How Effective and Efficient Are Student-Written Software Tests?

Amanda Showler, <u>Michael A. Miljanovic</u>, Jeremy S. Bradbury

Software Engineering & Education Research Lab
Ontario Tech University, Oshawa, ON, Canada
{amanda.showler, michael.miljanovic}@ontariotechu.net,
jeremy.bradbury@ontariotechu.ca

http://www.seerlab.ca

SEER LAB | Ontario Tech UNIVERSITY

SIGCSE
Technical Symposium 2025

Problem & Motivation

- Recent CS graduates have insufficient **software testing skills** [1]
  - Student tests have low **bug detection** rates
  - Likely to only test **valid** inputs ("**happy paths**")
- Most work in this area focuses on **White Box** / **Unit tests** and not **Black Box** Testing
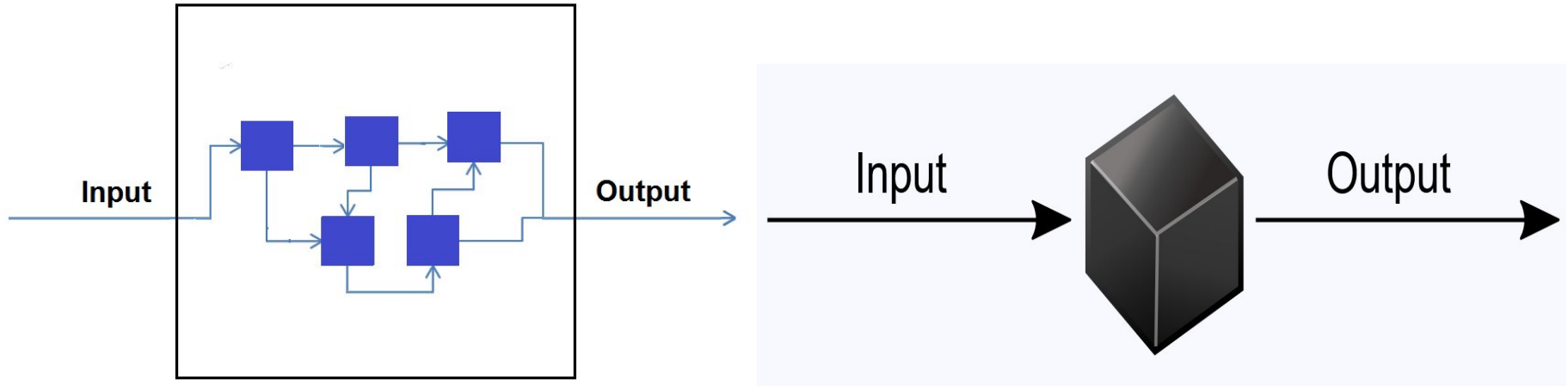
[1] Stephen H Edwards and Zalia Shams. Do student programmers all tend to write the same software tests? In Proceedings of the 2014 conference on Innovation & technology in computer science education, pages 171–176. ACM, 2014.

Test Suite Efficiency and Effectiveness

- **Test Bucket**
  - Group of related test cases for a specific **requirement**
- **Inefficiency**
  - High test **redundancy**
- **Ineffectiveness**
  - Low **variation** in **buckets** tested
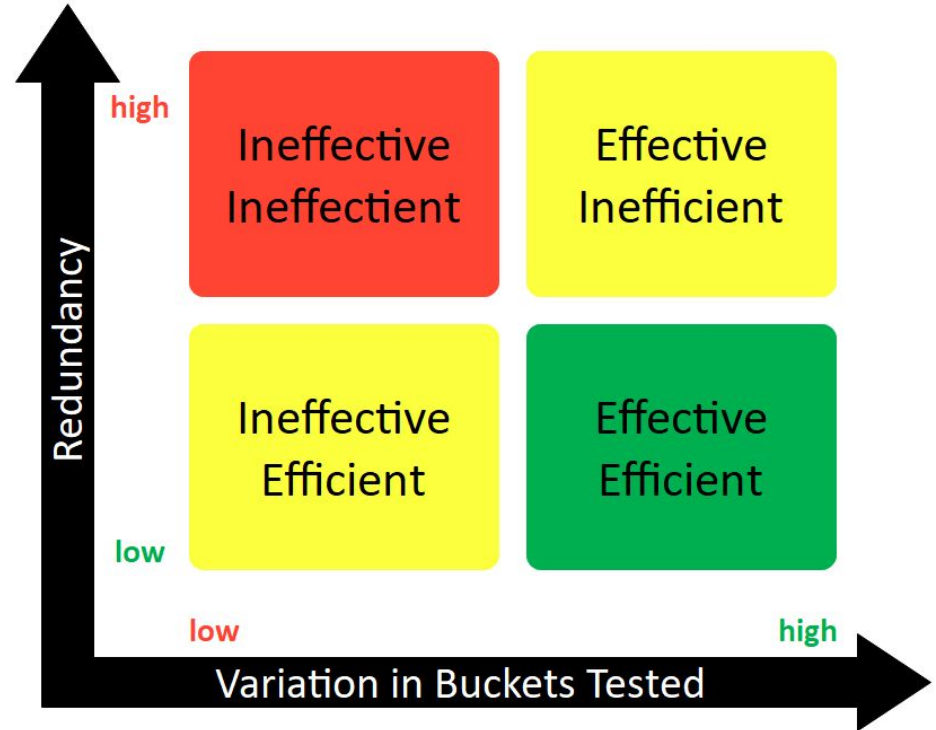
Test Features (example)

ATM Bank System
Requirement: ● A maximum of $1000.00 can be deposited to an account at a time.

1. Requirement
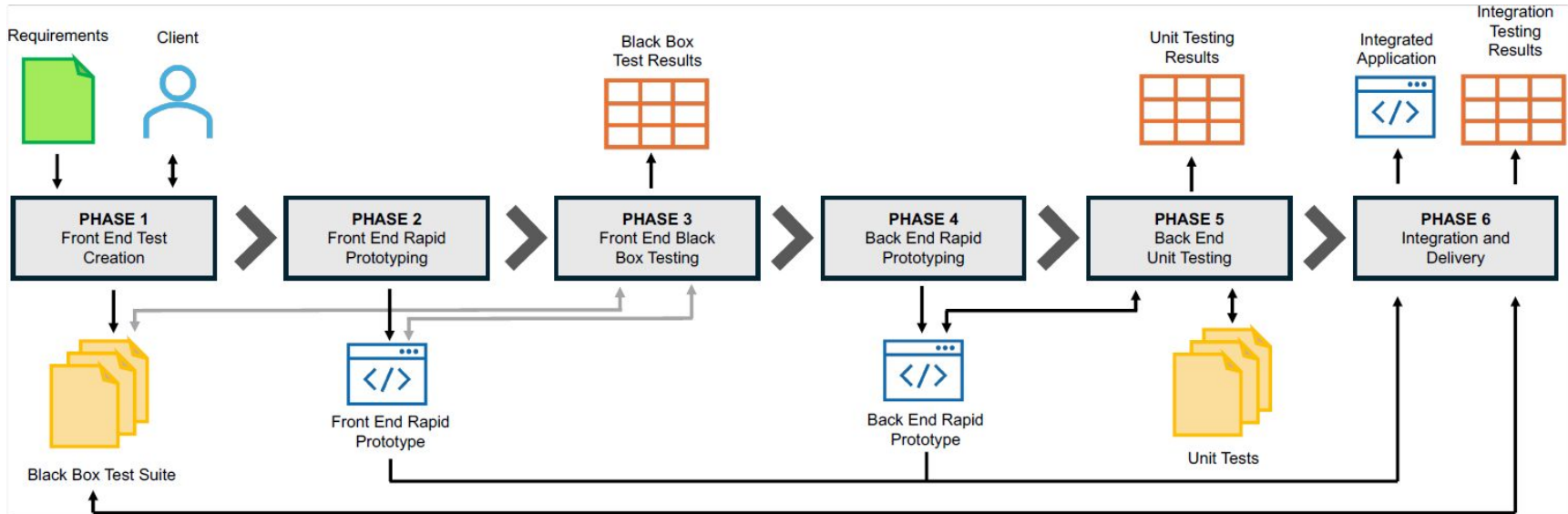   - **Explicit**: `Maximum Deposit is $1000`
   - **Implicit**: `Deposit cannot be negative`

2. Input
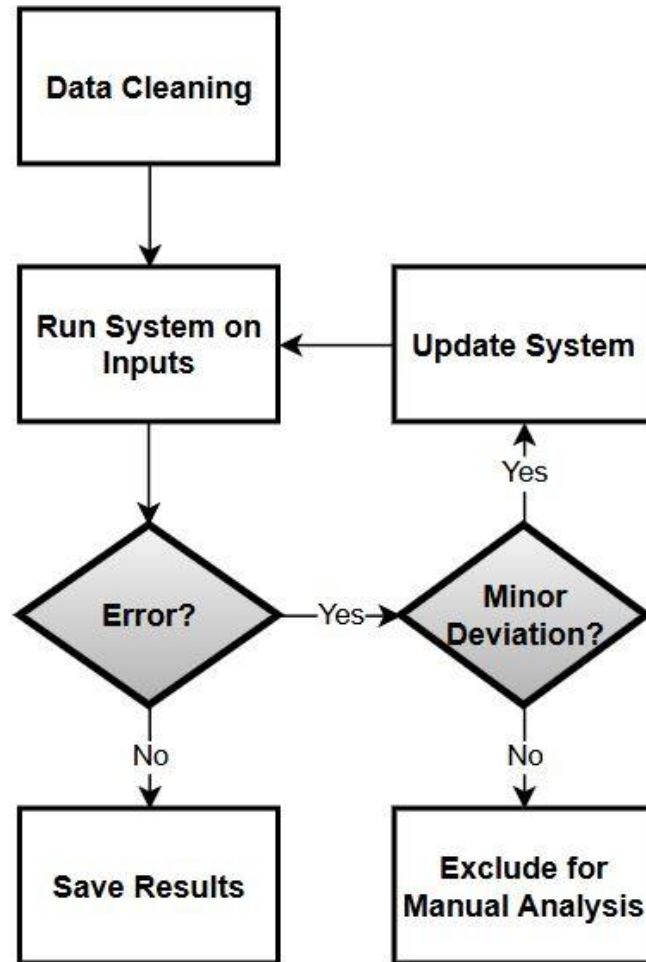   - **Happy Path**: `$0 < Deposit <= $1000`
   - **Invalid Input**: `Deposit > $1000`

Software Quality Assurance (SQA) Course Project

- Students develop both **system** and **test suite**
- Same set of **requirements** for all students

System

Data Exclusion Criteria

- Group's test suite **deviated** too much from expected format
- Test bucket was **missed** by the class
- Test bucket is only tested by **1 group**
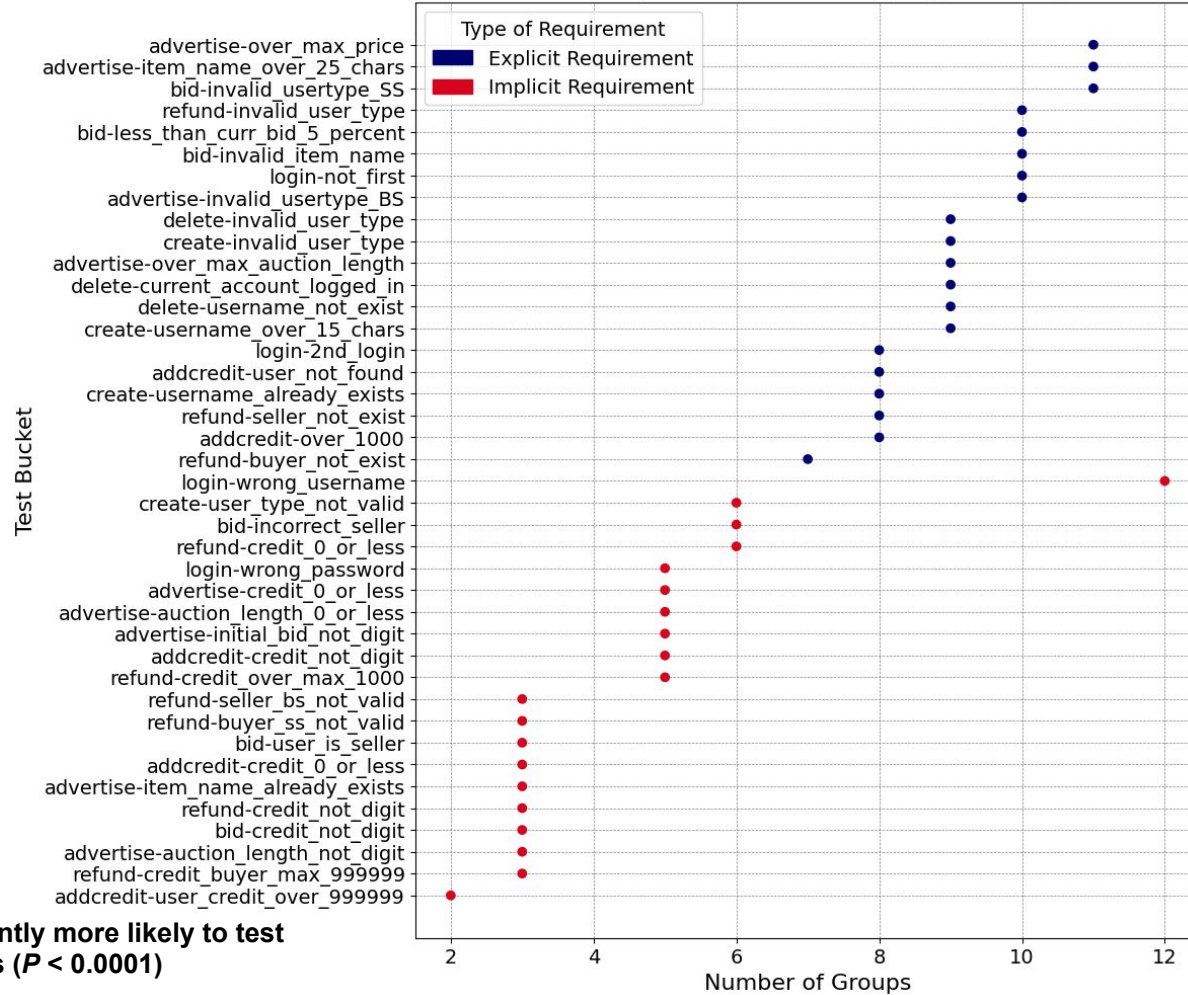
```
1  Login
2  Admin
3  Create
4  tempA
5  FS
6  Logout
```

Results

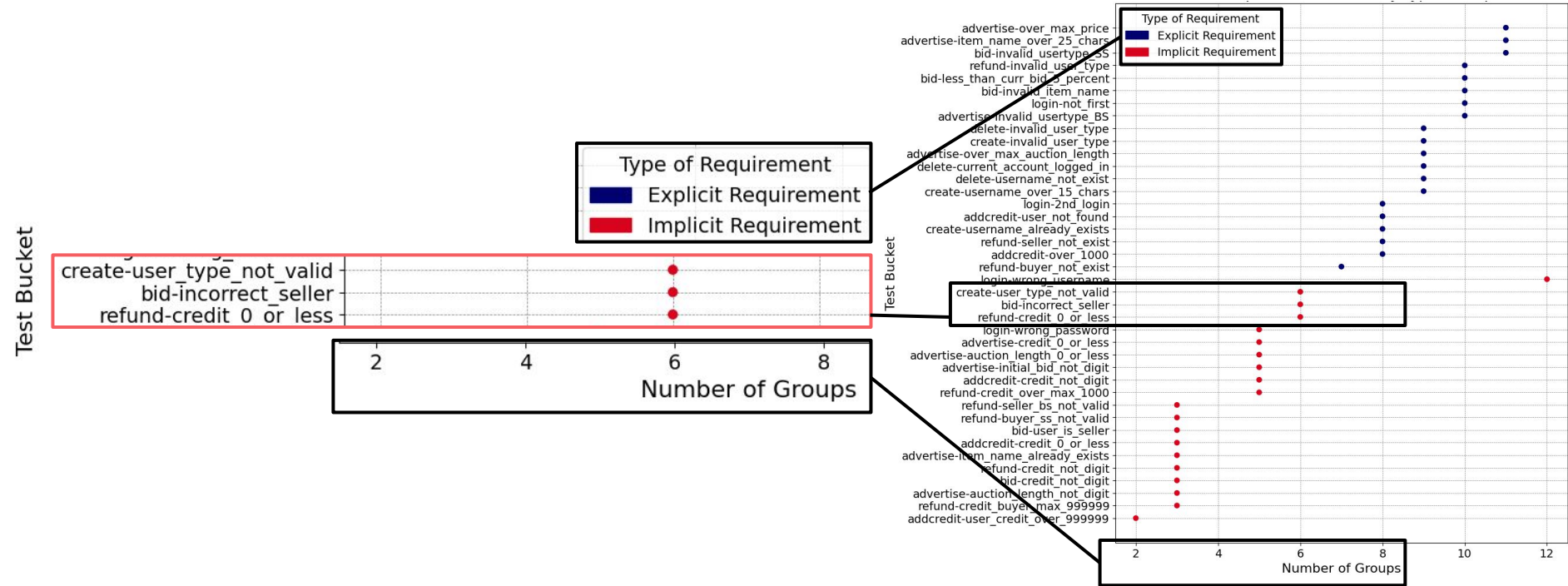| Winter 2023	Winter 2024
Auction Bidding System ● 12/17 test suites ● 49 test buckets ● 1014 tests	Game Distribution System ● 6/15 test suites ● 36 test buckets ● 381 tests

Results

| Winter 2023	
Auction Bidding System ● 12/17 test suites ● 49 test buckets ● 1014 tests	

Number of Groups vs. Test Bucket by Type of Requirement

***Students significantly more likely to test explicit requirements (*P* < 0.0001)

Number of Groups vs. Test Bucket by Type of Requirement

***Students significantly more likely to test explicit requirements ($P < 0.0001$)

11

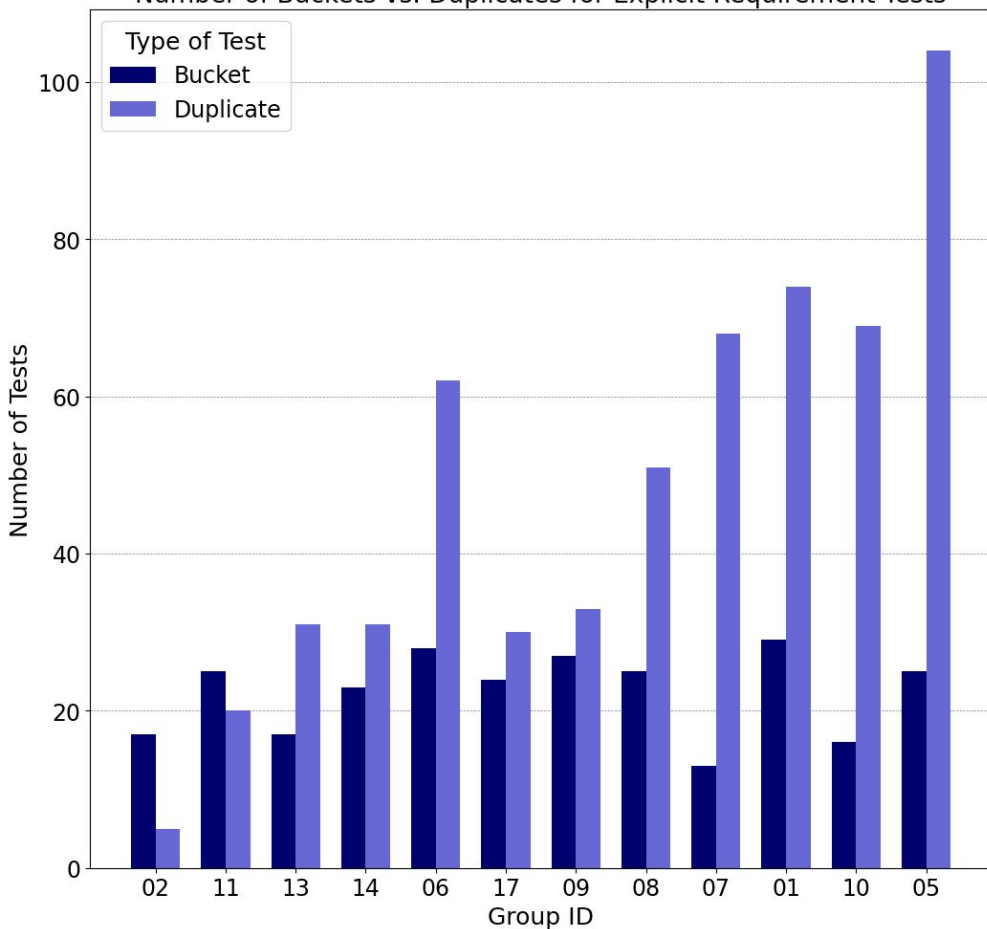Number of Buckets vs. Duplicates for Happy Path Tests
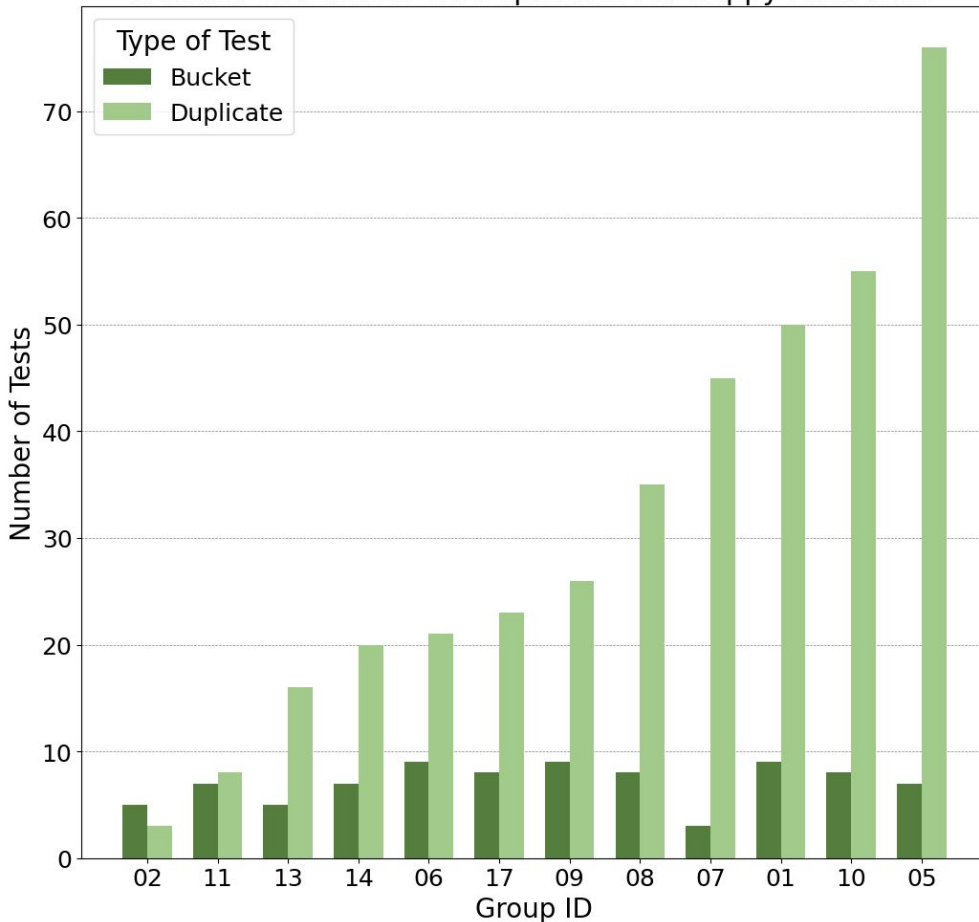
Number of Buckets vs. Duplicates for Invalid Input Test
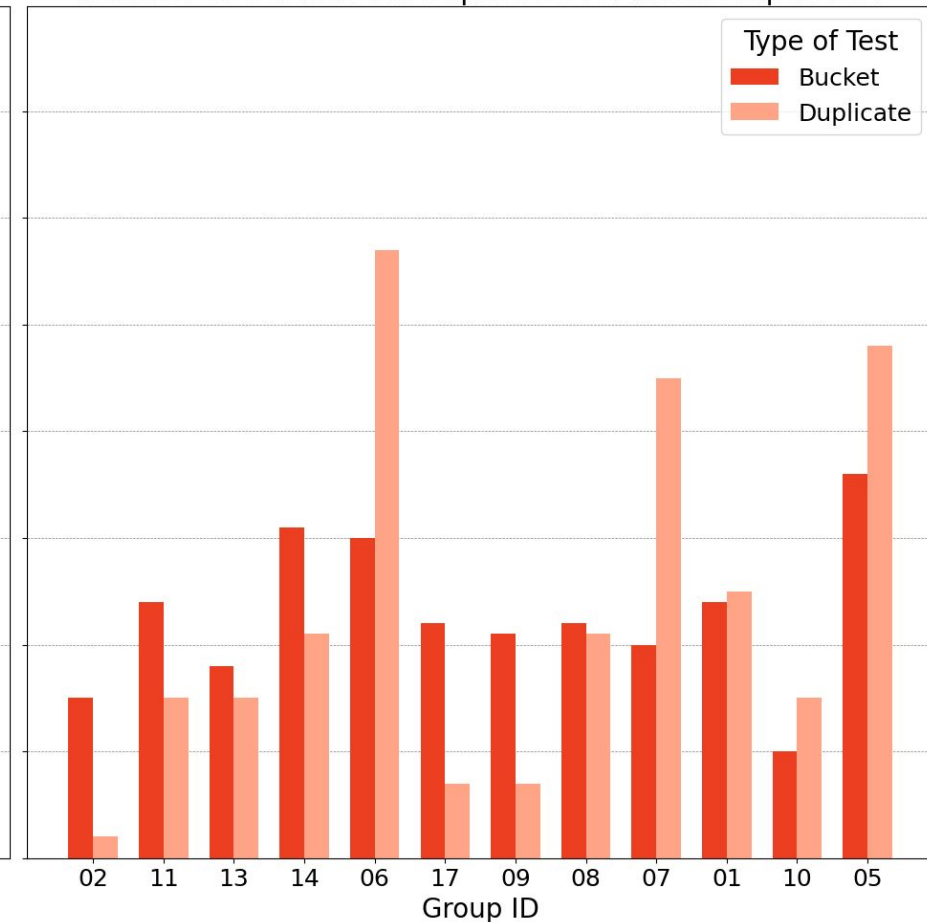
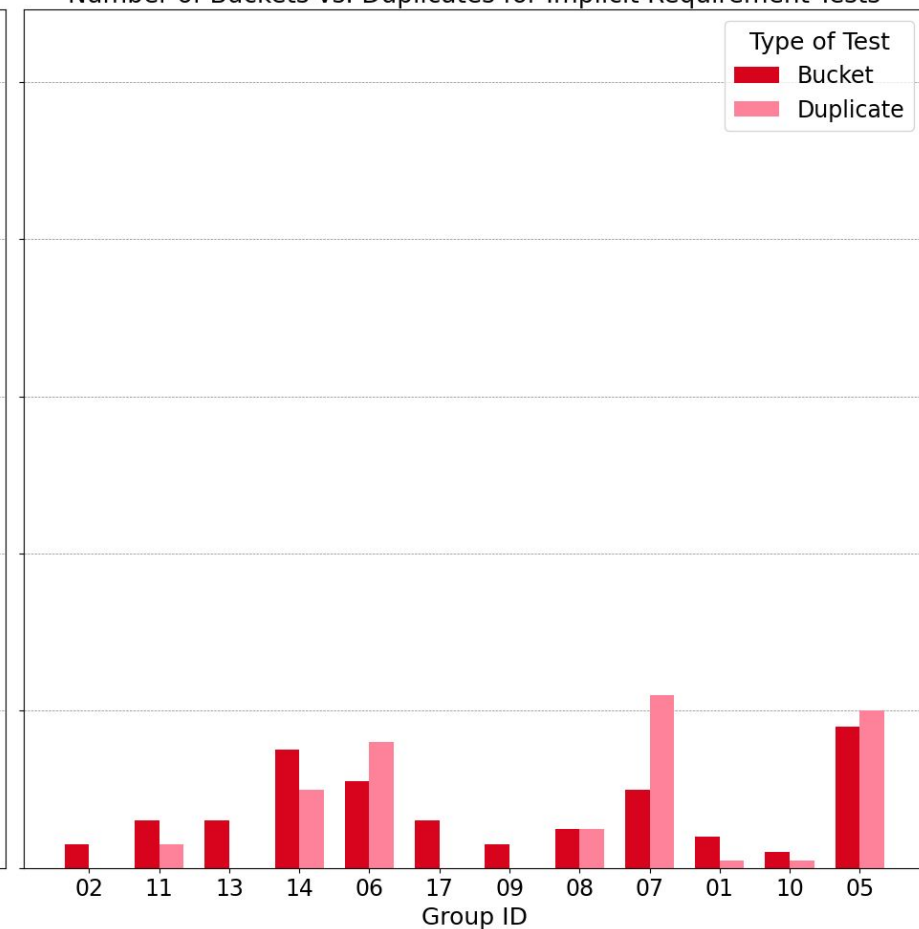***Students test invalid inputs more efficiently than
happy path inputs ($P < 0.0001$)

12

Number of Buckets vs. Duplicates for Explicit Requirement Tests

Number of Buckets vs. Duplicates for Implicit Requirement Tests

***Students test implicit requirements more efficiently than explicit (P = 0.0002)

13

% of Buckets Covered
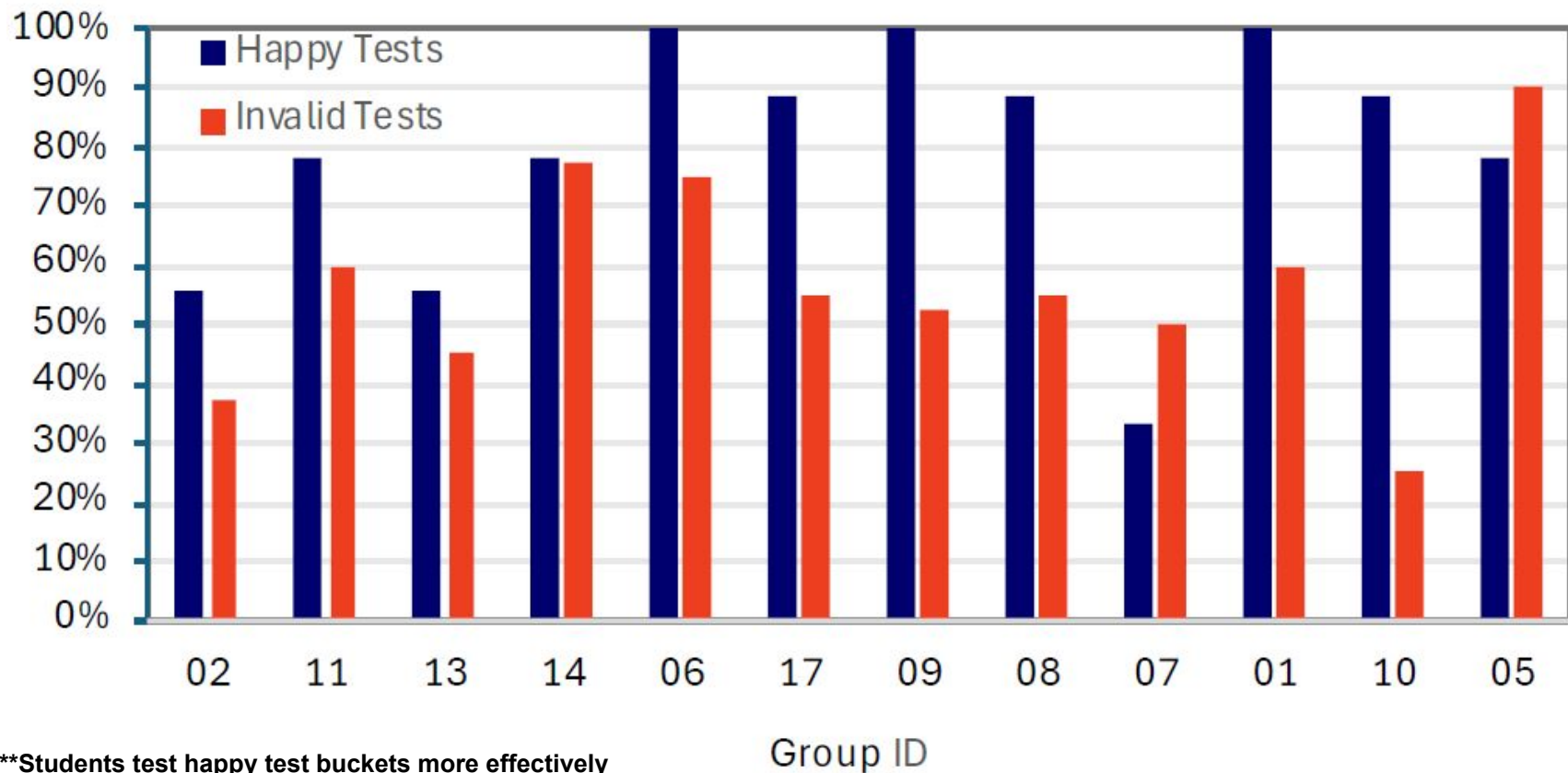
***Students test happy test buckets more effectively
than invalid tests (*P* = 0.0092)

Discussion

What types of tests do students test **inefficiently**?
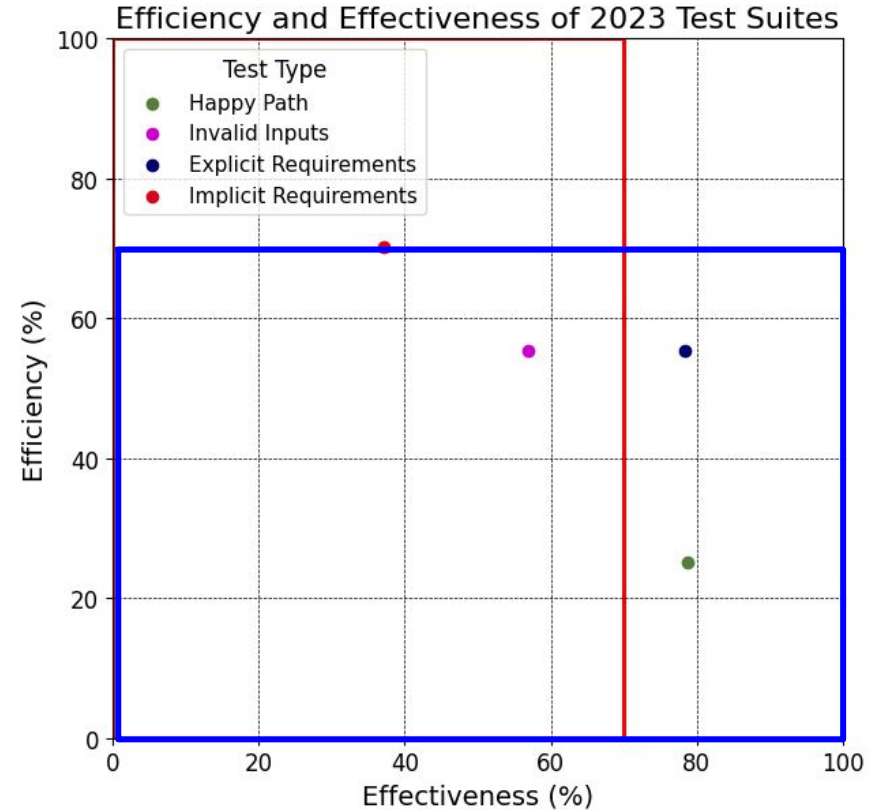
What types of tests do students test **ineffectively**?

Discussion

What types of tests do students test **inefficiently**?

What types of tests do students test **ineffectively**?

Efficiency and Effectiveness of 2023 Test Suites

Discussion

What types of tests do students test **inefficiently**?

What types of tests do students test **ineffectively**?

Efficiency and Effectiveness of 2023 Test Suites
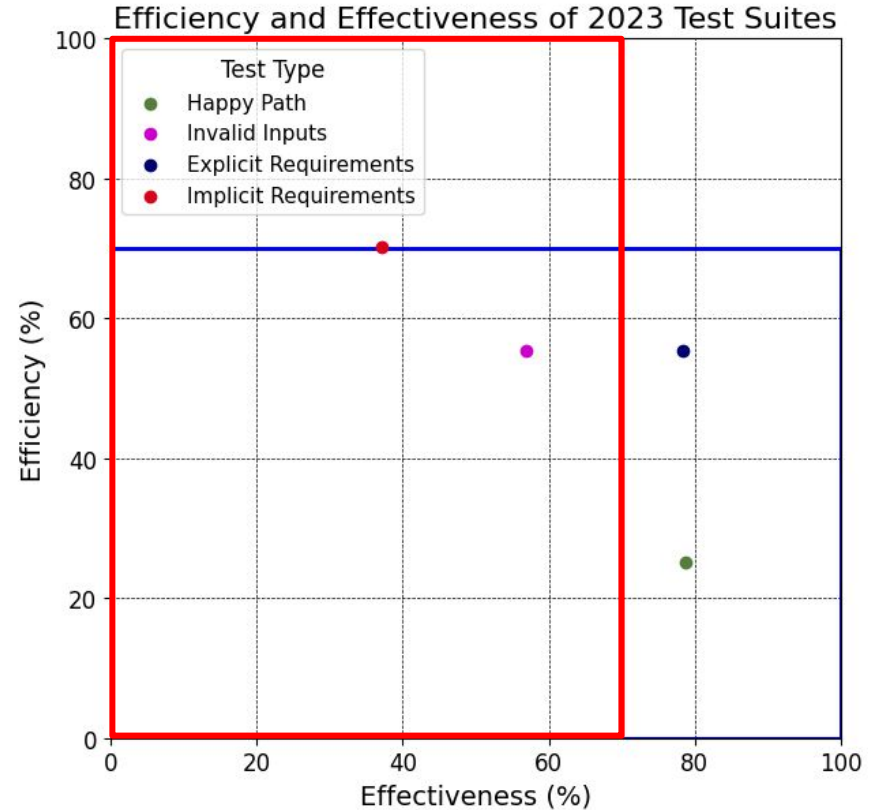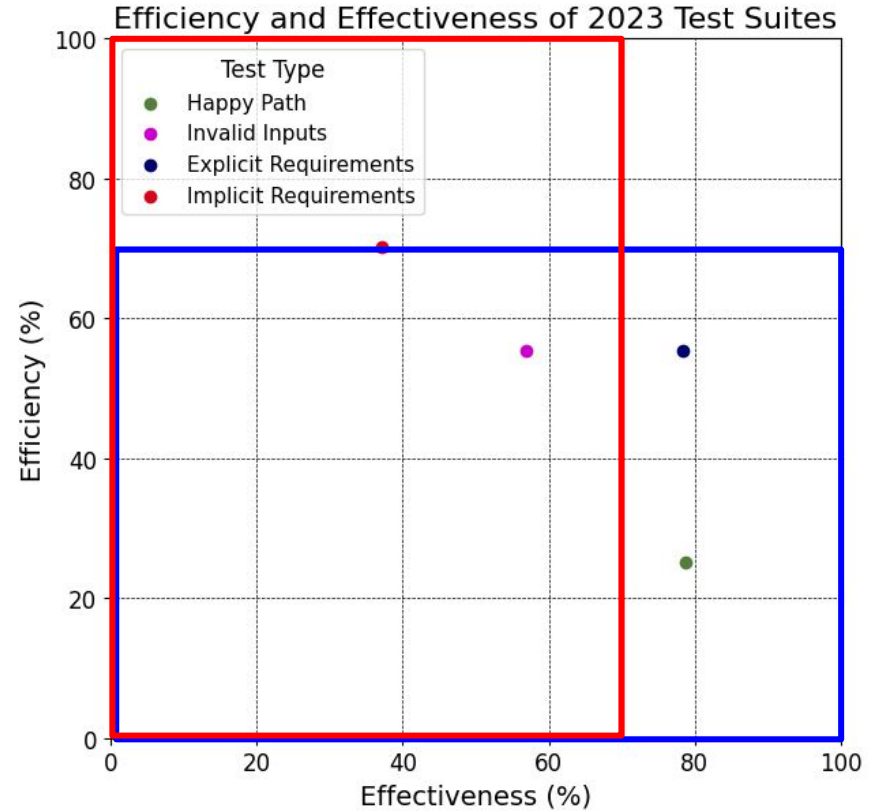
Contributions

- Analysis of students' **Black Box** tests rather than **White box** tests
- Insights valuable for **instructors** to address **knowledge gaps**
  - Students have difficulty writing tests for **implicit requirements** and reducing **redundancy**
- System to **categorize** student input files for the SQA project

Future Work

- Create public **repository** for researchers with tests from future instances of SQA courses
- Provide **feedback** to students using the automated system
- Expand the study to include other **types** of **test** (e.g. white box unit tests)

Test Suite Efficiency and Effectiveness

- Test Bucket
  - Group of related test cases for a specific requirement
- Inefficiency
  - High test redundancy
- Ineffectiveness
  - Low variation in buckets tested

(Quadrant diagram: Redundancy (vertical, low to high) vs. Variation in Buckets Tested (horizontal, low to high))
- high redundancy, low variation: Ineffective Ineffectient (red)
- high redundancy, high variation: Effective Inefficient (yellow)
- low redundancy, low variation: Ineffective Efficient (yellow)
- low redundancy, high variation: Effective Efficient (green)

Number of Buckets vs. Duplicates for Explicit Requirement Tests
Type of Test: Bucket, Duplicate

Number of Buckets vs. Duplicates for Implicit Requirement Tests
Type of Test: Bucket, Duplicate

***Students test implicit requirements more efficiently than explicit (*P* = 0.0002)

12

Number of Buckets vs. Duplicates for Happy Path Tests
Type of Test: Bucket, Duplicate

Number of Buckets vs. Duplicates for Invalid Input Test
Type of Test: Bucket, Duplicate

***Students test invalid inputs more efficiently than happy path inputs (*P* < 0.0001)

14

Discussion

What types of tests do students test **inefficiently**?
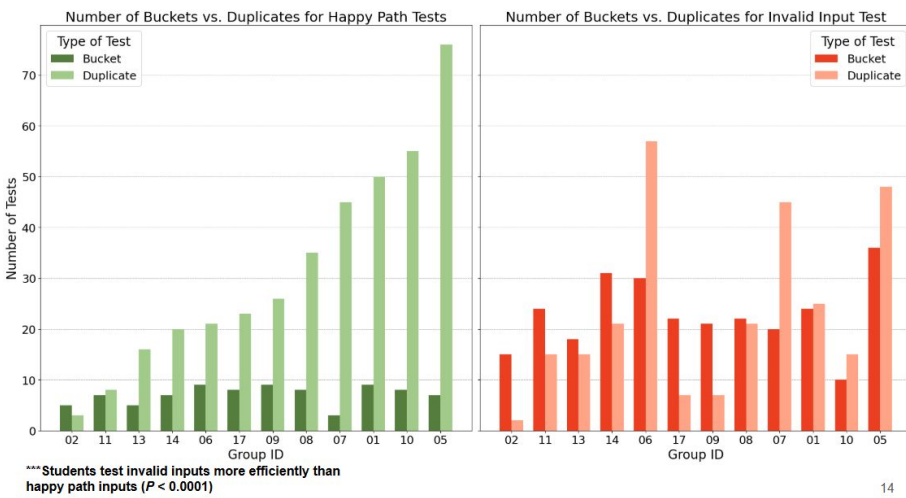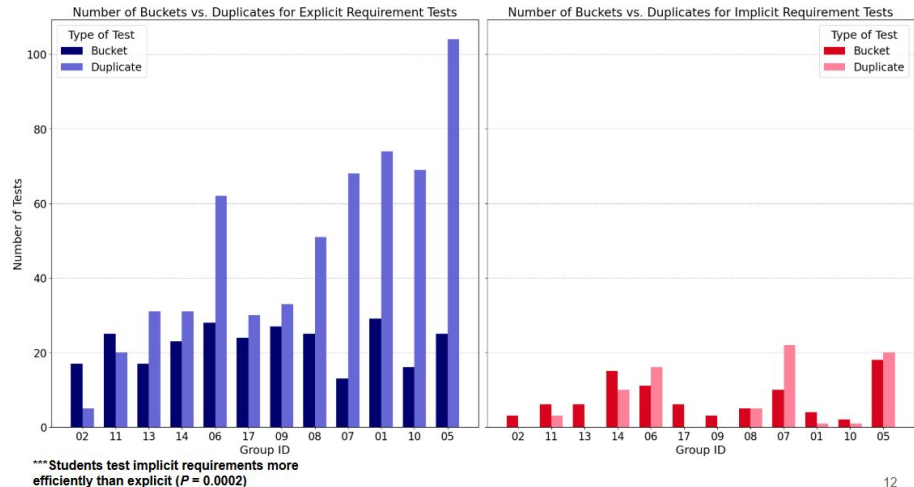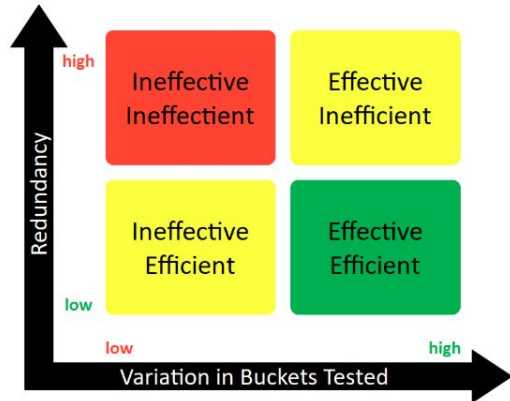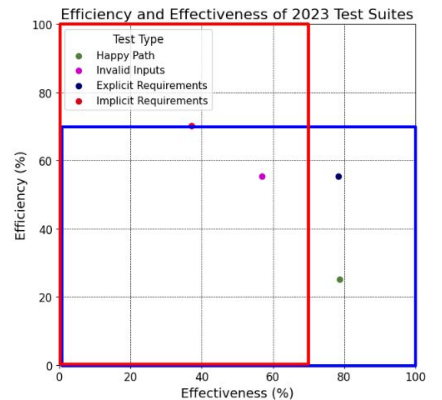
What types of tests do students test **ineffectively**?

Efficiency and Effectiveness of 2023 Test Suites
Test Type: Happy Path, Invalid Inputs, Explicit Requirements, Implicit Requirements

16

How Effective and Efficient Are Student-Written Software Tests?

Amanda Showler, <u>Michael A. Miljanovic</u>, Jeremy S. Bradbury

Software Engineering & Education Research Lab
Ontario Tech University, Oshawa, Canada
{amanda.showler, michael.miljanovic}@ontariotechu.net,
jeremy.bradbury@ontariotechu.ca

http://www.seerlab.ca

SEER LAB | Ontario Tech UNIVERSITY

SIGCSE
Technical Symposium 2025