PREDICTING (AND CREATING) SUCCESS IN CS1

Carl Farrell, Hawaii Pacific University, cfarrell@hpu.edu

ABSTRACT

This paper describes a testing process that has shown significant validity at measuring preparedness for success in a rigorous Computer Science 1 course. Basic skills, knowledge, and conceptual abilities are tested with a simple exam, and the results have proven to be very useful for predicting future success in CS1. The exam also provides prescriptive information about areas of deficiency in which students need improvement to be adequately prepared for CS1. This placement exam has been used to determine which students need additional preparation before entering CS1, and a CS0 course has been created to provide a way for students to get a stronger background in many of the areas that are needed for CS1. Details of the testing exam and its predictive abilities are described in the paper, along with recommendations for its most effective use.

Keywords: CS1, Placement Testing, Student Needs Assessment, Assessment, CS0.

INTRODUCTION

Learning computer science can be a lot like learning how to ski. If someone starts on a slope that is too difficult for them, they are very apt to go too fast, possibly get injured, and never master the fundamentals that are necessary for success. This situation has been a concern among university computer science programs for more than 20 years [1, 7]. Problems and challenges have included a wide diversity of skills and aptitudes among entering students, loss of students due to poor placement, and non-existence of necessary preliminary courses. Many attempts have been made to design tests that can reliably predict whether or not students are adequately prepared for CS1 courses, and a number of studies have attempted to determine specific testable attributes that correlate highly with success in CS1. Some studies have attempted to determine criteria that correlate highly with all final grade outcomes in CS1 [2, 4, 9]; while a number of others have focused more specifically on identifying students that are likely to fail CS1 [3, 5, 7, 8]. A summary of the results from a number of previous studies has shown that only 11 to 40 percent of final grade variability was explained by surveyed attributes [4]; while much more success has been

achieved in discriminating students that will be successful in CS1 from those that will not be successful.

This project combines the use of a placement test and the creation of a new lower level course to help students that are not adequately prepared for CS1. During the period of this study, students have not been required to pass the placement test to gain admission to CS1, and enrollment in the new CS0 course has been voluntary. This policy has allowed time to tune and validate the placement test.

The overall goal of this project is to contribute to the ultimate success of students in CS1. One specific sub-goal is the development of a placement test that can accurately identify three groups of students: those that are very likely to fail CS1 and should definitely take CS0 first; those that would benefit from CS0, but who might be able to pass CS1 if they are highly motivated; and those that are well-prepared for CS1. In addition, it is hoped that the placement test will be able to provide prescriptive information about individual student weaknesses that can possibly be addressed, and it is hoped that information from general results of the test may contribute to useful design ideas for the CS0 course.

DESIGN OF THE PLACEMENT EXAM

The setting for this research is the computer science program at Hawaii Pacific University (HPU), which has a CS1 course designed around the Computing Curricula 2001 recommendations [6]. The only required prerequisite for this CS1 course is proficiency in intermediate algebra. In planning the design of a placement test to determine student preparedness for CS1, many of the same questions were considered that others have faced concerning the kind of test to be used, what options should be offered to students with identified deficiencies, and whether or not to enforce assessment results [3].

The exam was designed so that it could be easily administered, easily taken, easily scored, and possible to offer online.

Although some studies have demonstrated significant correlations between success in CS1 and various self-assessed characteristics such as "comfort" with

programming assignments [2] or not feeling dependent upon luck [9], this study only focuses on knowledge and skills that have been identified previously as contributing to CS1 success and which could be targeted for prescriptive assistance to students.

A set of multiple-choice questions focusing on basic math, logic and problem-solving skills, similar to those studied by others [1, 4], has been used informally at HPU as a pretest in CS1 courses for a number of years. It has been observed that students scoring below 66% on these questions almost never succeed in the CS1 course. These questions were reviewed by a committee of the faculty, and 15 questions were chosen to be fairly evenly distributed across the following conceptual areas: algebra, logic, discrimination, algorithmic processes, attention to detail, applying abstract concepts, and solving simple word problems.

A previous study reported that "The hurdle in learning programming appears to be learning the basic concepts such as sequence, iteration, and decision" [5], and one of our questions specifically included those aspects. This question was written in "algorithmic English" rather than pseudocode, as used by others [3], since there is no required programming prerequisite for CS1 at HPU.

Questions and answers were edited carefully to minimize the likelihood of wrong answers due to minor errors. All questions were evenly weighted and there was no penalty for wrong answers.

A primary goal of the exam was to identify three separate groups of students: those that would almost certainly not pass CS1 without better preparation; those that might be able to pass CS1, but that would probably benefit by taking CS0; and those that are well prepared for CS1.

The exam was designed to be given online, and it was made available for students to take online if they wished during the period of this study. However, to ensure as complete a data sample as possible for this study, the exam was also given during the first class meeting of the CS1 course, and only those scores were used in this project.

A sample of the questions that were used on the placement exam is shown in Figure 1.

1. What value of X would be a correct solution to the equation $6 / X = 12$?

   A) 0.5
   B) 2
   C) 6
   D) 72
   E) none of the above

2. How many different solutions are there to the equation $|X| - 1 = 6$?

   A) 0 solutions
   B) 1 solution
   C) 2 solutions
   D) 4 solutions
   E) none of the above

3. This question is a bit tricky -- read it carefully. Students must score higher than 150 on a certain test to get an A, and the 10 students in a class received scores of 110, 130, 135, 139, 140, 145, 149, 150, 152, and 155. How many students got an A on the test?

   A) 0 students
   B) 1 student
   C) 2 students
   D) 3 students
   E) none of the above

4. A number is an even number if it is an integral multiple of 2 (e.g., 0, 2, 4, 6, 8, 10, etc. are even numbers). Perform the steps described below, and then answer the question at the end.
   Step 1: Write the number 1.
   Step 2: If the last number you have written is an even number, add 1 to your last number and write down the new number; otherwise, just continue to step 3.
   Step 3: Double the last number you wrote, and write that new number.
   Step 4: If the last number you wrote is less than 6, repeat steps 2, 3, and 4; otherwise, stop.

   What was the last number you wrote?

   A) 2
   B) 6
   C) 10
   D) 14
   E) none of the above

Figure 1. Sample of Questions Used on the Placement Exam

5. The expression (x mod y) means the integer remainder that would be left over if you divided x by y; for example, 7 divided by 2 goes 3 times with an integer remainder of 1, so (7 mod 2) is equal to 1.

   What is the value of the following expression?
   (7 mod 5) + (4 mod 5) + (10 mod 5)

   A) 0
   B) 6
   C) 8
   D) 10
   E) none of the above

Figure 1 (cont.). Sample of Questions Used on the Placement Exam

DESIGN OF THE NEW CS0 COURSE

The design of a new course to prepare weaker students for CS1 shared a goal that has been described by some others in the following way: "The view of this [CS0] course was not one of providing a particularly broad view of computing for non-majors (as is typical of most CS0 courses), but with the goal of providing *majors* with the appropriate mental models and conceptualizations of programming for future success in CS1" [3]. The name of the new course is not something like "Introduction to Computer Science," but instead "Foundations of Programming." Even though the CS1 course is as much about problem-solving as programming specifically, it was felt that the fundamental concepts of programming provided the most significant hurdle that weaker students needed to overcome in order to become adequately prepared for CS1.

The content for the new course includes a broad introduction to the discipline of computer science as well as object-oriented methodologies; however, emphasis is placed on aspects that are addressed by the placement exam but not covered in other mathematics courses, particularly the process of algorithmic problem-solving and the foundations of structured programming. This course is not expected to go as far or as fast as CS1, but to provide students with extensive exercises and practice in basic algorithmic constructs [5]. Hands-on lab work is an integral part of the course, and the Java language is used since that is what is currently used in the CS1 course.

RESULTS

The 15-point placement exam was first given in the Fall of 2005 to 45 students that registered and attended the first class meeting of CS1. These students consisted of approximately even groups of CS majors, CIS majors, Engineering majors, and other majors. Since most of the questions had 5 possible answers, and no deduction was made for incorrect answers, purely random answers would be expected to create a score of about 3. Consequently, scores below 3 weren't expected on the exam. In fact, the scores ranged from a low of 4 to a high of 14, with a mean of 10.36, a median of 11, and a standard deviation of 2.29. The same exam was given to the 25 students that attended the first class meeting of CS1 at the start of Spring 2006. In that case, the scores ranged from 4 to 15, with a mean of 10.36, a median of 11, and a standard deviation of 2.41. Obviously, the results had a very high degree of consistency across the two samples; although it is certainly not expected that the mean score for future samples will be so precisely similar again.

This project made no attempt to study any relationships between the declared majors of students and their outcomes in CS1. In the past, correlations between majors and CS1 outcomes has been very weak; and more importantly, the goal of the current project is to develop an exam that can identify students, regardless of major, that need additional preparation before taking CS1.

Tables 1 and 2 below show the final grades of students in CS1 for each of the two semesters that were studied, organized by the scores that those students received on the placement exam.

Table 1: Final Grades of Students in CS1 during Fall 2005 for each Score on the Placement Exam

Score	Grade Outcomes -- Fall, 2005			
15				
14	A	B		
13	AAA	BB		W
12	AA	BB	C	D*
11		BBB	CC	FFFF
10	AA	B	C	FFFWWWW
9	AA		C	F
8				
7			D**	FFW
6				W
5				W
4				W

* This student got a B on the CS1 final exam, but had chosen not to do required homework. He was deemed ready to continue into CS2.

** This student chose to retake the CS1 course.

Table 1: Final Grades of Students in CS1 during Spring 2006 for each Score on the Placement Exam

Score	Grade Outcomes -- Spring, 2006				
15	A				
14		B			
13	A		C		
12	AAAA		C		
11	AA				F
10	A			D	FW
9	A	B	C		F
8	A		C		F
7					W
6					
5					
4					W

Overall, a linear regression analysis of these data shows that 26% of the variability in final grade point outcomes is explained by scores on the placement exam. This is very similar to, and in some cases better than, the results that have been observed in previous studies [1, 2, 4, 7, 9]; however, this is not a high degree of explanation for all grade outcomes. On the other hand, the purpose of this study was not to predict all grade outcomes in CS1, but rather to identify students in need of additional preparation before entering CS1. Leeper [7] showed that a test with a 26% overall explanatory power for individual grades could explain 64% of the variability in whether students passed or failed CS1.

It was hoped that the exam would consistently be able to predict a bottom group of students that would not pass CS1, and a top group that would all be very likely to pass CS1. The tables of results have been separated with dashed lines to indicate the three groupings that seemed to emerge. Students with a score of 12 (80%) or higher on the placement exam nearly always passed the CS1 course in both data samples. Students with a score of 7 (47%) or lower essentially never passed CS1, and this result was also consistent across both data samples. The only exceptions to these predictions occurred in the case of one individual that withdrew from CS1 in Fall 2005 for external reasons, and in two cases that are footnoted in Table 1, which actually corroborated the prediction from the placement test. Therefore, the predictive reliability of the placement exam for students that scored above 11 or below 8 was essentially 100%.

The rather large group of students that fell in the middle range of approximately one standard deviation from the mean received no predictive value from the placement exam, and it was felt that their final grades tended to correlate much more highly with their personal motivation and effort in CS1 than with their score on the placement exam. Additional factors, beyond the questions used on the placement exam, will need to be studied in order to identify students in the middle group who are most at risk if they take CS1 without additional preparation.

Even though students were advised in both semesters that a score of 7 or lower on the placement exam almost certainly indicated that they needed to take CS0 before attempting CS1, only one student in Spring 2006 transferred into the CS0 course. The poor outcomes of those students that had placement scores below 8 show that the predictions of the placement test were accurate.

Detailed item analysis for each of the questions on the placement exam, along with categorization of the questions by CS1 topic area and Bloom's taxonomy, as was done by Soh [8], will provide further information in the future and may allow further tuning of the questions on the exam with the goal of accurate predictions for a larger percentage of students.

It was originally hoped that knowledge of the specific questions that students missed on the placement exam would provide prescriptive assistance in helping individual students and in tailoring the content of the CS0 course to address common problem areas. The first of these goals was mostly unmet, largely because even after students were informed about areas where they seemed to need improvement, they either ignored recommendations or didn't have good resources available for help. The second goal of tailoring content of CS0 to address common problems is being considered for the future.

The CS0 course was cancelled in Fall 2005 because not enough students self-selected into it, and it also had a low enrollment in Spring 2006. This problem of low enrollments may change if it becomes a requirement that students pass the CS1 placement exam (or pass CS0) before being admitted into CS1. The results of CS0 students when they later take CS1 will also be studied in the future.

CONCLUSIONS

Just as beginning skiers may get hurt and fail to master necessary fundamentals if they start on a slope that is too steep and fast, so too may beginning computer science students suffer similar problems if

they start in CS1 before they are adequately prepared for such a rigorous course. One goal of this project has been the prevention of some of the "injuries" and problems that have occurred in the past with students who were not ready for CS1.

A placement exam has been constructed, which provides significant value by accurately predicting success for students scoring above a certain level. The exam has also been shown to be very useful in identifying students that will not succeed in CS1 without additional preparation. Further work needs to be done to identify other factors that may contribute to student difficulty in CS1, so that a greater proportion of failures can be prevented.

During the period of this study, students were allowed to ignore the warnings presented by the placement exam, and they tended to do so. It is felt that these weaker students will be better served in the future if the results of the placement exam are enforced, so that students will be required to either pass the placement exam or take a preparatory course before being admitted into CS1.

Just identifying students that are inadequately prepared for CS1 is not sufficient to create success for those students. A new CS0 course has been created to provide a way for poorly prepared students to strengthen their knowledge, skills, and experience so they can later be successful in CS1. Prescriptive diagnoses are being developed from the results of the placement exam for the purpose of individually advising students and also to identify any common areas of weakness that need to be addressed in CS0.

REFERENCES

1. Barker, J. & Unger, A. (1983). A Predictor for Success in an Introductory Programming Class Based upon Abstract Reasoning Development. *Proceedings of the fourteenth SIGCSE technical symposium on Computer science education*, 154-158. Available at http://portal.acm.org/ citation.cfm?id=801037.

2. Bergin, S. & Reilly, R. (2005). Programming: Factors that Influence Success. *Proceedings of the thirty-sixth SIGCSE technical symposium on Computer science education*, 411-415. Available at http://portal.acm.org/citation.cfm?id= 1047453.

3. Dierbach, C., Taylor, B., Zhou, H. & Zimand, I. (2005). Experiences with a CS0 Course Targeted for CS1 Success. *Proceedings of the thirty-sixth SIGCSE technical symposium on Computer science education*, 317-320. Available at http://portal.acm.org/ citation.cfm?id=1047453.

4. Evans, G. E. & Simkin, M. G. (1989). What Best Predicts Computer Proficiency? *Communications of the ACM, 32*(11), 1322-1327.

5. Holden, E. & Weeden, E. (2004). The experience factor in early programming education. *Proceedings of the 5th conference on Information technology education,* 211-218. Available at http://portal.acm.org/ citation.cfm?id=1029585.

6. Joint Task Force on Computing Curricula (2001). Computing Curricula 2001, *Journal on Educational Resources in Computing, 1*(3es).

7. Leeper, R. & Silver, J. (1982). Predicting Success in a First Programming Course. *Proceedings of the thirteenth SIGCSE technical symposium on Computer science education*, 147-150. Available at http://portal.acm.org/ citation.cfm?id=801357.

8. Soh, L., Samal, A., Person, S. Nugent, G. & Lang, J. (2005). Designing, Implementing, and Analyzing a Placement Test for Introductory CS Courses. *Proceedings of the thirty-sixth SIGCSE technical symposium on Computer science education*, 505-509. Available at http://portal.acm.org/ citation.cfm?id=1047504.

9. Wilson, B. & Shrock, S. (2001). Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors. *Proceedings of the thirty-second SIGCSE technical symposium on Computer science education*, 184-188. Available at http://portal.acm.org/ citation.cfm?id=364581.