The Programming Performance Prophecies

Predicting Student Achievement in a First-Year Introductory Programming Course

Jeff Ringenberg, Marcial Lapp, and Apoorva Bansal
{jringenb, mlapp, apoorvab}@umich.edu

Presented at the *Sixth Annual Research and Scholarship in Engineering Education Poster Fair*. 02/07/12.

Research Question

Is it possible to predict student performance in an introductory programming course without testing any specific programming knowledge?

Abstract

Ensuring student success in first-year introductory programming courses presents a unique challenge when considering the diversity of educational backgrounds. In order to create a more equitable experience and to ensure that students are placed in a course whose difficulty is commensurate with their abilities, we present a methodology to predict student performance in first-year introductory programming courses such as ENGR 101 and 151.

Our Approach

- Give an online survey at the start of the semester before the students have received any formal instruction.
- Alternatively, have all incoming students take the survey during Summer Orientation.
- The survey contains a combination of algorithmic, math, and logic-based questions intended to test core, prerequisite knowledge of the course material.
- Questions ***must*** be programming language agnostic.
- Use the results of the survey to predict student performance and ultimately advise students which first-year programming courses they should take (e.g. a standard programming course such as ENGR 101 or an accelerated programming course such as ENGR 151).

Experimental Design

Fall 2009 – Initial release of survey (25 questions / 30 mins)
Class sample size: ~350 students

Stage I: Question Analysis and Refinement

Fall 2010 – Survey Administration (16 questions / 30 mins)
Class sample size: ~500 students

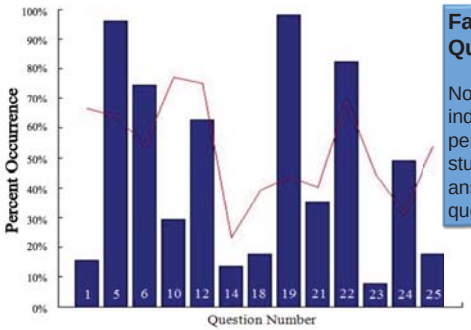
Stage II: Question Analysis

Fall 2011 – Survey Administration (16 questions / 45 mins)
Class sample size: ~550 students

Stage III: Final Analysis

Question Analysis

1. Ignore questions that over 80% of students answered correctly. These are considered ***ineffective***.
2. Create sets containing 7 ***effective*** questions.
3. Correlate set scores to overall course grades.
4. Ignore sets with a correlation of less than 0.4. These are considered ***ineffective***.
5. Calculate the percentage of the ***effective*** sets in which each ***effective*** question appears.



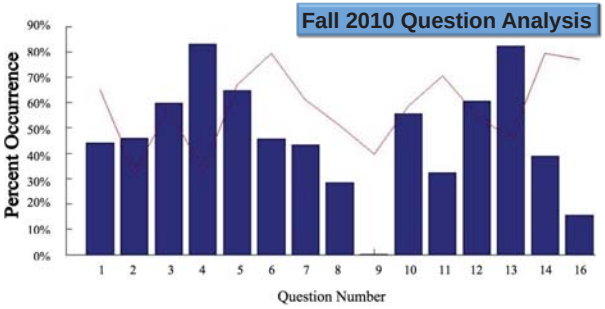
Fall 2009 Question Analysis

Note: The red line indicates the percentage of students that answered the question correctly.

Question Refinement

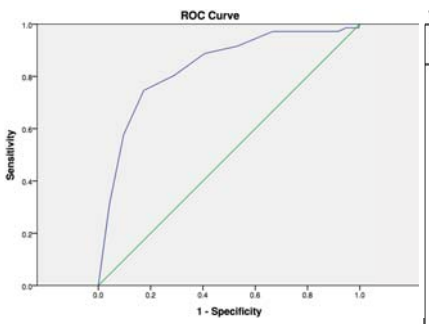
- The most effective questions from Fall 2009 (i.e. 5, 6, 12, 19 and 22) were used in Fall 2010.
- Inspired by the effective questions from Fall 2009, a new group of related questions was then added.

Fall 2010 Question Analysis



Final Analysis

- ROC Curve Analysis was used to pick a survey score to decide which course, ENGR 101 or 151, would be most appropriate for a student to take.



Coordinates of the Curve
Test Result Variable(s):Survey

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.00	1.000	1.000
1.00	.986	.996
2.50	.986	.990
3.50	.986	.978
4.50	.986	.947
5.50	.972	.917
6.50	.972	.840
7.50	.972	.764
8.50	.972	.667
9.50	.915	.531
10.50	.887	.406
11.50	.803	.289
12.50	.746	.174
13.50	.577	.097
14.50	.310	.042
15.50	.028	.004
17.00	.000	.000

Acknowledgements

We would like to thank Ella Atkins, Alex Bielajew, and Mike Wellman for their feedback on the initial pool of survey questions and Brady West from CSCAR for his help with the ROC Curve Analysis.