Programming Tutors, Practiced Concepts, and Demographics

Amruth N. Kumar
Computer Science
Ramapo College of New Jersey
Mahwah, NJ, USA
amruth@ramapo.edu

Lisa C. Kaczmarczyk
Independent Research Consultant
San Diego, CA
lisakacz@gmail.com

***Abstract*— A study was conducted to find out who needed online problem-solving tutors and who benefited from using them. In particular, the study focused on whether there were any significant differences between male and female students and between traditionally represented and under-represented racial groups. Data collected by two Computer Science tutors over multiple semesters was analyzed. The only significant differences found between sexes and racial groups were when female students practiced significantly more concepts because they had solved significantly fewer problems during pre-test, or when they demonstrated greater pre-post increase in score because they had scored significantly less on the pre-test. In both the cases, the tutors helped female students overcome differences in prior preparation vis-a-vis male students. No difference was found between the sexes or racial groups on the number of practice problems solved per practiced concept. Finally, students needed and benefited from the tutors in the same proportion, regardless of sex or racial group.**

Keywords- problem-solving; online learning; demographics; software tutor; computer programming

I. Introduction

Solving problems is an integral part of learning STEM (Science, Technology, Engineering and Mathematics). Traditionally, problems have been included at the end of each chapter in STEM text books. Instructors have given out problem-solving assignments, and used problems for assessment during tests.

Problem-solving activity is increasingly going online. The benefits of online problem-solving are several: students get their answers graded instantly; they can receive feedback on how to solve problems correctly; and they can solve problems as often as they please, at their own pace, and on their own time.

Some online problem-solving systems that are available for computer programming include CodeLab (turingscraft.com), myprogramminglab (myprogramminglab.com), codingbat (codingbat.com), CodeMate (pearsonhighered.com/mycodemate/), and CloudCoder (cloudcoder.org) - they provide practice writing code snippets focused on specific programming constructs. In contrast, we have been developing software tutors to help students learn programming concepts by solving code-tracing problems, called problets (problets.org)

Most of the studies of the effectiveness of learning Computer Science concepts from online tutors originate from research on Intelligent Tutoring Systems (ITS). For example, in one of the early studies, ITS researchers documented one standard deviation improvement in learning LISP using a problem-solving tutor [1]. However, many of the studies in Intelligent Tutoring Systems have been conducted *in-vitro* rather than *in-vivo*. More *in-vivo* evaluations of online tutors in Computer Science would better justify the wide-spread use of such tutors in Computer Science and the concomitant costs associated with developing the tutors. In one of our early *in-vivo* studies, we found that online tutors indeed helped students learn [2] and students who used the tutor for practice learned better than those who used a printed workbook [3].

Women and minorities are significantly under-represented in Computer Science, e.g., in 2011-2012, the latest year for which Taulbee survey figures are available (cra.org), women represented only 12.9% of Bachelor's degree graduates; Caucasians represented 64%, Asians 16.3%, and non-Asian minorities accounted for less than 20% of the graduates. Given the rising popularity of online tutors in Computer Science, it is appropriate and timely to study the differential effects, if any, of using online Computer Science tutors on women and minorities. In one of our earlier studies, we found that female students did learn using problets, and that there was no statistically significant difference between the improvement in the learning of female and male students using problets [4]. In another study, we found that using problets improved the self-confidence of female students to be on par with that of male students when female students started with lower prior self-confidence than male students [5]. We also found that female students assessed problets more positively than male students [6], and whenever there was a statistically significant difference between racial groups, under-represented racial groups assessed problets more favorably than positively stereotyped racial groups [7].

Two interesting questions about the differential effects of using online problem-solving tutors in Computer Science are: who *needs* to use them, and who *benefits* from using them. In particular, whether there is any significant difference between male and female students, and between traditionally represented and under-represented racial groups. In order to answer these questions, we conducted a study using problets, spanning multiple semesters. In this paper, we will present our findings, after first discussing the research methodology, and data analysis.

In the study, participants were asked to identify their sex (biological notion of male/female) rather than their gender (social/cultural notion of man/woman) [8]. Therefore, the analysis will be presented in terms of sex rather than gender. When comparing racial groups, two groups were considered: traditionally represented racial groups versus under-represented racial groups. Asians are positively stereotyped in quantitative domains such as Computer Science (e.g., [9]). Therefore, Asians were combined with Caucasians to form the traditionally represented racial group (referred to as "majority" in another study [10]). The other racial groups, viz., Black/African American, Hispanic/Latino, Native American, Native Hawaiian/Pacific Islander and Other designations were combined to form the under-represented racial group.

Given the significant under-representation of women and minorities in Computer Science, the results of this study will be of interest to educators who use problem-solving software tutors in their Computer Science courses. If women and/or minorities are found to need and/or benefit from using problem-solving tutors, the results would provide justification for greater development and use of such tutors in Computer Science.

The online problem-solving software used for this study, called problets, deals with introductory computer programming concepts. Each problet presents problems on a specific programming construct and has the student solve them. When the student submits the answer, the problet grades it, and provides feedback, including step-by-step explanation of the correct answer, which has been shown to help students learn [2].

Each problet is configured to administer pre-test-practice-post-test protocol as follows:

- **Pretest** – During the pretest, the problet presents one problem per concept. If a student solves a problem correctly, the student is given credit for the corresponding concept. No feedback is provided to the student, and no more problems on the concept are presented to the student. On the other hand, if the student solves a problem incorrectly, feedback is presented to the student immediately after the student submits his/her solution to the problem. Additional problems are presented on the concept during the subsequent stages.
- **Adaptive practice** – During this stage, additional problems are presented to the student on only the concepts on which the student made mistakes when solving problems during the pre-test. For each such concept, the student is presented multiple problems until the student masters the concept, i.e., solves at least 60% of the problems correctly. On each problem, the student receives feedback explaining the correct answer step by step.
- **Post-test** - During this stage, the student is presented test problems on the concepts that the student mastered during adaptive practice.
- **Demographics** - Students are provided the option to identify their demographic information, including sex and race. Demographic information is solicited after the pre-test-practice-post-test protocol to avoid the effects of stereotype threat [11,12,13].

The pre-test-practice-post-test protocol was limited to 30 minutes and was administered back-to-back, entirely over the web.

Given this set up, the following were the types of experiences students had using the problet:

- The student solved all the pre-test problems correctly. So, the student was presented no practice or post-test problems.
- The student solved at least one of the pre-test problems incorrectly. However, the student ran out of time solving practice problems and never solved any post-test problems.
- The student solved at least one pre-test problem incorrectly, solved sufficient practice problems to demonstrate mastery of at least one concept and solved a post-test problem on each such concept. For these students, **the number of practiced concepts** was the number of concepts on which the student solved pre-test problems incorrectly, solved sufficient practice problems to demonstrate mastery and solved the post-test problem.

For analysis purposes, we grouped students into four categories, and used the following measures to compare them:

- **Students who did not need to use the tutor** – these were the students who scored 100% on the pre-test. So, the tutors did not present any practice or post-test problems to these students. For these students, the time spent per pre-test problem was compared - those who solved problems faster had greater self-efficacy [14] when they started using the tutor.
- **Students who needed the tutor, but did not get to practice with it** – these students scored less than 100% on the pre-test. But, they spent all 30 minutes solving pre-test problems. So, they ran out of time and did not get to solve any practice or post-test problems. For these students, the score per pre-test problem was compared - those who scored more per problem were better prepared when they started using the tutor.
- **Students who solved practice problems, but not enough to practice any concept** – these students scored less than 100% on the pre-test, so they needed to use the tutor. They solved practice problems, but not in sufficient numbers to demonstrate mastery, i.e., score 60% or more on any one concept. So, the tutor did not present any post-test problems to these students. For these students, the number of practice problems was compared - the more practice problems a student solved without mastering any concept, the less the tutor helped the student learn.

- **Students who practiced one or more concepts** – these students solved less than 100% on the pre-test, solved sufficient number of problems during practice to score at least 60% on one or more concepts, and went on to solve post-test problems on those concepts. For these students, the number of concepts practiced and the pre-post change in score on the practiced concepts were compared – the more the number of practiced concepts and the more the increase in score from pre-test to post-test, the greater the learning. The number of practice problems solved per practiced concept was also compared – the more the problems solved, the more the effort needed by the student to learn with the tutor.

For this study, two tutors were used, both on introductory concepts:

- Arithmetic expressions tutor, designed to help students learn to evaluate arithmetic expressions as used in computer programming. Each problem contains an arithmetic expression that includes one or more operators. The student is asked to evaluate the expression one operator at a time, according to the rules of precedence and associativity.
- Selection tutor, designed to help students learn to trace programs containing if and if-else statements. Each problem contains a computer program that includes selection statements. The student is asked to predict all the outputs of the program, one output at a time, along with the line number of the code that generates the output.

These two tutors were selected because they promised large sample sizes – they are both typically used early in the semester and are used by larger numbers of students than tutors on more advanced topics such as functions and arrays. At the same time, the two tutors were not duplicative – they have different user interfaces and require different problem-solving skills.

The tutors were used by students enrolled in introductory programming courses. Students from multiple institutions used the tutors each semester – an average of 12 institutions for arithmetic tutor and 15 institutions for selection tutor. All the students used the tutors over the web. The entire protocol was administered back-to-back – any student considered in this study went through all the stages of the protocol within 30 minutes, with no breaks in between. So, any change in score from pre-test to post-test is attributable to the use of the tutor and not any extraneous factors.

II. ARITHMETIC EXPRESSION TUTOR RESULTS

During pre-test, arithmetic expressions tutor presents 16 problems. For analysis purposes, only those students who had solved at least 10 problems during the pre-test were considered. When a student had used the tutor multiple times, data from only the first time the student had solved at least 10 pre-test problems was considered.

Data was collected over five semesters – fall 2010 through fall 2012. All the students used the same version of the tutor during these five semesters. After eliminating duplicates and students who had not solved sufficient problems during the pre-test, 1236 students remained in the study:

- 242 students "did not need to use the tutor";
- 206 students "needed the tutor, but did not get to practice with it";
- 273 students "solved practice problems, but not enough to practice any concept"; and
- 515 students who "practiced one or more concepts".

Since providing sex and race information was optional, the value of N may vary in the following analyses.

A. Students who did not need to use the tutor

Univariate analysis of the time per pre-test problem yielded a significant main effect for sex [F(1,185) = 6.717, p = 0.01]: Female students spent statistically significantly more time to solve each pre-test problem than male students, as shown in Table I (In the table, means are listed with confidence intervals at 95% confidence level). Since both male and female students scored 100% on the pre-test, perceived self-efficacy [14] may be one possible explanation for why female students took more time – female self-efficacy is significantly lower than male self-efficacy [15]. Similar analysis did not yield a significant main effect for race [F(1,185) = 2.082, p = 0.151], i.e., the difference between the racial groups was not statistically significant.

TABLE I. TIME PER PRE-TEST PROBLEM ON ARITHMETIC EXPRESSIONS TUTOR FOR THE GROUP THAT DID NOT NEED TO USE THE TUTOR

	Male	**Female**
N	142	44
Mean	45.48 ± 3.66	56.83 ± 7.83
	Caucasians+Asians	**Under-represented**
N	135	51
Mean	47.99 ± 3.90	54.31 ± 7.71

B. Students who needed the tutor

Among the students who needed the tutor, but did not get to practice with it, there was no significant difference in the pre-test score per problem between the two sexes [F(1,145) = 0.115, p = 0.735] or racial groups [F(1,145) = 0.959, p = 0.329]. In other words, both the groups in each comparison were equally prepared before using the tutor.

Among the students who solved practice problems, but not enough to practice any concept, no significant difference was found in the number of practice problems between the sexes [F(1,213) = 0.3, p = 0.584] or racial groups [F(1,213) = 0.128, p = 0.72].

C. Students who practiced one or more concepts

Analysis of the number of concepts practiced yielded significant difference between the sexes [F(1,392) = 5.52, p = 0.019], but not between the racial groups [F(1,392) = 0.357, p = 0.551], as shown in Table II. Female students practiced significantly more concepts than male students, possibly because they solved statistically significantly fewer problems on the pre-test than male students [F(1,392) = 5.931, p =

0.015], and therefore, demonstrated proficiency on fewer concepts during pre-test.

TABLE II. CONCEPTS PRACTICED ON ARITHMETIC EXPRESSIONS TUTOR

	Male	Female
N	300	93
Mean	1.588 ± .14	1.933 ± .253
	Caucasians+Asians	**Under-represented**
N	289	104
Mean	1.716 ± .147	1.804 ± .249

No statistically significant difference was found on the pre-post change in score on practiced concepts, whether between the sexes [F(1,392) = 0.45, p = 0.503] or racial groups [F(1,392) = 0.031, p = 0.861]. Similarly, on the number of practice problems solved per practiced concept, no significant difference was found between the sexes [F(1,392) = 1.865, p = 0.173] or racial groups [F(1,392) = 0.101, p = 0.751].

In summary, among the students who did not need to use the tutor, female students may have exhibited lower self-efficacy. Among the students who practiced one or more concepts, female students practiced significantly more concepts than male students, possibly because they were less prepared than male students when they started using the tutor. No other difference was found between the sexes or racial groups on the number of practice problems solved per practiced concept, or the pre-post change in score on the practiced concepts.

III. SELECTION TUTOR RESULTS

During pre-test, selection tutor presents 12 problems in C++ and 9 problems in Java/C#. For analysis purposes, only those students who had solved at least 8 problems during the pre-test were considered. When a student had used the tutor multiple times, data from only the first time the student had solved at least 8 pre-test problems was considered.

Data was collected over four semesters – fall 2010 through spring 2012. All the students used the same version of the tutor during these four semesters. After eliminating duplicates and students who had not solved sufficient problems during the pre-test, 893 students remained in the study:

- 370 students "did not need to use the tutor";
- 91 students "needed the tutor, but did not get to practice with it" or "solved practice problems, but not enough to practice any concept". No analysis was done on these groups because of their small sample size.
- 432 students "practiced one or more concepts".

Since providing sex and race information was optional, the value of N may vary in the following analyses.

A. Students who did not need to use the tutor

Univariate analysis of the time per pre-test problem did not yield a significant main effect for sex [F(1,251) = 0.474, p = 0.492], but did so for racial groups [F(1,251) = 6.908, p = 0.009], as shown in Table III. Under-represented students spent statistically significantly more time to solve each pre-test problem than Caucasians+Asians. Again, since both racial groups scored 100% on the pre-test, perceived self-efficacy [14] may be one possible explanation for why under-represented students took more time – under-represented racial groups were found to have significantly lower prior self-confidence in another study [7].

TABLE III. TIME PER PRE-TEST PROBLEM ON SELECTION TUTOR FOR THE GROUP THAT DID NOT NEED TO USE THE TUTOR

	Male	Female
N	194	58
Mean	52.85 ± 4.60	56.03 ± 7.83
	Caucasians+Asians	**Under-represented**
N	203	49
Mean	48.38 ± 4.20	60.50 ± 8.05

B. Students who practiced one or more concepts

There was no significant difference in the number of concepts practiced by the two sexes [F(1,323) = 1.608, p = 0.206] or racial groups [F(1,323) = 1.321, p = 0.251], as shown in Table IV.

TABLE IV. CONCEPTS PRACTICED ON SELECTION TUTOR

	Male	Female
N	253	71
Mean	1.686 ± .152	1.896 ± .289
	Caucasians+Asians	**Under-represented**
N	240	84
Mean	1.696 ± .165	1.886 ± .281

The pre-post change in learning was significantly different between the sexes [F(1,323) = 4.661, p = 0.032], but not the racial groups [F(1,323) = 0.002, p = 0.962]. Female students had a greater improvement from pre-test to post-test on the practiced concepts as shown in Table V, likely because they scored statistically significantly less on the pre-test than male students [F(1,323) = 4.252, p = 0.04]. The post-test scores were not significantly different. The interaction between the sexes and racial groups was also significant [F(1,323) = 5.223, p = 0.023] as shown in Table V, i.e., pre-post change was less for under-represented males and more for under-represented females than Caucasians+Asians.

TABLE V. PRE-POST IMPROVEMENT ON PRACTICED CONCEPTS ON SELECTION TUTOR

	Caucasians+Asians	Under-represented	Total
Male	.887 ± .033 (N=187)	.809 ± .054 (N=66)	.848 ± .032
Female	.882 ± ..060 (N=53)	.963 ± .104 (N=18)	.923 ± .061
Total	.884 ± .034	.886 ± .059	

On the number of practice problems solved per practiced concept, no significant difference was found between the sexes

[F(1,323) = 0.328, p = 0.567] or the racial groups [F(1,323) = 2.744, p = 0.099].

In summary, among the students who did not need to use the tutor, students in under-represented racial groups may have exhibited lower self-efficacy. Among the students who practiced one or more concepts, the change in score from pre-test to post-test was significantly more for female students than male students, and this made up for the fact that female students scored significantly less on the pre-test. No other difference was found between the sexes or racial groups in the number of concepts practiced, or the number of practice problems solved per practiced concept.

IV. DISTRIBUTION OF STUDENTS AMONG CATEGORIES

As mentioned earlier, students were grouped into four categories:

1. Students who did not need to use the tutor
2. Students who needed the tutor, but did not get to practice with it
3. Students who solved practice problems, but not enough to practice any concept
4. Students who practiced one or more concepts using the tutor.

We analyzed the distribution of students among these four categories by sex and race. Univariate analysis of the category by sex and racial group for arithmetic expressions tutor yielded no significant main effect for sex [F(1,938) = 0.185, p = 0.667] (Figure 1) or racial group [F(1,938) = 0.28, p = 0.597] (Figure 2). In other words, no statistical difference was observed between male and female students and between traditional and under-represented racial groups in their distribution among the four categories, including the proportions that needed the tutor and that benefited from the tutor, i.e., practiced one or more concepts using the tutor.

Similar analysis of the data collected by selection tutor again yielded no significant main effect for sex [F(1,630) = 0.373, p = 0.542] (Figure 3) or racial group [F(1,630) = 1.956, p = 0.162] (Figure 4). *So, students needed and benefited from the tutors in the same proportion, regardless of sex or racial group.*

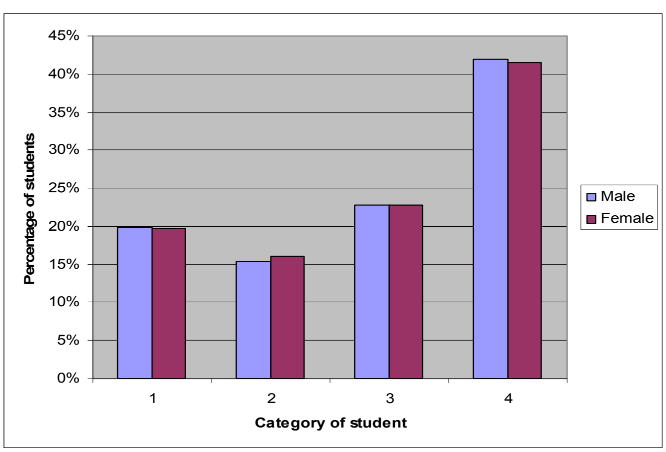

Figure 1. Arithmetic Expressions Tutor: Percentages of male and female students distributed among the four categories

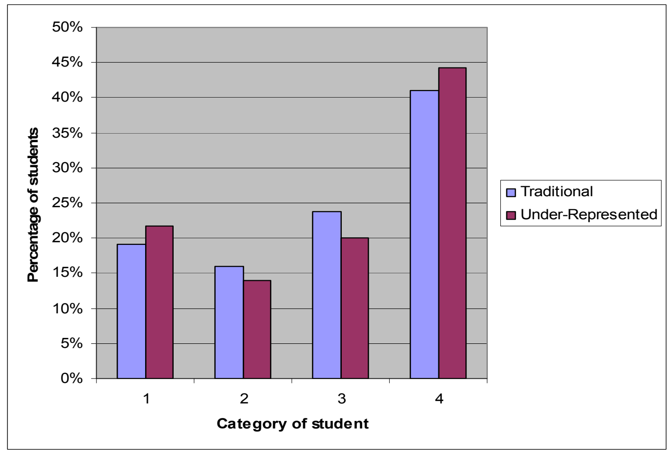

Figure 2. Arithmetic Expressions Tutor: Percentages of traditional and under-represented racial groups distributed among the four categories

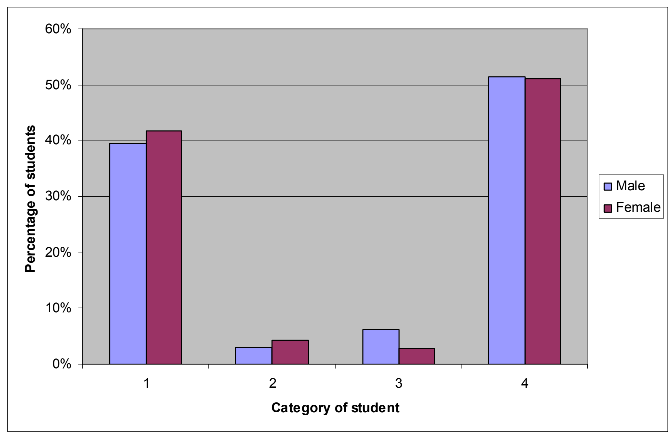

Figure 3. Selection Tutor: Percentages of male and female students distributed among the four categories

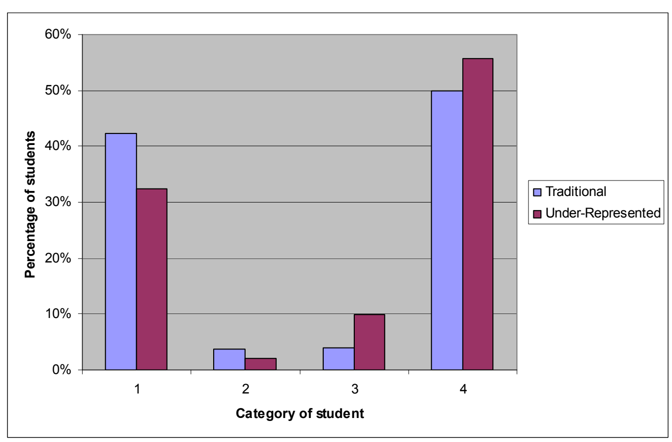

Figure 4. Selection Tutor: Percentages of traditional and under-represented racial groups distributed among the four categories

V. Discussion

Two tutors were used in this study. The following summarizes the analysis of the data collected by these two tutors:

- Students who did not need to use the tutor – Among these students, whenever any difference was found, either female students or students in under-represented racial groups spent longer solving problems, possibly because of lower perceived self-efficacy [14].
- Students who practiced one or more concepts – Among these students, whenever any difference was found, it was female students who either practiced significantly more concepts when they had solved significantly fewer problems during pre-test, or demonstrated greater pre-post increase in score when they had scored significantly less on the pre-test. These results bode well – one study found that the most significant cognitive factor predicting attrition in Science was low grades earned in introductory courses [16]. Attrition of female students is of particular concern in Computer Science, a phenomenon popularly referred to as the shrinking pipeline [17]. Given that the tutors help female students improve their scores from pre-test to post-test, they might help reduce the attrition of female students in introductory Computer Science courses.

No other differences were found between the sexes or racial groups. In particular:

- No difference was found on the number of practice problems solved by the sexes and racial groups per practiced concept. So, the effort needed to practice with problets was the same regardless of sex- or racial group.
- Students needed and benefited from the tutors in the same proportion, regardless of sex or racial group.

Both these results are reassuring for the use of problem-solving tutors in Computer Science, given the significant under-representation of women and minorities in Computer Science and the push to broaden participation in the discipline.

Socially relevant themes, and teamwork and collaboration, are increasingly being emphasized along with hands-on learning as effective pedagogical approaches for engaging women in computing.. [18]. That female students learned just as well as male students by using problets seems to suggest that hands-on learning is effective on its own, even without the incorporation of socially relevant themes or teamwork and collaboration, neither of which is part of problets.

The tutors used in this study dealt with basic topics in programming. We plan to conduct additional studies with tutors on more advanced programming topics to see whether these results can be replicated or different patterns will emerge.

Acknowledgment

Partial support for this work was provided by the National Science Foundation under grant DUE-0817187.

References

[1] Anderson J.R., Corbett A.T., Koedinger K.R. and Pelletier R., "Cognitive Tutors: Lessons Learned", The Journal of the Learning Sciences, Vol No 4(2), 1995, Lawrence Erlbaum Associates, Inc., pp 167-207.

[2] Kumar, A.N., Explanation of step-by-step execution as feedback for problems on program analysis, and its generation in model-based problem-solving tutors, Technology, Instruction, Cognition and Learning (TICL) Journal, Vol 4(1), 2006.

[3] Kumar, A.N. Results from the Evaluation of the Effectiveness of an Online Tutor on Expression Evaluation", *Proc.* of The 36th SIGCSE Technical Symposium on Computer Science Education, SIGCSE 2005, St. Louis, MO, 2/23-27/2005, 216-220.

[4] Kumar, A.N. Software Tutors Help Female Students Learn Programming Concepts Just as Well as Male Students, E-LEARN 2007, Quebec City, Canada, 10/15-19/2007 , 6840-6845.

[5] Kumar, A.N. The Effect of Using Problem-Solving Software Tutors on the Self-Confidence of Female Students. *Proc.* The 39th SIGCSE Technical Symposium on Computer Science Education, SIGCSE 2008, Portland, OR, 3/12-15/2008, 523-527.

[6] Kumar, A.N. "Female Students Assess Software Tutors More Positively Than Male Students". *Proc.* of Frontiers in Education Conference (FIE 2008)},Saratoga Springs, NY, 10/22-25/2008, Session S4F.

[7] Kumar, A.N. "Results from Repeated Evaluation of an Online Tutor on Introductory Computer Science'". *Proc* of Frontiers in Education Conference (FIE 2011)}, Rapid City, SD, 10/12-15/2011.

[8] Sears, J. (1999). Teaching queerly: Some elementary propositions. In W. J. Letts & J. Sears (Eds.), Queering elementary education: Advancing the dialogue about sexualities and Schooling (97-110). Lanham, MD: Rowman & Littlefield, Inc.

[9] Kao, G., "Asian Americans as model minorities? A look at their academic performance", American Journal of Education, 103, 1995, 121-159.

[10] Katz, S., Aronis, J., Allbritton, D., Wilson, C., and Soffa, M.L. Gender and Race in Predicting Achievement in Computer Science. *IEEE Technology and Society Magazine.*Fall 2003. 21 – 27.

[11] Steele, C.M. and Aronson, J., "Stereotype threat and the intellectual test performance of African Americans", Journal of Personality and Social Psychology, 69, 1995, 797-811.

[12] Kumar, A.N. A Study of Stereotype Threat in Computer Science. Proc. Innovation and Technology in Computer Science Education (ITiCSE 2012), Haifa, Israel, 7/3-5/2012, 273-278.

[13] Kumar, A.N. The Effect of Soliciting Demographic Data on the Performance of Students on Online Tests. Proc. Frontiers in Education Conference (FIE 2010), Washington, D.C., 10/27-30/2010.

[14] Bandura, A. Self-Efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review. 8(2). 1977. 191-215.

[15] Beckwith, L., Burnett, M., Grigoreanu, V. and Wiedenbeck, S. Gender HCI: What About the Software? IEEE Computer. November 2006. 97-101.

[16] Strenta, A.C., Elliott, R., Adair, R., Matier, M., and Scott, J. Choosing and Leaving Science in Highly Selective Institutions. *Research in Higher Education.* 35(5). 1994. 513—547.

[17] Camp, T. The Incredible Shrinking Pipeline. *Communications of the ACM.* 40(10). Oct 1997. 103-110.

[18] Margolis, J and Fisher, A. Unlocking the Bluc House. MIT Press, 2002.