g., Beyer, 2014; Cheryan et al., 2017): Boys reported statistically significantly higher self-efficacy of programming, and utility value that is suggestive of significant difference between genders.

These results were more pronounced and interesting when we took the students' pretest performance into account and investigated the interaction. We observed that despite their high pretest performance, girls reported lower self-efficacy and utility value than boys with similar pretest performance. The reason for this reported difference could be, as previously explained in Cheryan et al. (2017), related to perceiving computer science fields as masculine (Master et al., 2017), encountering opportunities for CS rarely, or lacking role models, which, in combination, contribute to lower self-efficacy and representation of girls in CS fields. Our results indicate that despite acceptable performance in CS, girls still report lower utility value and self-efficacy. Although such an effect has been previously reported in undergraduate engineering contexts (e.g., Huang & Brainard, 2001) where women reported lower self-efficacy despite higher grades, we know of no other studies that investigated this issue in K-12 CS contexts. An explanation can, however, be found in a recent study conducted in an undergraduate CS setting. Lishinski et al. (2016) found that females self-efficacy improved significantly when they were able to learn about their CS performance. With the lack of such feedback (e.g., project work with slower feedback vs. exams with quicker feedback), females' self-efficacy stayed the same, while it significantly increased for males. When we interpret our finding within the light of Lishinski's work, we can argue that a similar mechanism could have been at play in our context. The girls in our study took the pretest but did not know how they did (because of the nature of the study design, the results of the pretest were not reported). They did not receive performance feedback. This finding from our study, in corroboration with Lishinski's work (2016), points to the importance of self-efficacy interventions and the impact of feedback for female students' self-efficacy development and participation in CS activities.

Limitations and future directions for research

The study took place at a Title-1 rural middle school in the southeastern United States. Therefore, the results of the present study may be more applicable to schools in a similar context. Replicating the study in other contexts that are different from the current study can help explain the contextual limitations (see Hodges, 2015 and Spector et al., 2015). Although we have identified differences between boys and girls, due to the descriptive nature of this study, we cannot speak to the causes of these differences. Based on previous studies (e.g., Master et al., 2017), a stereotype threat seems to be a possible causal factor. Finally, our study design can be improved by a new quasi-experimental design. In such a design, in addition to the pretest and control groups, a third posttest-only group could be included. With such an addition, the effect of pretest could be more accurately modeled.

Conclusion

Given the increased demand (and occupational and societal opportunities) for computer science students and professionals, knowing how motivated students are in this domain is important in its own right and is necessary to develop instruction and interventions to overcome potential inequities or barriers. The results from the current study helped confirm some of the previous findings in the context of computer programming motivation and added new knowledge to K-12 CS education. We found that taking a pretest did not lead to significant differences, but it positively impacted the test-takers depending on their performance. Given that it might be working as a relevance connection, students might benefit from these simple interventions, helping them anchor their thoughts. In addition to finding gender differences in reporting, we also noted a difference between genders despite their actual performance. Girls, even when they performed well in the pretest, reported lower self-efficacy and utility value compared to boys with similar performance. This is an alarming finding pointing to the importance of CS interventions to decrease the gender gap both while advertising and teaching CS courses. For example, interventions can be designed for the purpose of increasing girls' belonging perceptions in these fields by creating early positive experiences and providing positive feedback about their performance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Akcaoglu, M. (2013). *Cognitive and motivational impacts of learning game design on middle school children* (Unpublished doctoral dissertation). Michigan State University.

Akcaoglu, M. (2014). Learning problem-solving through making games at the game design and learning summer program. *Educational Technology Research and Development*, *62*(5), 583–600. doi:10.1007/s11423-014-9347-4

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, *84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Beyer, S. (2014). Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education*, *24*(2–3), 153–192. doi:10.1080/08993408.2014.963363

Blad, E. (2016). Students help shape measures of 'soft skills. *Education Week*, *27*(35), 10–11.

Brophy, J. (2010). *Motivating students to learn*. Routledge.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045–1060. doi:10.1037/a0016239

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. doi:10.1037/bul0000052

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. doi:10.1146/annurev.psych.53.100901.135153

Google Inc. & Gallup Inc. (2016, March). *Computer science learning: Closing the gap: Girls. Findings from diversity gaps in computer science: Exploring the underrepresentation of girls, Blacks and Hispanics* (Issue Brief No. 2). https://goo.gl/dF7shN and https://nsf.gov/statistics/2016/nsb20161.

Google Inc. & Gallup Inc. (2017, December). *Images of computer science: Perceptions among students, parents and educators in the U.S. Results from the 2015-2016 Google-Gallup Study of Computer Science in U.S. K-12 Schools*. (Issue Brief No. 5). https://edu.google.com/latest-news/research.

Gutierrez, A. P., & Price, A. F. (2017). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *The Journal of Experimental Education*, *85*(3), 486–500. doi:10.1080/00220973.2016.1180278

Hecht, C. A., Harackiewicz, J. M., Priniski, S. J., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Promoting persistence in the biological and medical sciences: An expectancy-value approach to intervention. *Journal of Educational Psychology*, *111*(8), 1462–1477. doi:10.1037/edu0000356

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. doi:10.1207/s15326985ep4102_4

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women.

Hodges, C. B. (2015). Replication studies in educational technology. *TechTrends*, *59*(4), 3–4. doi:10.1007/s11528-015-0862-2

Hodges, C. B. (Ed.). (2018). *Self-efficacy in instructional technology contexts*. Springer. doi:10.1007/978-3-319-99858-9

Huang, P. M., & Brainard, S. G. (2001). Identifying determinants of academic self-confidence among science, math, engineering, and technology students. *Journal of Women and Minorities in Science and Engineering*, *7*(4), 24–337. doi:10.1615/JWomenMinorScienEng.v7.i4.40

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412.

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology*, *109*(3), 387–404. doi:10.1037/edu0000146

JASP Team. (2019). JASP (Version 0.11.1)[Computer software].

Jerrim, J., Parker, P., & Shure, N. (2019, April). Bullshitters: Who are they and what do we know about their lives? *Discussion paper series. IZA – Institute of Labor Economics.* http://ftp.iza.org/dp12282.pdf.

Lawanto, K., Close, K., Ames, C., & Brasiel, S. (2017). Exploring strengths and weaknesses in middle school students' computational thinking in Scratch. In P. Rich & C. B. Hodges (Eds.), *Emerging research, practice, and policy on computational thinking* (pp. 307–326). Springer. doi:10.1007/978-3-319-52691-1_19

Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology*, *38*(4), 424–430. doi:10.1037/0022-0167.38.4.424

Lenz, G. S. (2009). Learning and opinion change, not priming: Reconsidering the priming hypothesis. *American Journal of Political Science*, *53*(4), 821–837. doi:10.1111/j.1540-5907.2009.00403.x

Lishinski, A., & Yadav, A. (2019). Motivation, attitudes, and dispositions. In S. Fincher & A. Robins (Eds.), *The Cambridge handbook of computing education research* (pp. 801–826). Cambridge University Press. doi:10.1017/9781108654555.029

Lishinski, A., Yadav, A., Good, J., & Enbody, R. (2016, August). Learning to program: Gender differences and interactive effects of students' motivation, goals, and self-efficacy on performance. *Proceedings of the 2016 ACM Conference on International Computing Education Research* (pp. 211–220). ACM.

Maeda, J. (2001). *Design by numbers*. MIT press.

Maehr, M. L., & Meyer, H. A. (1997). Understanding motivation and schooling: Where we've been, where we are, and where we need to go. *Educational Psychology Review*, *9*(4), 371–409. doi:10.1023/A:1024750807365

Margulieux, L., Ketenci, T. A., & Decker, A. (2019). Review of measurements used in computing education research and suggestions for increasing standardization. *Computer Science Education*, *29*(1), 49–78. doi:10.1080/08993408.2018.1562145

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, *108*(3), 424–437. doi:10.1037/edu0000061

Master, A., Cheryan, S., & Meltzoff, A. N. (2017). Social group membership increases STEM engagement among preschoolers. *Developmental Psychology*, *53*(2), 201–209. doi:10.1037/dev0000195

Midgley, C., Maehr, M. L., Ludmila, Z. H., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (1996). *Patterns of adaptive learning survey (PALS)*. Ann Arbor, MI: Center for Leadership and Learning.

Muis, K. R., Duffy, M. C., Trevors, G., Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Education Research*, *2*(1), 17–32. doi:10.12735/ier.v2i1p17

Nicholls, J. G., Patashnick, M., & Nolen, S. B. (1985). Adolescents' theories of education. *Journal of Educational Psychology*, *77*(6), 683–692. doi:10.1037/0022-0663.77.6.683

Rasinski, K. A., Visser, P. S., Zagatsky, M., & Rickett, E. M. (2005). Using implicit goal priming to improve the quality of self-report data. *Journal of Experimental Social Psychology*, *41*(3), 321–327. doi:10.1016/j.jesp.2004.07.001

Spector, J. M., Johnson, T. E., & Young, P. A. (2015). An editorial on replication studies and scaling up efforts. *Educational Technology Research and Development*, *63*(1), 1–4. doi:10.1007/s11423-014-9364-3

Stajkovic, A. D., Bandura, A., Locke, E. A., Lee, D., & Sergent, K. (2018). Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: A meta-analytic path-analysis. *Personality and Individual Differences*, *120*, 238–245. doi:10.1016/j.paid.2017.08.014

Tran, Y. (2018). Computer programming effects in elementary: Perceptions and career aspirations in STEM. *Technology, Knowledge and Learning*, *23*(2), 273–299. doi:10.1007/s10758-018-9358-z

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306. doi:10.1126/science.2296719

Vitale, D. C., Armenakis, A. A., & Feild, H. S. (2008). Integrating qualitative and quantitative methods for organizational diagnosis: Possible priming effects? *Journal of Mixed Methods Research*, *2*(1), 87–105. doi:10.1177/1558689807309968

Yadav, A., Hong, H., & Stephenson, C. (2016). Computational thinking for all: Pedagogical approaches to embedding 21st century problem solving in K-12 classrooms. *TechTrends*, *60*(6), 565–568. doi:10.1007/s11528-016-0087-7

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, *142*(5), 472–497. doi:10.1037/bul0000030

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). Chapter 3: The development of children's motivation in school contexts. *Review of Research in Education*, *23*(1), 73–118.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. doi:10.1006/ceps.1999.1015