International Society of the Learning Sciences

CS1 Student Assessments of Themselves Relative to Others: The Role of Self-Critical Bias and Gender

Jamie Gorson, Northwestern University, jgorson@u.northwestern.edu
Eleanor O'Rourke, Northwestern University, eorourke@northwestern.edu

Abstract: Introductory computer science courses at the university level (CS1) present a number of challenges. Students enter CS1 with varying backgrounds, and often assess their ability while choosing a major. Studies show that students often negatively self-assess in response to natural programming moments, such as getting a syntax error, but we have a limited understanding of the mechanisms that drive these self-assessments. In this paper, we study differences in student assessments of themselves and others in response to particular programming moments. We analyze survey data from 214 CS1 students, finding that many have a *self-critical bias*, evaluating themselves more harshly than others. We also found that women have a stronger self-critical bias, and that students tend to be more self-critical relative to women. These insights can help us reduce the impact of negative self-assessments on student experiences.

Introduction

The rapid growth of computing has prompted learning scientists to study approaches for broadening participation in the field and creating positive learning experiences for students. In this paper, we study student experiences in introductory computer science courses at the university level, often referred to as CS1. CS1 courses are the entry point to the computer science major, but often suffer from learning challenges as they serve students with a wide range of prior experience in computing. Specifically, students who are new to programming often struggle when grouped in classes with more experienced peers (Ott et al., 2018), and feel pressure to quickly evaluate their ability to succeed in the field due to the need to choose a major (Lewis et al., 2011). These challenges are often amplified for women and students of color, who are underrepresented in computer science and who drop out of the major at higher rates (Buzzetto-More et al., 2010; Cohoon, 2006; Fischer & Margolis, 2002).

Recent research in computing education has established that students in CS1 frequently assess their own ability (Gorson & O'Rourke, 2019; Kinnunen & Simon, 2012). While evaluating progress is important for self-regulation (Butler & Winne, 1995), these studies show that students negatively assess themselves in response to programming moments that are natural parts of professional practice (LaToza et al., 2006; Perscheid et al., 2017), which therefore are not helpful indicators of ability. For example, many students believe that they are performing poorly when they stop to think or plan, get a compiler error, or forget syntax (Gorson & O'Rourke, 2019; Gorson & O'Rourke, 2020). Furthermore, students who negatively self-assess more frequently and more strongly in response to these moments have a lower self-efficacy on average (Gorson & O'Rourke, 2020), an important factor that predicts student persistence in computer science (Lewis et al., 2011; Miura, 1987).

To address these issues in student experience and persistence, researchers and practitioners need to design environments and curricula to reduce the prevalence and impact of overly negative self-assessments. In order to work towards these designs, we first need a better understanding of the mechanisms that drive these self-assessments. While there are likely many factors that contribute to student self-assessments, in this paper we specifically explore how students assess themselves in comparison to how they assess others. Studies show that people tend to have a self-enhancement bias when evaluating themselves compared to others, rating themselves more favorably (Alicke, 1986; Kwan et al., 2004). At the same time, women tend to under-evaluate their performance in science (Ehrlinger & Dunning, 2003), and have a weaker self-enhancement bias (Kurman, 2004). Given the challenging learning context that CS1 presents for many students, we were interested in understanding how factors such as gender might shape negative self-assessments.

Student self-assessments in CS1

Multiple studies have shown that students frequently assess their own ability in CS1 (Gorson & O'Rourke, 2019; Lewis et al., 2011; Kinnunen & Simon, 2012). Through an interview study, Lewis et al. (2011) identified that self-assessments play a large role in CS1 students' decisions to major in CS. Kinnunen & Simon (2012) conducted an interview study exploring how CS1 students experience the programming process and found that students often have negative evaluations of programming episodes after successful programming experiences, particularly when the experience does not match their expectations. More recently, we built on this research through two in-depth studies of the criteria that students use to self-assess. We found that many of the criteria that students believe are indicators of programming ability refer to moments that arise while programming, such as writing code that runs on the first try or debugging easily (Gorson & O'Rourke, 2019). Furthermore, most of these moments are natural parts of professional practice (LaToza et al., 2006; Perscheid et al., 2017) and are therefore not accurate measures of ability. To explore these programming moments, we conducted a survey study with 214 students from three universities, and found that many students at all three schools reported that they negatively self-assess during the thirteen moments included in the survey, despite differences in curriculum and population (Gorson & O'Rourke, 2020). Furthermore, these moments may influence student experience, since students who report stronger negative self-assessments tended to have lower self-efficacy, suggesting a relationship between these two factors.

This body of research highlights the prevalence of negative self-assessments in CS1. However, we still have a limited understanding of the factors that drive these assessments. Our previous study revealed correlations between students' beliefs about professional practice and self-assessments for a few of the moments (Gorson & O'Rourke, 2020), however these effects do not fully explain students' negative views of their own ability. In this paper, we explore another factor that could explain negative self-assessments, namely differences in how students assess themselves in comparison to others and the role of gender in these self-assessments.

Differences in assessments of the self and assessments of others

Social psychologists have studied the differences in the ways individuals assess themselves and others extensively; see Kwan et al. (2004) for a review of the literature. In many domains, people tend to hold overly positive views of their own abilities, an effect that is referred to as *self-enhancement bias* (Alicke, 1986; Brown, 1986; Kwan et al., 2004). For example, in an early study Alicke (1985) asked college students to rate the degree to which a set of trait adjectives characterized themselves and the average college student, finding that students rated themselves significantly higher than others for desirable traits. Studies show that this overestimation arises in part due to a lack of metacognitive awareness of one's own weaknesses (Kruger & Dunning, 1999), as well as a natural tendency for humans to be overly optimistic about their own abilities (Eva & Regehr, 2008).

Self-enhancement bias during self-assessments does not necessarily arise as strongly for students who experience stereotype threat. Ehrlinger and Dunning (2003) gave college students a pop quiz on scientific reasoning and found that female students rated themselves more negatively than male students on scientific skills and estimated performance on the quiz, even though there were no gender differences in actual performance. While we are not aware of any studies of self-enhancement in the domain of computer science, we might expect to see similar effects as in other STEM domains. Women are notably underrepresented in computer science (Cohoon, 2006; Fisher & Margolis, 2002) and prevalent stereotypes depict computer scientists as male, technologically oriented, and socially awkward (Master et al., 2016). Given this context and the frequent negative self-assessments of CS1 students (Gorson & O'Rourke, 2019), we wondered if there are any differences in student assessments of themselves and others, and how gender might shape these assessments.

Research questions

The goal of this paper is to study differences in how students assess themselves and others in response to particular moments that arise during the programming process. Towards this end, we conducted a secondary analysis of the data from our previous survey (Gorson & O'Rourke, 2020) to answer two research questions: (1) are there differences in students' assessments of themselves and their assessments of others? (2) Are the differences in these assessments impacted by the gender of the student or the other? We aim to understand whether any self-assessment biases might help explain the prevalence of the negative self-assessments while programming in CS1.

Methods

This paper reports on a secondary analysis of data we collected in February 2019 (Gorson & O'Rourke, 2020). We recruited participants from three universities of different types and with different levels of selectivity in the midwestern United States. All participants were enrolled in an introductory course at their university and 36% of participants identified as female. See our previous paper for a complete description of the data collection methods. In this section, we review the subset of the survey that is most relevant for understanding the present analysis.

The survey was designed to uncover student self-assessments in response to specific moments that might arise during the programming process. When observing students programming, moments such as struggling with a compiler error or stopping to plan will occur sporadically and inconsistently. To address this concern, we designed a set of thirteen vignettes that each describe a fictional character encountering one of the programming moments that may prompt negative self-assessments. This is an example of a vignette:

> *Diego starts working on a programming problem. He writes a few lines of code. He realizes that he is confused about what to do next. He pauses and plans his next steps. Diego wishes that he did not have to stop writing code to plan.*

After each vignette, students were asked how much they agree with two statements, one about the character and one about themselves. For this vignette, the statements were: *"Since Diego had to stop and think, he didn't do well on the problem"* and *"When I have to stop programming to plan, I feel like it means that I'm not doing well on the problem."* Students were asked to rate their agreement with each statement on a six-point forced-choice Likert scale. Through these survey questions, we hoped to elicit students' self-assessments in response to a wide range of moments without requiring that these moments arise naturally during an observation.

The gender of the vignette character was communicated through the character's name and the pronouns used in the vignette. To control for any biases in participant responses based on the character's gender, we randomized the names of the characters across vignettes and participants.

Findings

Students evaluate themselves more critically than they evaluate others

To answer our first research question, we measured whether there were differences in students' assessments of themselves and their assessments of the vignette characters. We first converted the responses to the two forced-choice Likert-scale questions following each vignette to a numerical scale ranging from -3 (strongly disagree) to 3 (strongly agree). By agreeing to a statement that follows a vignette, participants demonstrate a belief that they (or the characters) are performing poorly during that moment. Therefore, to calculate self-critical bias we subtracted their response to the question about themselves from their response to the question about the character for each vignette question. For example, a participant may *slightly agree* (1) that the character is performing poorly in a particular moment, and *slightly disagree* (-1) that they are preforming poorly. We would calculate 1 minus -1 resulting in a self-enhancement bias of 2 for that vignette. A positive value indicates a self-enhancement bias in participants' responses. After calculating these self-enhancement biases, we grouped participants into three categories for each vignette: those who exhibited a positive self-enhancement bias, those who exhibited no bias, and those who exhibited a negative self-enhancement bias.

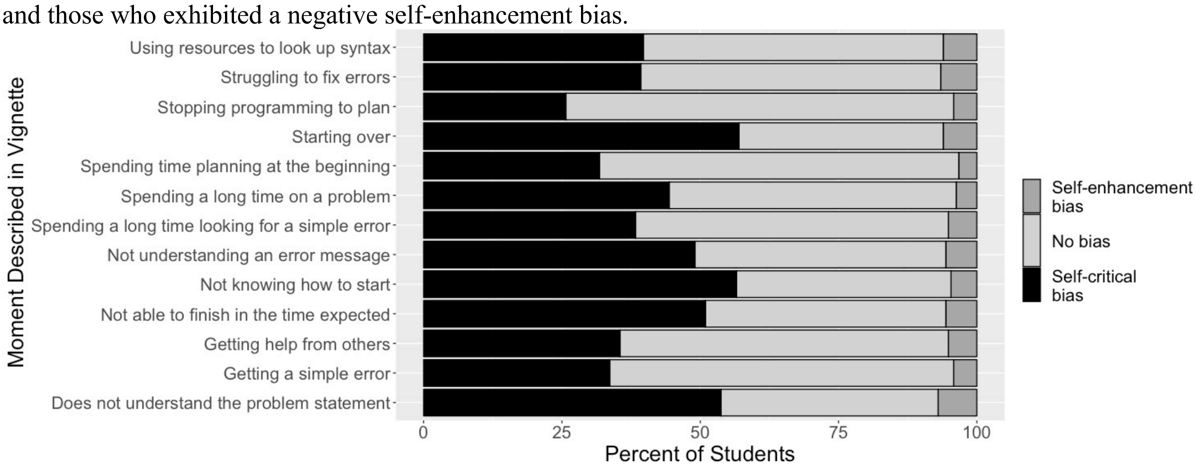

Figure 1. Graph showing the percentage of participants who exhibited a self-enhancement bias, a self-critical bias, and no bias, when comparing their responses to the questions following each of the thirteen vignettes.

Surprisingly, we found that very few students exhibited a self-enhancement bias. As shown in Figure 1, only 3-5% of participants had a positive self-enhancement bias for each question, while 25-60% of students had a negative self-enhancement bias. For the rest of this paper, we refer to a negative self-enhancement bias as a *self-critical bias*. Overall, these findings suggest that CS1 students tend to be more critical of themselves than of others, which is surprising given previous findings on the prevalence of self-enhancement biases.

This self-critical bias could manifest in two degrees of severity. Students could believe that both they and the character are performing poorly or well, but differ in the strength of the assessments (e.g. *slightly agree* for the character, *agree* for themselves). Or, students could believe that they are performing poorly (e.g. *slightly agree*) but make no negative assessment of the character (e.g. *disagree*), and vice versa. We were interested in measuring how often students assessed themselves in a different direction than the character, so for each vignette, we counted how often participants' responses to the two vignette questions fell on different sides of the Likert scale. We found that only 1-4% of participants negatively assessed the character but not themselves for each vignette, while 9-36% of participants negatively assessed themselves but not the character for each vignette. Furthermore, we found that 85% of students negatively assessed themselves but not the character for at least one vignette, while only 18% of students assessed the character but not themselves for at least one vignette. This effect was most common for the moments *starting over* and *does not understand the problem statement*, in which 35% and 33% of students assessed themselves but not the character negatively. These findings show that a significant number of students negatively assess themselves at moments that they think are acceptable for others.

Self-critical bias is stronger when the student or the vignette character is female

Next, we analyzed whether self-critical bias was influenced by gender. We used non-parametric methods because the Shapiro-Wilk test showed that our data has a non-normal distribution. For the remaining analyses, we did not include the five students who reported non-binary gender identities because we feared the size of the group would result in an inaccurate representation of their experience. For the binary students, we conducted a Mann-Whitney U test, and found that female students were significantly more likely to have a self-critical bias than male students ($Z = 3484.5$, $p < 0.001$), with a median bias of 0.46 for male students and 0.92 for female students.

Women are underrepresented in computer science, and prevalent stereotypes depict computer scientists as technologically-oriented males (Master et al., 2016). As a result, we wondered whether students might assess themselves differently in relation to female and male vignette characters. We averaged the self-critical bias that each participant reported for the vignettes with female characters and the vignettes with male characters. Then, we conducted a Wilcoxon signed-rank test, a non-parametric paired t-test, between these two scores and found that participants were significantly more self-critical when the vignette character was female ($Z = 8688$, $p < 0.05$). The median self-critical bias was 0.57 when the vignette character was male, and 0.71 when they were female.

After finding that students are generally more critical of themselves in comparison to female vignette characters, we wondered whether this effect is influenced by the gender of the participant. To answer this question, we conducted an Aligned Rank Transform, a non-parametric ANOVA. We used the same self-critical bias scores for female and male vignette characters described above. We found that the gender of the character ($F(1, 206) = 4.39$, $p < 0.05$) and the gender of the participant ($F(1,206) = 14.90$, $p < 0.001$) both had significant effects on self-critical bias. While this effect appears to be stronger for male students, we did not find a significant interaction ($F(1, 206) = 0.36$, n.s.). We often think of female students as being most affected by stereotypes about who belongs in computer science, however these findings show that male students are also influenced by these narratives.

Discussion and design implications

The goal of this research was to study the differences in how students assess themselves and others in response to particular moments that arise during the programming process. Through a secondary analysis of our survey data (Gorson & O'Rourke, 2020), we found that very few students exhibit a self-enhancement bias in this domain. Instead, we found that many students exhibit what we call a *self-critical bias*, with 96% of participants rating themselves more harshly than the vignette character in response to at least one vignette. We also found that female students are significantly more likely to have a self-critical bias than male students. Finally, we found that both male and female students were more self-critical when the vignette character was female. This research has some limitations that should be addressed through future work. Due to the quantitative nature of our data, we cannot explain why some students evaluate themselves more critically than the vignette characters. Furthermore, while this study focused on the role of gender on self-assessments, many other factors might influence students' self-assessments, including race, ethnicity, sense-of-belonging, and perceptions of professionals.

Given the established phenomena of self-enhancement bias in other domains, we were surprised to see the prevalence of self-critical bias for both male students and female students in this context. We believe these findings have important implications for the design of CS1 curricula and interventions. For example, we previously argued that CS1 courses should explicitly teach students about professional programming practices to help students develop accurate expectations about the programming experience and to reduce their negative self-assessments (Gorson & O'Rourke, 2020). While this could help some students, our new findings suggest that some students who negatively self-assess in response to these moments do not view them as universal signs of poor performance. Instead, they view these moments as more problematic for themselves than for others. To better support this group of students, CS1 teaching staff could call attention to moments when students may be assessing themselves particularly harshly, and help students reframe their perceptions of these moments. Additionally, our findings reveal that students have a stronger self-critical bias when the vignette character is female, suggesting that the stereotypes about who belongs in computer science may lead students to have lower expectations of women. We believe this provides compelling evidence for designing diversity events and initiatives that help both male and female students shift their expectations to see women as belonging and excelling in computer science.

Acknowledgments

This work was supported by the NSF Grant IIS-1755628. Thank you Delta Lab and Mesmin Destin for feedback.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of personality and social psychology, 49*(6), 1621.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social cognition, 4*(4), 353-376.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research, 65*(3), 245-281.

Buzzetto-More, N. A., Ukoha, O., & Rustagi, N. (2010). Unlocking the barriers to women and minorities in computer science and information systems studies: Results from a multi-methodical study conducted at two minority serving institutions. *Journal of Information Technology Education: Research, 9*(1), 115-131.

Cohoon, J. M. (2006). Just get over it or just get on with it: Retaining women in undergraduate computing. *In J. Cohoon & W. Aspray (Eds.), Women and information technology: Research on underrepresentation*, 205-238.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of personality and social psychology, 84*(1), 5.

Eva, K. W., & Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions, 28*(1), 14-19.

Fisher, A., & Margolis, J. (2002). Unlocking the clubhouse: the Carnegie Mellon experience. *ACM SIGCSE Bulletin, 34*(2), 79-83.

Gorson, J., & O'Rourke, E. (2019, July). How Do Students Talk About Intelligence? An Investigation of Motivation, Self-efficacy, and Mindsets in Computer Science. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (pp. 21-29).

Gorson, J., & O'Rourke, E. (2020, August). Why do CS1 Students Think They're Bad at Programming? Investigating Self-efficacy and Self-assessments at Three Universities. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 170-181).

Kinnunen, P., & Simon, B. (2012). My program is ok–am I? Computing freshmen's experiences of doing programming assignments. *Computer Science Education, 22*(1), 1-28.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology, 77*(6), 1121.

Kurman, J. (2004). Gender, self-enhancement, and self-regulation of learning behaviors in junior high school. *Sex Roles, 50*(9-10), 725-735.

Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological review, 111*(1), 94.

LaToza, T. D., Venolia, G., & DeLine, R. (2006, May). Maintaining mental models: a study of developer work habits. In Proceedings of the 28th international conference on Software engineering (pp. 492-501).

Lewis, C. M., Yasuhara, K., & Anderson, R. E. (2011, August). Deciding to major in computer science: a grounded theory of students' self-assessment of ability. In *Proceedings of the Seventh International Workshop on Computing Education Research* (pp. 3-10).

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.

Miura, I. T. (1987). The relationship of computer self-efficacy expectations to computer interest and course enrollment in college. *Sex roles, 16*(5-6), 303-311.

Ott, L., Bettin, B., & Ureel, L. (2018, July). The impact of placement in introductory computer science courses on student persistence in a computing major. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 296-301).

Perscheid, M., Siegmund, B., Taeumel, M., & Hirschfeld, R. (2017). Studying the advancement in debugging practice of professional software developers. Software Quality Journal, 25(1), 83-110.