DOI:
10.1145/3265757.3265773

Document Version
Peer reviewed version

Citation for published version (APA):
Kallia, M., & Sentance, S. (2018). Are boys more confident than girls? The role of calibration and students' self-efficacy in programming tasks and computer science. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education: WIPSCE '18* https://doi.org/10.1145/3265757.3265773

Are boys more confident than girls? The role of calibration and students' self-efficacy in programming tasks and computer science

Maria Kallia
King's College London
London
maria.kallia@kcl.ac.uk

Sue Sentance
King's College London
London
sue.sentance@kcl.ac.uk

ABSTRACT

Computer programming is regarded as a difficult subject at both school and university. There have been a vast amount of studies with a focus on identifying students' difficulties, common errors and misconceptions in programming, and on the development and design of instructional techniques that could potentially help students overcome these difficulties. Nevertheless, there are few studies that explore students' performance in programming under the prism of self-regulation theory. To this end, the current study considers girls' and boys' calibration and how it is related with their performance in programming, self-evaluation, and self-efficacy in computer science. Calibration is a measure of the accuracy with which people assess their confidence in their own performance. The results of our study suggest that boys feel significantly more efficacious in computer science than girls, as well as make significantly more accurate predictions (better calibrated) of their programming performance than girls. The implications of these findings for the current education practices are outlined and discussed.

CCS CONCEPTS

• **Social and professional topics → CS1**; **K-12 education**;

KEYWORDS

self-efficacy, computer science education, computer programming, calibration, gender differences

ACM Reference Format:
Maria Kallia and Sue Sentance. 2018. Are boys more confident than girls? The role of calibration and students' self-efficacy in programming tasks and computer science. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education (WiPSCE '18), October 4–6, 2018, Potsdam, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3265757.3265773

1 INTRODUCTION

Learning computer programming is a task that requires effort, practice and resilience and can be characterised as an arduous journey where the learner encounters multifaceted problems. It is true that students who learn to program make lots of errors ranging from simple syntactic errors to deeper ones whose sources are often hidden and difficult to be identified and explained. While identifying students' common errors and misconceptions or misunderstandings is the focus of many research studies (e.g. [4, 7–9, 12, 17, 22]), there is limited research that examines if and how students' programming performance can impact students' perceived ability (self-efficacy) in computer science in general and the role of students' calibration and self-evaluation abilities in computer programming; this pedagogical issue is of high importance. Calibration is a measure of the relationship between confidence in performance and accuracy of performance [24].

The current study aims to address the above gaps by exploring the relationship between boys' and girls' performance in programming, self-evaluation, and self-efficacy in computer science and by investigating students' calibration in computer programming tasks and its potential relationship with the above variables. The main research questions of the study are the following:

- RQ1: Are boys better calibrated than girls in computer programming?
- RQ2: Do boys feel more efficacious in computer science than girls?
- RQ3: Does boys' and girls' self-efficacy in computer science correlate with their performance and self-evaluation in computer programming?

2 BACKGROUND

2.1 Self-efficacy

The concept of self-efficacy is situated in the social cognitive theory of Albert Bandura [2] who defines self-efficacy as *"people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives"* [3, p. 71]. It is a central powerful belief about one's capability to accomplish and perform a task and designates the *"mediation between social experience, individual thinking and behavior"* and, thus, is a cognitive function that supports individual's behavior [5, p. 3]. Self-efficacy beliefs are shaped by the following sources: mastery experience, vicarious experience, verbal persuasion and psychological states [3].

Studies that evaluate students' self-efficacy beliefs in computer programming are limited. Ramalingam and Wiedenbeck [18] were among the first researchers to develop a self-efficacy scale (Computer programming self-efficacy scale, CPSES) for programming courses. Studies that explore students' self-efficacy in computer programming (e.g. [25], [1], [13], [16]) generate controversial results

WiPSCE '18, October 4–6, 2018, Potsdam, Germany

ACM ISBN 978-1-4503-6588-8…$15.00
https://doi.org/10.1145/3265757.3265773

Table 1: Descriptive statistics by Gender

	Performance Mean/Median	Self-Evaluation Mean/Median	Self-Efficacy in CS Mean/Median	Bias Mean/Median	Accuracy Mean/Median
Girls	5.84/6.00	5.28/5.00	5.92/6.00	-.56/-1.00	8.16/8.00
Boys	6.34/7.00	6.47/7.00	7.02/7.00	.13/.00	8.66/9.00

with some supporting self-efficacy's impact on students' performance while others fail to support this relationship.

2.2 Self-evaluation and Calibration

In addition to self-efficacy, self-evaluation is one of the central processes in self-regulation and it refers to the individual's evaluation of his/her performance [19]. These judgments must be as accurate as possible in order for the self-regulated process to be effective [26]. As a meta-cognitive judgment of an individual's perceived performance and the actual one on a task, calibration has a critical role in learning [19]. Specifically, calibration refers to the degree that (students') evaluations echo an actual performance. Measures of calibration that are usually used in research are the bias or "direction of the errors in judgment" and it is calculated by subtracting the actual performance from the students' self-evaluation score [15] [19, p. 29]. A zero bias indicates perfect calibration, a positive suggests overconfidence and a negative under-confidence [19]. The second measure is the calibration accuracy which is the extent of the judgment error and it is calculated by subtracting the absolute value of the bias score from its maximum possible value [19]. For instance, the calibration accuracy for a self-evaluation Likert scale (0 to 10) will range from 0 suggesting a complete inaccuracy to 10 suggesting complete accuracy. Accurate calibration is focal for functioning and it is important in meta cognition and, therefore, in self-regulation [21].

3 METHODOLOGY

To explore the relationship between students' performance, self-evaluation on a programming test, self-efficacy in computer science and calibration, we employed a correlation and a causal-comparative group research design. The students were asked to solve 6 programming tasks and after that to respond to two 11-point (0-10) Likert-scale questions about their self-evaluation on the test (How do you think you did on the test from 0 to 10?) and their self-efficacy in computer science (How do you think you get on with the computer science course from 0 to 10?). These two questions were adapted from the study of Román-González et al. [20] who used the same questions to examine the relationship between computational thinking and students' self-efficacy perceptions on their performance in computational thinking tasks and about computers in general.

3.1 Participants

To answer the research questions of the study, we considered a homogeneous purposive sample. We were specifically interested in finding students who had experience in computer programming in any programming language. In total, 123 students from 7 different UK schools participated in the study, from which 98 were males and 25 were females [1]. 117 students were in 11th grade and 6 students in 12th grade. The programming tasks were written in C# and Python; 19 students completed the tasks written in C# and 104 in Python.

3.2 Instruments and Data collection

To collect the data, we created 6 programming exercises that included multiple choice questions, fill in the gaps, parson puzzles, and open-ended questions. All the questions had a fixed way of marking (0 to 10), decided between the researchers, and, thus, only one of the researchers involved in marking them. The data was collected through an online survey tool. The researchers had created two versions of the test, one in Python and one in C#, and the links to the test were distributed to the teachers. The teachers then gave the link to their students who completed the tasks during the computer science class under the supervision of their teachers.

3.3 Limitations

The main limitation of the study concerns the small sample size, especially the size of female participants. It should be noted that the number of girls who attend A level computer science courses in England is limited. For example, last year, girls represented 10% (816) of the students in UK that sat A level computing exams [23]. This impacted our ability to recruit female participants for our study. Nevertheless, in our sample, girls represent 20% of the participants and therefore we consider it as a good representative percentage for our purposes. The second limitation of the study concerns with the self-efficacy question. For our research, and based on Román-González et al.'s study [20], we used an one-item question which was employed ad-hoc for the purposes of our study. Finally, the results presented here, could only be interpreted having in mind that these students have chosen to attend computer science courses. Thus, the findings do not generalise to other populations.

4 RESULTS

4.1 Gender differences

Table 1 shows the descriptive statistics by gender and Table 2 presents the correlations between the examined variables. For the girls, the Pearson r correlation was used as the data were normally distributed while for boys the Spearman's rho. For boys, there is a strong and significant correlation between their performance and self-evaluation in the test (r_s=.741, p=.000, r^2=.549), a strong and significant correlation between self-evaluation of their test performance and efficacy in computer science (r_s=.651, p=.000, r^2=.423), and a moderate significant correlation between their performance and self-efficacy in computer science (r_s=.499, p=.000, r^2=.249). For

[1] Four of the seven schools had girls attending A level computer science

Table 2: Correlations by Gender

				Performance	Self-Evaluation	Efficacy in CS	Bias	Accuracy
Pearson's	Female	Performance	Correlation Coefficient	1	.319	.340	-.839**	-.229
			Sig. (2-tailed)	.	.120	.097	.000	.271
			N	25	25	25	25	25
		Self-Evaluation	Correlation Coefficient	.319	1	.623**	.249	.097
			Sig. (2 tailed)	.120	.	.001	.230	.646
			N	25	25	25	25	25
Spearman's rho	Male	Performance	Correlation Coefficient	1.000	.741**	.499**	-.483**	.058
			Sig. (2 tailed)	.	.000	.000	.000	.572
			N	98	98	98	98	98
		Self-Evaluation	Correlation	.741**	1.000	.651**	.172	.040
			Sig. (2 tailed)	.000	.	.000	.090	.694
			N	98	98	98	98	98

**Correlation is significant at the 0.01 level (2-tailed). Further correlations that are not significant or are already depicted in the table are not shown

the girls, there is only one significant and strong correlation between self-evaluation on test and self-efficacy in computer science (r=.623, p=.000, r^2=.388).

With the calibration correlations, there is a strong, negative and significant correlation between girls' performance and calibration bias (r=.-839, p=.000, r^2=.70) which is less intense for the boys (r_s=.-483, p=.000, r^2=.233).

To examine if boys or girls are better calibrated in computer programming and whether boys or girls are more efficacious in computer science we used the Mann-Whitney non-parametric test since boys' data were not normally distributed. The null hypotheses that we tested are listed below:

- H01: There is not a significant difference in students' performance between girls and boys.
- H02: There is not a significant difference in students' self-evaluation of their test performance between girls and boys.
- H03: There is not a significant difference in students' self-efficacy in computer science between girls and boys.
- H04: There is not a significant difference in students' calibration bias between girls and boys.
- H05: There is not a significant difference in students' calibration accuracy between girls and boys.

The results presented in Table 3 indicate that statistically significant differences are evident between girls' and boys' scores in the self-evaluation (U =793.0, z = -2.738, p = .006, d=0.24), the self-efficacy in computer science (U = 739.0, z = -3.107, p = .002, d=0.28) and the calibration accuracy (U = 916.5, z = -2.015, p = .044, d=0.18). Therefore, the null hypotheses H02, H03 and H05 could be rejected.

5 DISCUSSION

The first research question of the study referred to whether boys are better calibrated than girls in computer programming. The research findings suggest that boys indeed are better calibrated than girls in computer programming since the results were statistically significant. This finding suggests that boys make more accurate estimations of their programming performance than do girls. If we also look at their bias score, we can support that girls tend to underestimate their performance in comparison with boys. Similar studies conducted in other fields showed that boys (grade 5) tend to overestimate their achievements in regards with girls [6] who generally, tend to underestimate their performance [11]. Looking at research in mathematics education, the results are controversial with some studies supporting that boys are better calibrated than girls and others supporting no differences (e.g. [10], [14]).

The second research question was whether girls or boys feel more efficacious in computer science. The results indicate that there are significant differences between girls' and boys' self-evaluation and self-efficacy in computer science with boys scoring higher than girls. These findings suggest that 11th grade boys feel significantly more efficacious in programming and computer science than 11th grade girls even though there is no statistical difference in their actual performance. Askar and Davenport (2009) [1] have also highlighted the differences between females' and males' self-efficacy in programming and their study's result showed a statistically significant difference between the self-efficacy scores of male and female students with the latter being lower than the former.

The third research question was about correlations between the examined variables. Interestingly, when we examine the correlations between efficacy in computer science and the other variables per gender, it is obvious that the self-efficacy in computer science is highly correlated with self-evaluation in programming both for boys and girls. This indicates that as both boys' and girls' performance perceptions in programming tasks increase so does their efficacy in computer science. This is an interesting relationship as it may depict that students generalise their perceptions of ability from one part of computer science to the whole subject. Another important correlation is the negative significant relationship between performance and bias. In fact, this correlation is much stronger for girls (r=-.830) than boys (r=-.479) indicating that higher performance is associated with lower calibration bias (possible underestimations) both for girls and boys but it is much stronger for girls. Finally, a positive association between performance and self-evaluation as well as between performance and self-efficacy in computer science is only evident in the boys' population. This needs further investigation as it may be caused by a mediator variable.

Table 3: Mann-Whitney Test

	Test Statistics[a]				
	Performance	Self-Evaluation	Efficacy in CS	Bias	Accuracy
Mann-Whitney U	1074.000	793.000	739.000	973.500	916.500
Wilcoxon W	1399.000	1118.000	1064.000	1298.000	1241.000
Z	-.957	-2.738	-3.107	-1.605	-2.015
Asymp.sig. (2 tailed)	.339	.006	.002	.109	.044

a. Grouping variable: gender

6 CONCLUSIONS

The current paper presents a correlation and causal-comparative group research study which examines gender differences and the relationship between 11th grade students' self-efficacy in computer science, self-evaluation, and calibration achievement. The results of the study suggest that boys are better calibrated in computer programming than girls and in addition feel significantly more efficacious in computer science than girls. The study indicates that girls tend to underestimate their performance which, in comparison with boys, was not correlated with their self-evaluations nor their self-efficacy in computer science.

The results highlight a difference between boys and girls which reflects the way they evaluate their capabilities and the confidence they have of their capacity to succeed in computer science courses and not their actual performance. This strongly suggests that imprecisions in self-judgments in programming may be a liability for 11th and 12th grade students, especially for girls, and may impact the perceptions of their abilities in computer science.

The results of the study suggest the need for further and more comprehensive investigation and analysis. From the perspectives of teachers, the findings highlight the importance of incorporating opportunities for enhancing students' self-efficacy in computer science courses and for self-evaluation in lessons, while monitoring them for any inconsistencies. From students' perspectives, this training will engage them with self-assessment opportunities that will help them evaluate their performance and capabilities more objectively, take control of their learning and become self-regulated learners.

Our next research goal is to further investigate factors (predictors) that influence students' perceived capabilities in computer science as well as to explore differences on the calibration ability and self-efficacy between low and high performing students, with the gender variable being a central part of this investigation.

REFERENCES

[1] Petek Askar and David Davenport. 2009. An investigation of factors related to self-efficacy for Java programming among engineering students. *TOJET: The Turkish Online Journal of Educational Technology* 8, 1 (2009).

[2] Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84, 2 (1977), 191.

[3] Albert Bandura. 1994. Self-efficacy. In *Encyclopedia of Human Behavior*, VS Ramachaudran (Ed.). Vol. 4. Academic Press, New York, 71–81.

[4] Piraye Bayman and Richard E Mayer. 1983. A diagnosis of beginning programmers' misconceptions of BASIC programming statements. *Commun. ACM* 26, 9 (1983), 677–679.

[5] Jenny R Billings and Jan Macvarish. 2010. Self-Efficacy: Addressing Behavioural Attitudes Towards Risky Behaviour-An International Literature Review. (2010).

[6] Monique Boekaerts and Jeroen S Rozendaal. 2010. Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction* 20, 5 (2010), 372–382.

[7] Jeffrey Bonar and Elliot Soloway. 1985. Preprogramming knowledge: A major source of misconceptions in novice programmers. *Human-Computer Interaction* 1, 2 (1985), 133–161.

[8] Michael Clancy. 2004. Misconceptions and attitudes that interfere with learning to program. *Computer science education research* (2004), 85–100.

[9] Holger Danielsiek, Wolfgang Paul, and Jan Vahrenhold. 2012. Detecting and understanding students' misconceptions related to algorithms and data structures. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education.* ACM, 21–26.

[10] Cynthia A Ewers and Nancy L Wood. 1993. Sex and ability differences in children's math self-efficacy and prediction accuracy. *Learning and Individual Differences* 5, 3 (1993), 259–267.

[11] Eleftheria N Gonida and Angeliki Leondari. 2011. Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research* 50, 4 (2011), 209–220.

[12] Maria Kallia and Sue Sentance. 2017. Computing Teachers' Perspectives on Threshold Concepts: Functions and Procedural Abstraction. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education.* ACM, 15–24.

[13] Geetha Kanaparan, Rowena Cullen, and David Mason. 2017. Effect of Self-efficacy and Emotional Engagement on Introductory Programming Students. (2017). https:www.acis2017.org/wp-content/uploads/2017/11/ACIS2017_paper_234_FULL.pdf

[14] Frank Pajares and Laura Graham. 1999. Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary educational psychology* 24, 2 (1999), 124–139.

[15] Frank Pajares and M David Miller. 1997. Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *The Journal of Experimental Education* 65, 3 (1997), 213–228.

[16] Sarantos Psycharis and Maria Kallia. 2017. The effects of computer programming on high school studentsâĂŹ reasoning skills and mathematical self-efficacy and problem solving. *Instructional Science* 45, 5 (2017), 583–602.

[17] Yizhou Qian and James Lehman. 2017. StudentsâĂŹ misconceptions and other difficulties in introductory programming: a literature review. *ACM Transactions on Computing Education (TOCE)* 18, 1 (2017), 1.

[18] Vennila Ramalingam and Susan Wiedenbeck. 1998. Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *Journal of Educational Computing Research* 19, 4 (1998), 367–381.

[19] Darshanand Ramdass and Barry J Zimmerman. 2008. Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of advanced academics* 20, 1 (2008), 18–41.

[20] Marcos Román-González, Juan-Carlos Pérez-González, Jesús Moreno-León, and Gregorio Robles. 2017. Extending the nomological network of computational thinking with non-cognitive factors. *Computers in Human Behavior* 80, 2018 (2017), 441e459.

[21] Richard Sheldrake, Tamjid Mujtaba, and Michael J Reiss. 2014. Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16. *International Journal of Educational Research* 64 (2014), 49–61.

[22] Teemu Sirkiä and Juha Sorva. 2012. Exploring programming misconceptions: an analysis of student mistakes in visual program simulation exercises. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research.* ACM, 19–28.

[23] Royal Society. 2017. After the Reboot: Computing Education in UK schools. https://royalsociety.org/topics-policy/projects/computing-education/

[24] Nancy J. Stone. 2000. Exploring the Relationship between Calibration and Self-Regulated Learning. *Educational Psychology Review* 12, 4 (01 Dec 2000), 437–475.

[25] Susan Wiedenbeck, Deborah Labelle, and Vennila NR Kain. 2004. Factors affecting course outcomes in introductory programming. In *16th Annual Workshop of the Psychology of Programming Interest Group.* 97–109.

[26] Barry J Zimmerman. 1998. Academic studing and the development of personal skill: A self-regulatory perspective. *Educational psychologist* 33, 2-3 (1998), 73–86.