The Effect of Using Problem-Solving Software Tutors on the Self-Confidence of Female Students

Amruth N. Kumar
Ramapo College of New Jersey
505 Ramapo Valley Road
Mahwah, NJ 07430

amruth@ramapo.edu

ABSTRACT

We examined whether using problem-solving software tutors in *Computer Science I* can help improve the self-confidence of female students. We analyzed the data collected by five software tutors in spring 2006. We found that 1) the self-confidence of female *Computer Science I* students before using the software tutors was in many cases lower than that of male students, as has been stated in prior literature; 2) Using problem-solving software tutors 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. Since researchers have suggested that self-confidence is one of the factors contributing to the shrinking pipeline, problem-solving software tutors can be used to improve the retention of female students in Computer Science.

Categories and Subject Descriptors

K.3.1. [Computing Milieux]: Computer-Assisted Instruction

General Terms

Experimentation

Keywords

Problem-solving software tutors, Programming, Shrinking pipeline, Self-confidence, Learning

1. INTRODUCTION

Women are under-represented in Computer Science. Not only does the number of women in Computer Science drop from high-school through graduate school, but it has also been dropping at the undergraduate level over the years, a phenomenon referred to as the shrinking pipeline [6].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGCSE'08, March 12–15, 2008, Portland, Oregon, USA.
Copyright 2008 ACM 978-1-59593-947-0/08/0003...\$5.00.

Mathematics is an area where women have done well – they have received almost half of all undergraduate degrees in Mathematics. De Palma [9] argues that since the study of Mathematics relies on numerous practice problems, similar practice problems should be provided to improve the success of women in Computer Science. In this vein, we have been developing problem-solving software tutors for various topics in Computer Science. The software tutors are designed to help students learn programming concepts by solving problems. They are meant to be used as supplements to classroom instruction and complements to programming projects traditionally assigned in *Computer Science I*. Multiple evaluations have shown that these software tutors are effective at helping students learn programming concepts [14,15].

Self-confidence is one of the factors contributing to the shrinking pipeline [11]. Female students have lower self-confidence than male students in computer-related abilities [4,10]. Women are less confident than men in their ability to achieve their educational goals in computing at the undergraduate [5] and the graduate level [8]. Girls consistently rate their ability at a lower level than boys despite evidence to the contrary [7]. Women have a tendency to enter computing classes with considerably less confidence than men [12]. This can be particularly attributed to self-efficacy [2] - "a person's judgment about one's ability to carry out specific actions to achieve a goal." [3] found that female self-efficacy was significantly lower than male self-efficacy.

One might expect that solving practice problems can positively affect self-efficacy. Performance / accomplishment and freedom from anxiety are two of the components of self-confidence [1]. Solving practice problems might help reduce the anxiety of female students and instill a sense of accomplishment in them, thereby improving their self-confidence. This begs the question - does using problem-solving *software* tutors help improve the self-confidence of students, particularly female students?

In order to answer this question, we evaluated the data collected from five software tutors in spring 2006. In this paper, we will describe the data collection process and the results of analyzing the data. Participants of our studies were asked to identify their sex (biological notion of male/female) rather than their gender (social/cultural notion of man/woman) [17]. Therefore, our analysis will be in terms of sex rather than gender.

2. THE METHOD

In spring 2006, we evaluated five software tutors in *Computer Science I* – on arithmetic expressions, relational expressions, selection statements, while loops and for loops. Several faculty from different schools used the software tutors in their classes. The tutors were set up as Java applets accessible any time, anywhere over the Web. Teachers assigned the use of each software tutor as and when they covered its topic in class. Students using the tutors went through the following stages, all administered online:

1. Registration – students entered their name, status, sex and other demographic information;
2. Pre-Survey – students rated their knowledge about the various concepts covered by the software tutor on a 5-point Likert scale (1=Very Well, 2=Well, 3=Average, 4=Not well, 5=Not at all). Typical questions included “How well do you know the remainder operator?”, “How well do you know operator precedence concepts?”, “How well do you know if-else statements?”, “How well do you know nested 'if' statements?”, “How well do you know for loops?” and “How well do you know nested while loops?”
3. Problem-solving session - the software tutor presented problems to the students, had them solve the problems, and provided them feedback from which they could learn [14]. During the problem-solving session, the software tutor used the pre-test-treatment-post-test protocol: pre-test to evaluate the prior knowledge of students, treatment to help them learn the concepts they didn't already know, and post-test to evaluate the posterior knowledge of students. The entire problem-solving session lasted 30-35 minutes.
4. Post-Survey – the software tutor asked the students to again rate their knowledge about the concepts covered by the tutor on a 5-point Likert scale. The same questions were asked on the post-survey as on the pre-survey.

We used the student responses on the pre and post-survey as measures of their self-confidence in the topic before and after using the software tutor. Since the pre-survey and post-survey were conducted immediately before and after the problem-solving session, and students had little time in between to engage in any other activity that may have influenced their self-confidence, we can ascribe any difference in pre and post-survey scores to the intervening problem-solving session.

3. THE RESULTS AND DISCUSSION

We will present two aggregate analyses: of the average response per software tutor, followed by the average response per student. We will follow up with post-hoc analyses of each software tutor, and within each tutor, of each survey question.

Aggregate analysis of the average response per tutor: Recall that students answered 4-5 questions on the pre- and post-survey of each software tutor. We calculated the average response of each student on all the questions of each software tutor. We tabulated these average responses for all the tutors, and separated them by sex. We then calculated the average and standard deviation of male and female responses on the pre-survey, post-survey and the pre-post change. We used paired-sample t-test 2-tailed p -value to calculate the statistical significance of the

difference between pre-survey and post-survey responses. Finally, we used independent t-test 2-tailed p -value to calculate the statistical significance of the difference between male and female responses. We have listed these figures in Table 1. Due to the design of the Likert scale we used, higher scores in the table correspond to lower self-confidence.

Responses	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 266 responses)				
Ave	2.693	2.151	0.542	0.000
Std-Dev	0.901	0.847	0.796	
Female (N = 110 responses)				
Ave	3.004	2.211	0.793	0.000
Std-Dev	0.935	0.793	0.860	
Male Vs Female <i>p</i>	0.003	0.512	0.009	

Table 1: Average response on all the questions of a tutor – Female self-confidence significantly lower than male self-confidence on the pre-survey, but not the post-survey

Note that the pre-post change in response is statistically significant for both males and females – using problem-solving software helps improve the self-confidence of both male and female students. Female responses are statistically significantly lower than male responses on the pre-survey, as was to be expected from a review of the literature [4,7,10,12]. There was no statistically significant difference between male and female responses on the post-survey – *using problem-solving software helped improve female responses to be on par with the male responses.*

Aggregate analysis of the average response per student: For the analysis in Table 1, we treated the average response of each student on each software tutor as a separate item of data. So, if a student used four different tutors, we considered four items of data from that student in the analysis. For the next analysis, for each student we calculated a single item of data corresponding to the average of the student's responses for all the software tutors, bearing in mind that not all the students had used all the tutors. We wanted to find out if there were differences between male and female students, this time treating student responses as indicative of who the respondents were, more than which software tutor(s) they had used.

We conducted a mixed-factor 2 X 2 ANOVA with pre-post as within-subjects and sex as between-subjects factors. There was a significant main effect for pre versus post [$F(1,161) = 94.411$, $p = 0.000$], and for sex [$F(1,161) = 4.445$, $p = 0.037$]. The interaction between pre-post and sex was not significant [$F(1,161) = 1.628$, $p = 0.204$]. The results are summarized in Table 2. The results are similar to those in Table 1, except that the difference in the pre-post change between male and female responses is not statistically significant. The results presented in tables 1 and 2 warrant post-hoc analysis of the data of each tutor, which we will present next.

Students	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 125 students)				
Ave	2.740	2.159	0.598	0.000
Std-Dev	0.908	0.757	0.727	
Female (N = 38 students)				
Ave	3.114	2.357	0.757	0.000
Std-Dev	0.829	0.715	0.757	
Male Vs Female <i>p</i>	0.020	0.144	0.256	

Table 2: Average of the responses on all the tutors for each student – Female self-confidence again significantly lower than male self-confidence on the pre-survey, but not the post-survey

Post-hoc analysis of the average response for each tutor: We did a post-hoc analysis of the data by individual software tutors to see if the difference between male and female students varied among the tutors. We found no difference between male and female students on the software tutors on arithmetic expressions and `for` loops, the tutors that the students had used first and last in the semester. On the other three software tutors, we found the same pattern: statistically significant or nearly significant difference between male and female students on the pre-survey, but no significant difference on the post-survey. We have summarized the results in Tables 3, 4 and 5 respectively. These results warrant post-hoc analysis of the individual survey questions presented by these three software tutors. We will present this analysis next.

Relational Tutor	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 51 students)				
Ave	2.750	1.755	0.995	0.000
Std-Dev	0.911	0.757	0.867	
Female (N = 24 students)				
Ave	3.177	1.833	1.344	0.000
Std-Dev	0.948	0.751	0.890	
Male Vs Female <i>p</i>	0.072	0.676	0.118	

Table 3: Relational tutor – Female self-confidence significantly lower than male self-confidence on pre-survey, but not post-survey

Selection Tutor	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 67 students)				
Ave	2.693	2.176	0.516	0.000
Std-Dev	0.787	0.777	0.738	
Female (N = 20 students)				
Ave	3.080	2.270	0.810	0.000
Std-Dev	0.981	0.890	0.835	
Male Vs Female <i>p</i>	0.117	0.674	0.168	

Table 4: Selection tutor - Female self-confidence significantly lower than male self-confidence on pre-survey, but not post-survey

while Tutor	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 38 students)				
Ave	2.753	2.332	0.421	0.000
Std-Dev	0.746	0.823	0.681	
Female (N = 21 students)				
Ave	3.190	2.486	0.705	0.000
Std-Dev	0.891	0.725	0.706	
Male Vs Female <i>p</i>	0.064	0.460	0.142	

Table 5: while tutor - Female self-confidence significantly lower than male self-confidence on pre-survey, but not post-survey

Post-hoc analysis of the response per question: When we did a post-hoc analysis of the questions of the relational tutor individually, we found the same pattern - significantly lower self-confidence of female students on pre-survey, no significant difference between male and female responses on the post-survey - on 3 of the 4 survey questions.

Originally, we had used the relational tutor to test the effectiveness of providing error detection, but not error-correction support during the problem-solving session. This differential treatment could not have influenced the self-confidence responses on the pre-survey since the pre-survey was conducted before the problem-solving session. We conducted a 2 X 2 mixed factor ANOVA with pre-post survey as within-subjects factor and treatment (without versus with error-detection support) as between-subjects factor. We found that there was *no* significant interaction between the type of feedback and pre-post scores [$F(1, 73) = 0.066$, $p = 0.799$]. So, the pattern that we observed with the self-confidence cannot be attributed to the differential treatment provided during the problem-solving session.

When we did a post-hoc analysis of the questions of the selection tutor individually, we found the same pattern on 2 of the 5 questions. Originally, we had used the selection tutor to test the effectiveness of adaptation – changing the problems to suit the learning needs of the student during the problem-solving session. We conducted a 2 X 2 mixed factor ANOVA with pre-post survey as within-subjects factor and treatment (without versus with adaptation) as between-subjects factor. We found that there was *no* significant interaction between the treatment and pre-post scores [$F(1, 83) = 0.038$, $p = 0.847$]. So, the pattern that we observed with the self-confidence cannot be attributed to the differential treatment provided during the problem-solving session.

When we did a post-hoc analysis of the questions of the while loop tutor individually, we found the same pattern on 4 of the 5 questions. Originally, we had used the while loop tutor to test the effect of reflection: whether asking students to reflect on the concept behind each question influenced their learning. We conducted a 2 X 2 mixed factor ANOVA with pre-post survey as within-subjects factor and treatment (without versus with reflection) as between-subjects factor. We found that there was *no* significant interaction between treatment and pre-post scores [$F(1, 55) = 1.507$, $p = 0.225$]. So, the pattern that we observed with the self-confidence cannot be attributed to the differential treatment provided during the problem-solving session.

To summarize, on the 24 questions we asked in the five software tutors:

- On 21 of the 24 questions, the self-confidence of female students was lower than that of male students on the pre-survey. On 9 of these 21 questions, the difference was statistically significant. But, the difference on the corresponding post-surveys was not statistically significant. On 2 of the 24 questions, the self-confidence of female students was higher than that of male students, but the difference was not statistically significant. On the remaining question, the self-confidence of females and males was the same (correct to the third decimal place).
- The difference in the self-confidence of male and female students on the post-survey was statistically significant on only 2 of the 24 questions. On one of these, the self-confidence of female students was lower than that of male students. The self-confidence of females was lower than that of males on the corresponding pre-survey also; and the difference in the pre-post increase of self-confidence between males and females was not statistically significant. This was for the question “How well do you know the boolean data type” asked in the survey of the selection tutor. This is not one of the 9 cases we listed earlier. On the other of the 2 questions, females were more self-confident than males on both the pre-survey and the post-survey, but the question was not related to the topic of the tutor, and was included as a control question: “How well do you know recursion?” presented by the `for` loop tutor.

Except for these last two cases, the recurring pattern that we observed, starting with the two aggregate analyses, and ending with the two post-hoc analyses is that whenever there is a difference between male and female students, the pattern is one of female students starting with lower self-confidence than males, but ending with comparable self-confidence as males after the problem-solving session with the software tutors.

Correlation of self-confidence with learning: [7] observed that girls consistently rate their ability at a lower level than boys despite evidence to the contrary. We wanted to find out whether there was any difference in the prior preparation of male versus female students that could explain the lower self-confidence of female students before the problem-solving session. Could the improvement in the self-confidence of females be explained based on improved learning during the problem-solving session?

In order to answer this question, we considered the pre-post improvement in learning of male and female students during the problem-solving session – recall that the problem-solving session was administered using pre-test-treatment-post-test protocol. We considered the score per problem rather than the raw score in order to eliminate the effect of the increase in familiarity of the students with the user interface from the pre-test to the post-test [14]. We found that:

- The difference in the pre-test scores of male and female students was not statistically significant on the relational tutor, or the `while` loop tutor.
- On selection tutor, the average of females was statistically significantly lower than that of males on the pre-test, but it was also statistically significantly lower on the post-test.

- There was no correlation between the pre-post change in self-confidence and pre-post change in learning of female students ($r = -0.357$ for relational tutor, $r = 0.138$ for selection tutor and $r = 0.221$ for `while` loop tutor).

In other words, the pattern we observed in the self-confidence of females was not similarly reflected in significantly lower prior preparation of females, combined with improvement of females that resulted in post-test scores being no different from those of males. This result is not surprising – in earlier studies, we had found that there is no correlation between the pre-post change in self-confidence and pre-post change in learning of students who used our software tutors [13]. Again, literature in Psychology suggests that we should not expect a correlation between learning and self-confidence because testimonials can mislead [16].

Fall 2006 data: We collected additional data in fall 2006. We repeated the aggregate analysis of the average response of students per software tutor after including this data. We observed the same pattern as in Table 1. We have summarized the results in Table 6.

Responses	Pre	Post	Change	Pre-Post <i>p</i>
Male (N = 534 responses)				
Ave	2.565	2.105	0.460	0.000
Std-Dev	0.858	0.844	0.781	
Female (N = 253 responses)				
Ave	2.735	2.158	0.577	0.000
Std-Dev	0.906	0.858	0.809	
Male Vs Female <i>p</i>	0.013	0.417	0.055	

Table 6: Average response on all the questions of a tutor – spring and fall 2006 combined – Female self-confidence significantly lower than male self-confidence on pre-survey, but not on post-survey

4. CONCLUSIONS

Our quantitative results confirm what had been stated in earlier literature – that female students entering Computer Science have lower self-confidence than male students [4,7,10,12]. We observed that the self-confidence of female students prior to using our problem-solving software tutors was in many cases lower than that of male students, and significantly so.

Using our problem-solving software tutors 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. Since researchers have suggested that self-confidence is one of the factors contributing to the shrinking pipeline [11], problem-solving software tutors can be used to improve the retention of female students in Computer Science.

The improvement in the self-confidence of female students is not correlated with a similar improvement in learning. The improvement in self-confidence seems to result from the very act of using problem-solving software tutors, regardless of whether female students learn from using them. We speculate that using problem-solving software tutors could be improving female students’ self-efficacy [2], reducing their anxiety and/or instilling a sense of accomplishment in them. In any case, using problem-

solving software tutors bestows learning and self-confidence benefits, providing two reasons for using them in introductory Computer Science courses.

We plan to continue to analyze the self-confidence data collected by our software tutors over the next several semesters. Since the prior self-confidence of female students is not lower on all the tutors/questions, we will analyze the data to see if there is a discernible pattern in the questions/tutors on which female students report lower prior self-confidence. We plan to use delayed post-surveys to see whether the improvement in the self-confidence of female students will last.

5. ACKNOWLEDGMENTS

Partial support for this work was provided by the National Science Foundation under grant CNS-0426021.

6. REFERENCES

- [1] Ambrose, S., Lazarus, B. and Nair, I. No Universal Constants: Journeys of Women in Engineering and Computer Science. *Journal of Engineering Education*. 87(4). 1998. 363-368.
- [2] Bandura, A. Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*. 8(2). 1977. 191-215.
- [3] Beckwith, L., Burnett, M., Grigoreanu, V. and Wiedenbeck, S. Gender HCI: What About the Software? *IEEE Computer*. November 2006. 97-101.
- [4] Beckwith, L. and Burnett, M. Gender: An Important Factor in Problem-Solving Software? *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing Languages and Environments*. IEEE Press. 2004. 107-114
- [5] Beyer, S., Rynes, K., Perrault, J., Hay, K. and Haller, S. Gender differences in Computer Science Students. *SIGCSE Bulletin*, Vol 35, 2003, 49-53.
- [6] Camp, T. The Incredible Shrinking Pipeline. *Communications of the ACM*. 40(10). Oct 1997. 103-110.
- [7] Chen, M. Gender and Computers: The Beneficial Effects of Experience on Attitudes. *Educational Computing Research*. 2(3). 1986
- [8] Cohoon J.M., Gendered Experiences of Computing Graduate Programs. *Proceedings of The 38th SIGCSE Technical Symposium on Computer Science Education (SIGCSE 2007)*, Covington, KY, March 2007, 546-550.
- [9] De Palma, P. Why Women Avoid Computer Science. *Communications of the ACM*. 44(6). June 2001, 27-29
- [10] Fisher, A. and Margolis, J. Unlocking the Clubhouse: The Carnegie Mellon Experience. *SIGCSE Bulletin Special Issue on Women and Computing*, Vol. 34(2), June 2002.
- [11] Gurer, D. and Camp, T. An AMC-W Literature Review on Women in Computing. *SIGCSE Bulletin Special Issue on Women and Computing*, Vol. 34(2), June 2002. 121-127
- [12] Klawe, A. Girls, Boys and Computers. *SIGCSE Bulletin Special Issue on Women and Computing*. Vol 34(2), June 2002, 16-17
- [13] Kumar, A.N. The Effect of Using Problem-Solving Tutors on the Self-Confidence of Students. *Proceedings of the 18th Annual Psychology of Programming Workshop (PPIG 06)*, Brighton, U.K., 9/7-8/2006, 275-283.
- [14] 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*, Sept 2006.
- [15] Kumar, A.N., Learning Programming by Solving Problems, in *Informatics Curricula and Teaching Methods*, L. Cassel and R.A. Reis ed., Kluwer Academic Publishers, Norwell, MA, 2003, 29-39.
- [16] McCord, J. A thirty-year follow-up on treatment effects. *American Psychologist*. Vol 33. 284-289. 1978.
- [17] 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.