

Teaching Software Testing with Automated Feedback

James Perretta and Andrew DeOrio

EECS/CSE Division, University of Michigan

Abstract

- CS courses use automated grading to evaluate student code with a variety of feedback policies.
- There is little research on the impact of these policies on student learning.
- We examine the association between different types of feedback and student learning of software testing.

Research questions:

- Does automated feedback improve students' ability to write high-quality test cases?
- What type of automated feedback best encourages student learning of software testing?

Methods

- 1,556 students enrolled in CS2 (EECS 280) over two semesters. The first semester comprised the control group, and the second semester comprised the experiment group.
- Students worked alone or with a partner.
- Students submitted test cases for 3 projects. Tests case quality was evaluated as number of buggy instructor solutions exposed.
- Control and experiment groups received different feedback on their test cases for Project 3 and Project 4, same feedback for Project 5.

Test Case Evaluation

Student test cases were evaluated as follows:

- Each student test is run against a correct instructor solution. Tests that fail are marked as false positives.
- Each test that was free of false positives is run against a series of intentionally buggy instructor solutions. Buggy solutions for which any tests fail are marked as exposed.
- Students are awarded points based on how many buggy solutions their tests exposed.

Results

Independent variables

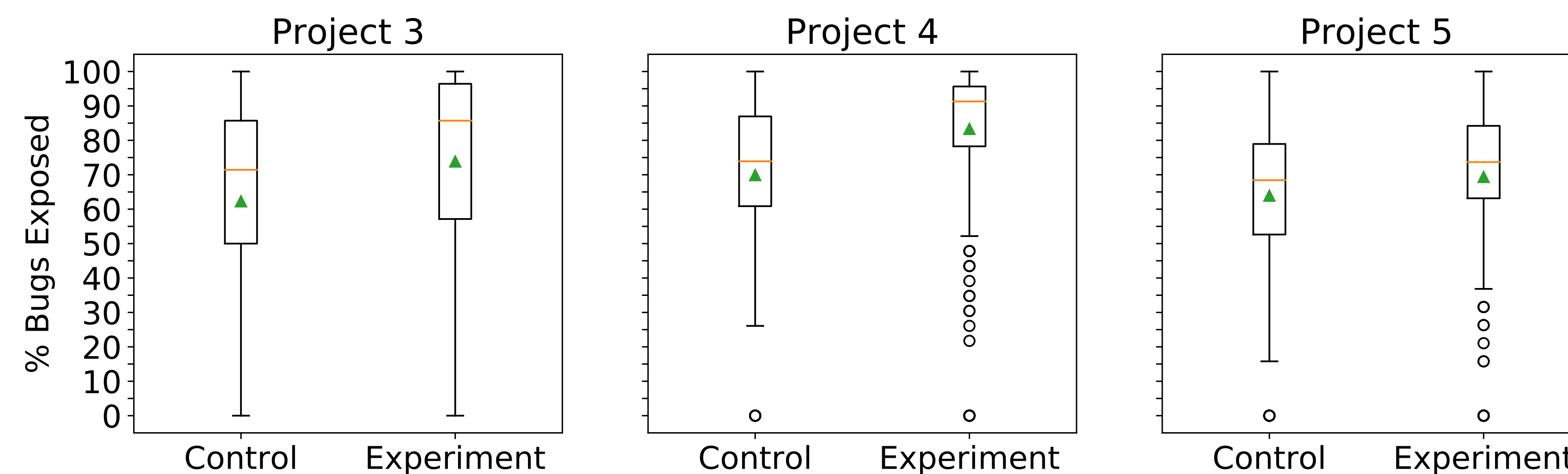
- Test case feedback type (control and experiment groups)
- Partnership status (alone or with a partner)
- GPA

Dependent variables

- Student test case quality (number of instructor buggy solutions exposed)

	Project 3				Project 4				Project 5			
	df	Sum Sq.	F	PR(>F)	df	Sum Sq.	F	PR(>F)	df	Sum Sq.	F	PR(>F)
Feedback	1	2.2	40.95	2.34e-10	1	3.43	114.92	1.64e-25	1	0.46	12.04	5.44e-04
Partner	1	3.03	56.32	1.31e-13	1	1.59	53.38	5.45e-13	1	1.24	32.29	1.75e-08
Feedback x Partner	1	0.01	0.11	7.39e-01	1	0.27	8.97	2.81e-03	1	0.14	3.6	5.82e-02
GPA	1	25.91	481.46	3.19e-88	1	11.76	394.25	1.08e-74	1	9.66	251.18	1.36e-50
GPA x Feedback	1	0.02	0.34	5.60e-01	1	0.0	0.12	7.26e-01	1	0.04	1.02	3.14e-01
GPA x Partner	1	0.0	0.0	9.63e-01	1	0.15	4.9	2.71e-02	1	0.0	0.02	8.88e-01
GPA x Feedback x Partner	1	0.0	0.07	7.87e-01	1	0.07	2.4	1.21e-01	1	0.06	1.56	2.11e-01
Residual	1056.0	56.83			1045.0	31.17			991.0	38.12		

Test Case Quality and Feedback Type



Feedback Types

Control group

Test Case	Passed	Score
Student List test validity check	✗	0/1

Test case List_test_bad.cpp incorrectly exposed the correct solution as buggy

Shown only which student test cases had false positives.

Experiment group

Test Case	Passed	Score
Buggy List solution 1	✓	1/1
Buggy List solution 2	✓	1/1
Buggy List solution 3	✓	1/1
Buggy List solution 4	✓	1/1

In addition to what the control group received, shown how many instructor buggy solutions their tests exposed (P3 and P4 only).

Discussion

- Students in experiment group wrote higher-quality test cases on all 3 projects, even when additional feedback was removed.
- Students who worked with a partner wrote higher quality test cases on all 3 projects.
- Associations were smaller in magnitude than the association between GPA and test case quality.

Limitations

- Difficulty of exposing instructor buggy solutions may have varied between assignments.
- Control and experiment groups came from different semesters.
- Students chose whether to work with a partner.

Conclusions

- Students who received feedback on the number of buggy solutions exposed wrote higher-quality test cases, even after that feedback was removed.
- Students who worked with a partner wrote higher-quality test cases.