Evaluation of Student-Written Test Suites

Amanda Showler
Supervisor: Dr. Michael Miljanovic

Problem & Motivation

- Recent CS graduates have insufficient testing skills [1]
- Findings from Edwards & Shams [1]
  - Student tests had low bug detection
  - Likely to only test valid inputs
- Identify the weaknesses of student-written test suites

[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.

Contributions

- Insights valuable for instructors to address knowledge gaps
- System to categorize student input files

Test Suite Effectiveness and Efficiency

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

4
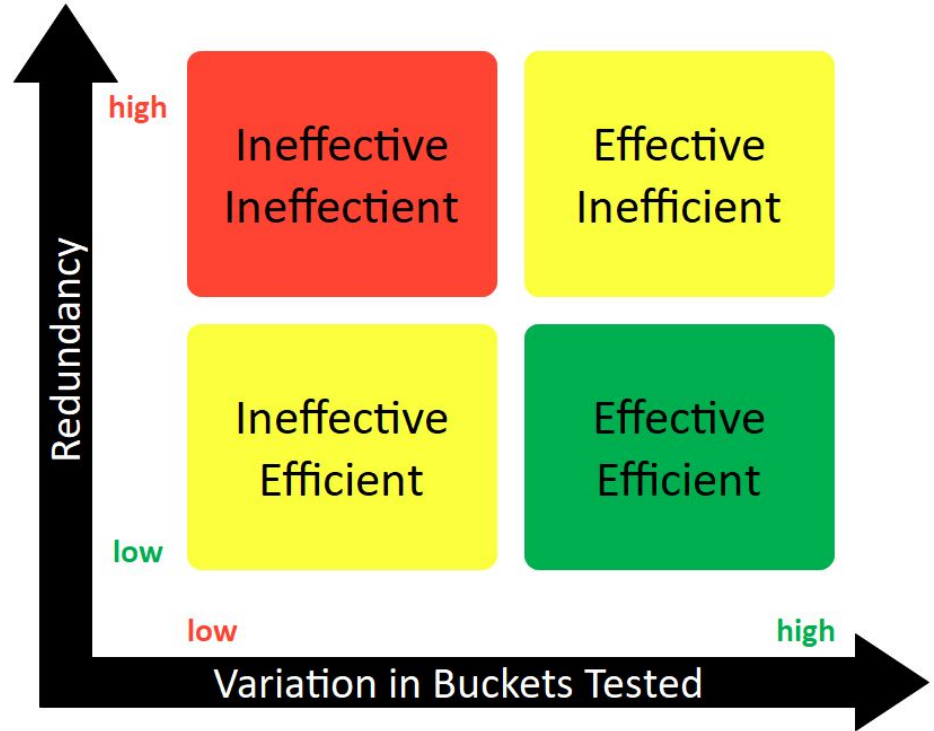
Test Type

1. Requirement Type
   - <u>Explicit:</u> Maximum Deposit is $1000
   - <u>Implicit:</u> Deposit cannot be negative

2. Input Type
   - <u>Happy Path:</u> $0 **<** Deposit **<=** $1000
   - <u>Invalid Inputs:</u> Deposit **>** $1000

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

Research Questions

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

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

SQA Course Project

- Students develop system and test suite based on program specifications
- Test suites are GitHub repositories from the CSCI 3060 SQA project

```
Welcome to the Online Auction Sales Service!

Please enter login to Login.
login - Login
login

***LOGIN***
Enter Username: testFSUser

Enter Password: password

Welcome Full-Standard User: testFSUser!

***MENU***
logout - Logout
advertise - Put up item for Auction
bid - Make a bid on an available item
addcredit - Add Credit
accounts - Display all active accounts
listitems - List Items Up For Auction
logout
```
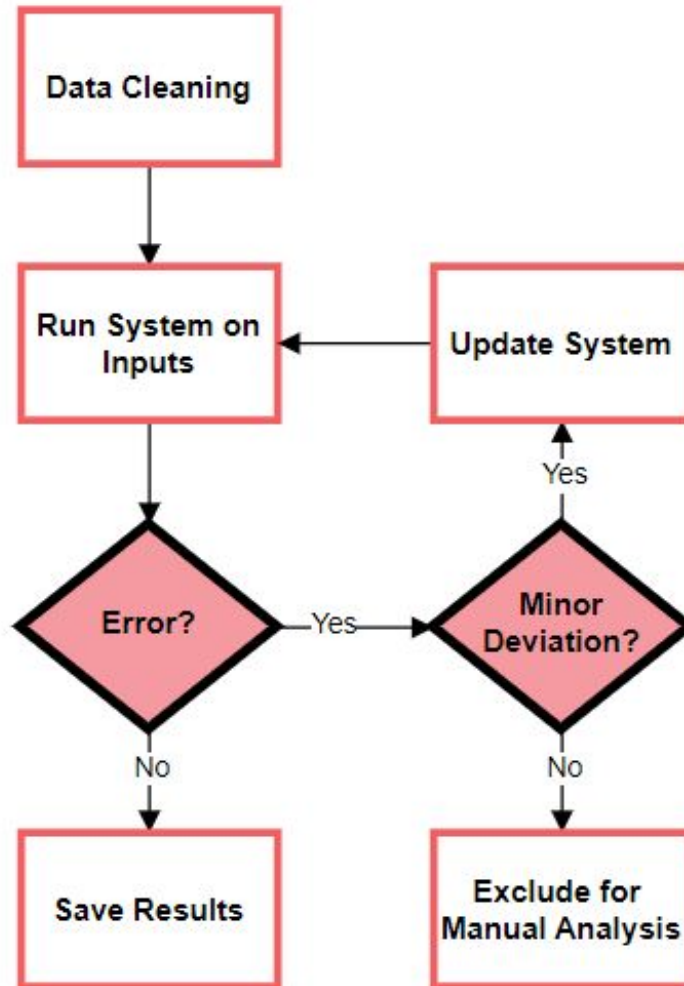
System

Data Exclusion Criteria

- Group's test suite deviated too much from expected format
- Test bucket was missed by the class
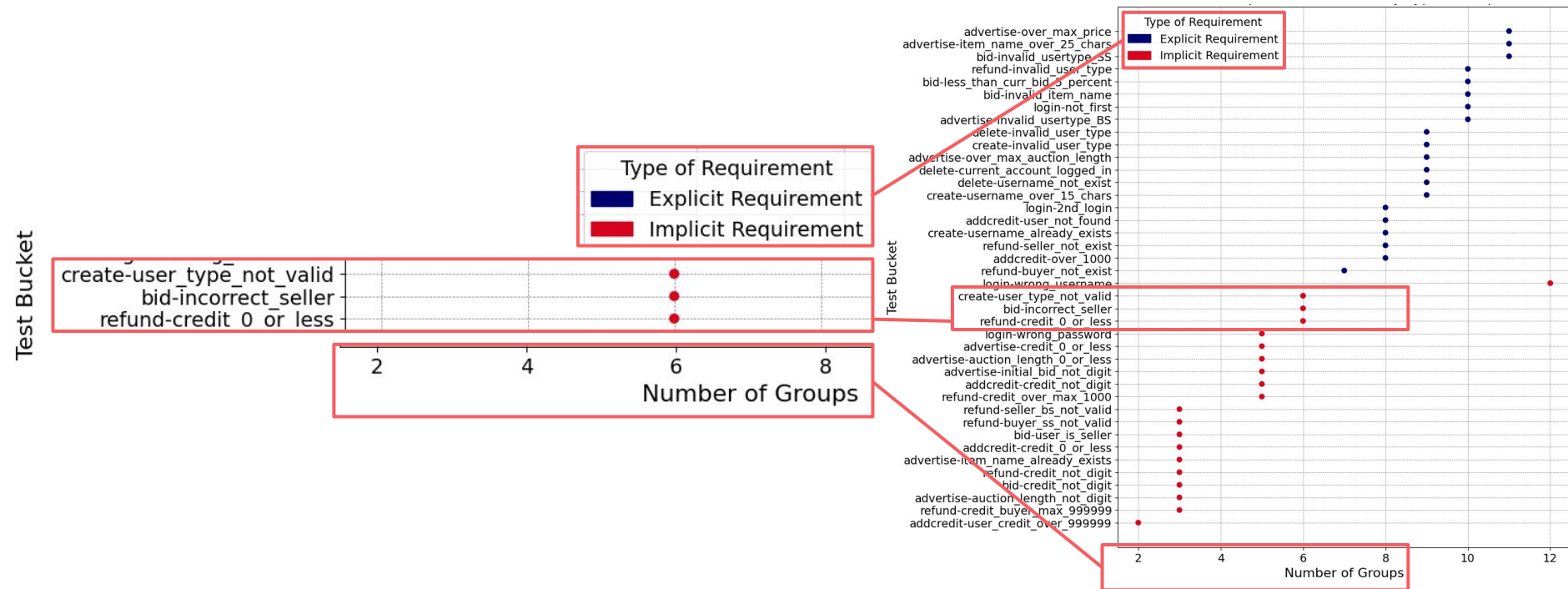- Test bucket is only tested by 1 group

Results

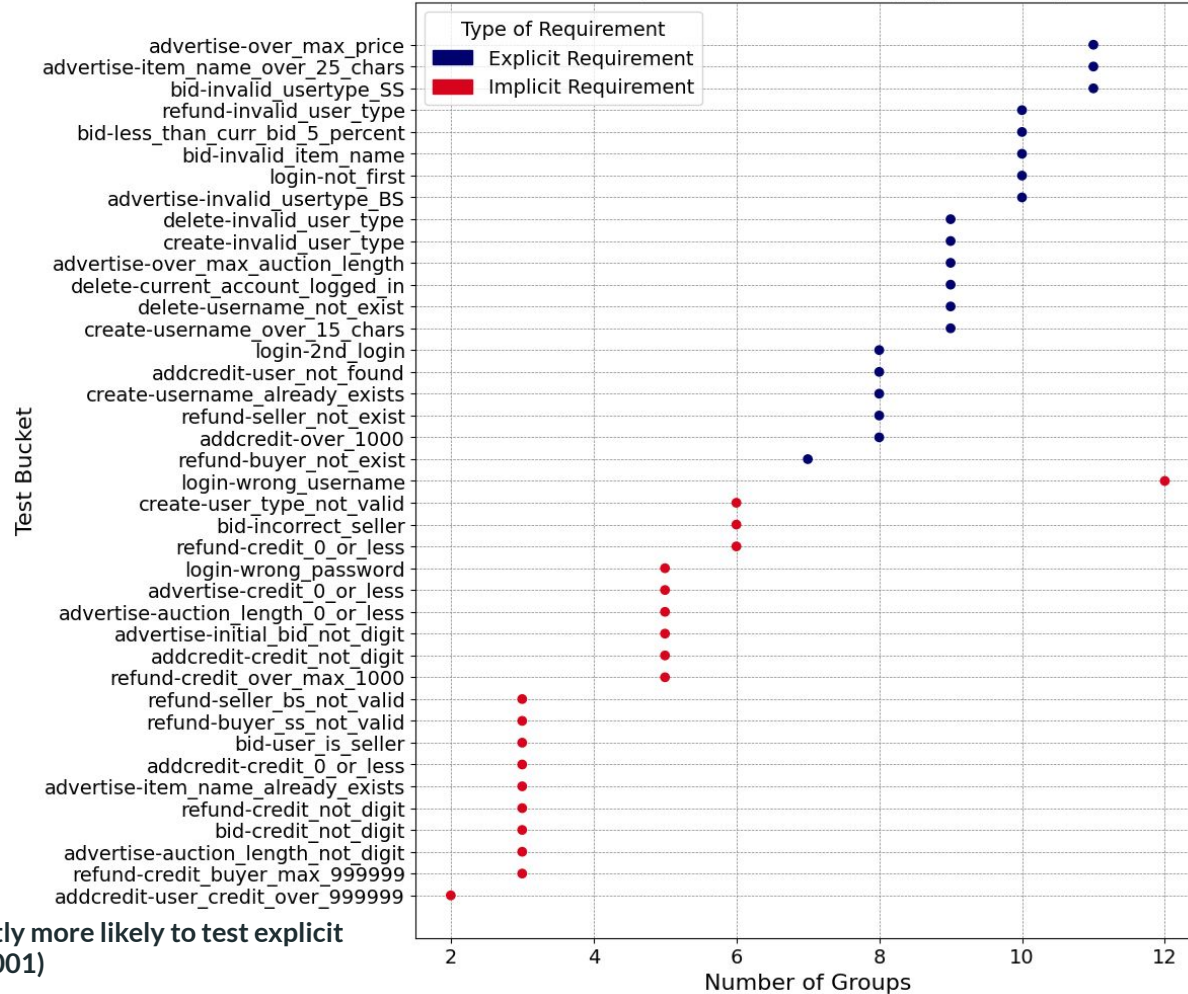| Winter 2023	Winter 2024
Auction Bidding System ● 12/17 test suites ● 49 test buckets ● 1014 tests	Steam Game Engine System ● 6/15 test suites ● 36 test buckets ● 381 tests

Results

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

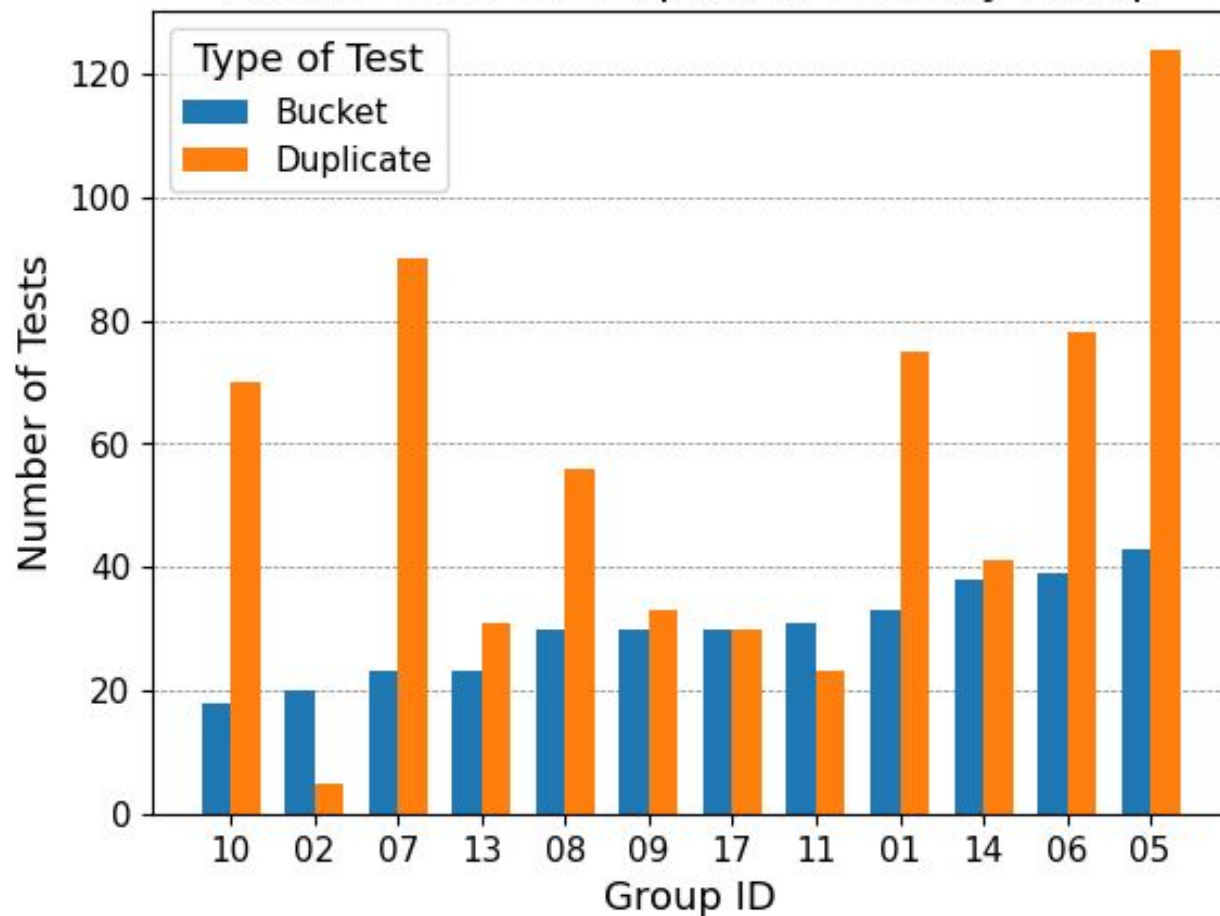# Number of Groups vs. Test Bucket by Type of Requirement

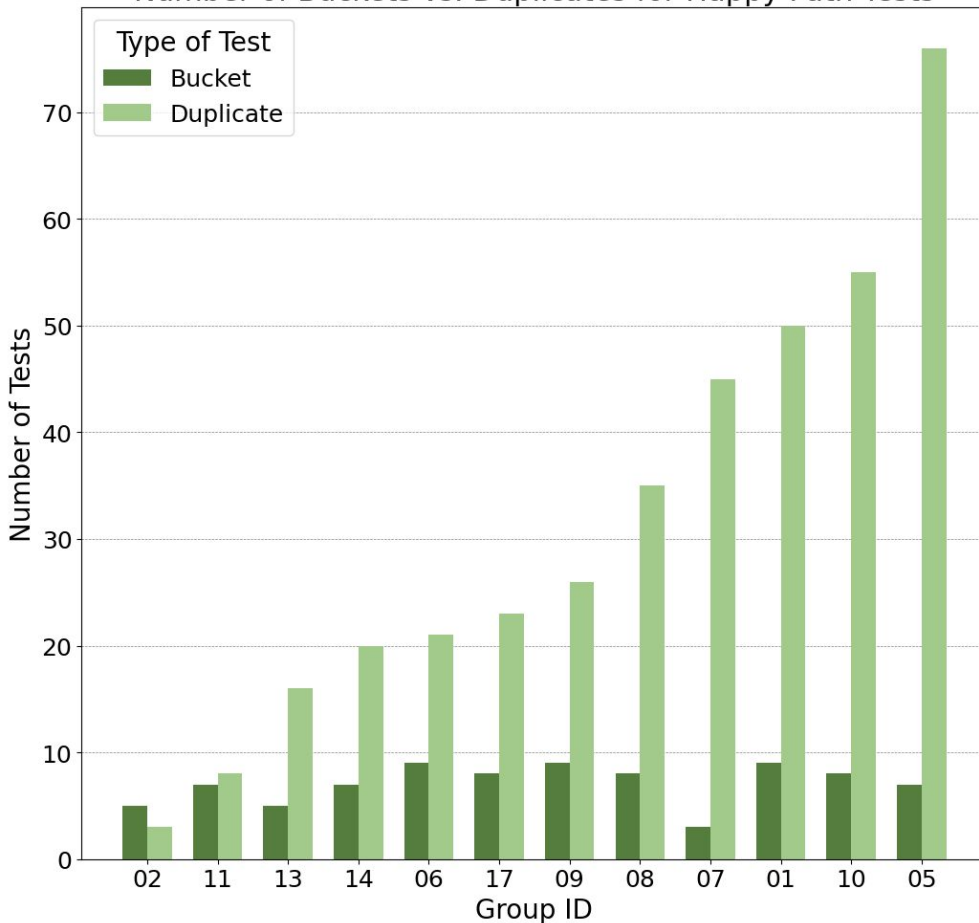Number of Groups vs. Test Bucket by Type of Requirement

***Students significantly more likely to test explicit requirements (P = 0.0001)
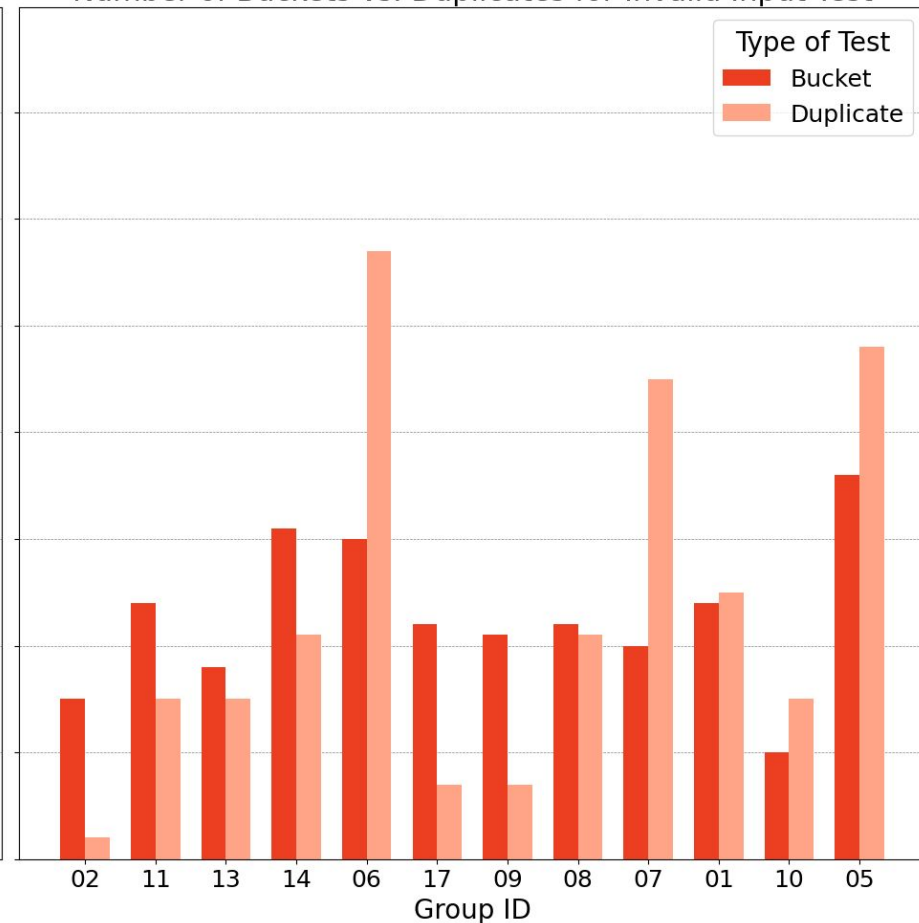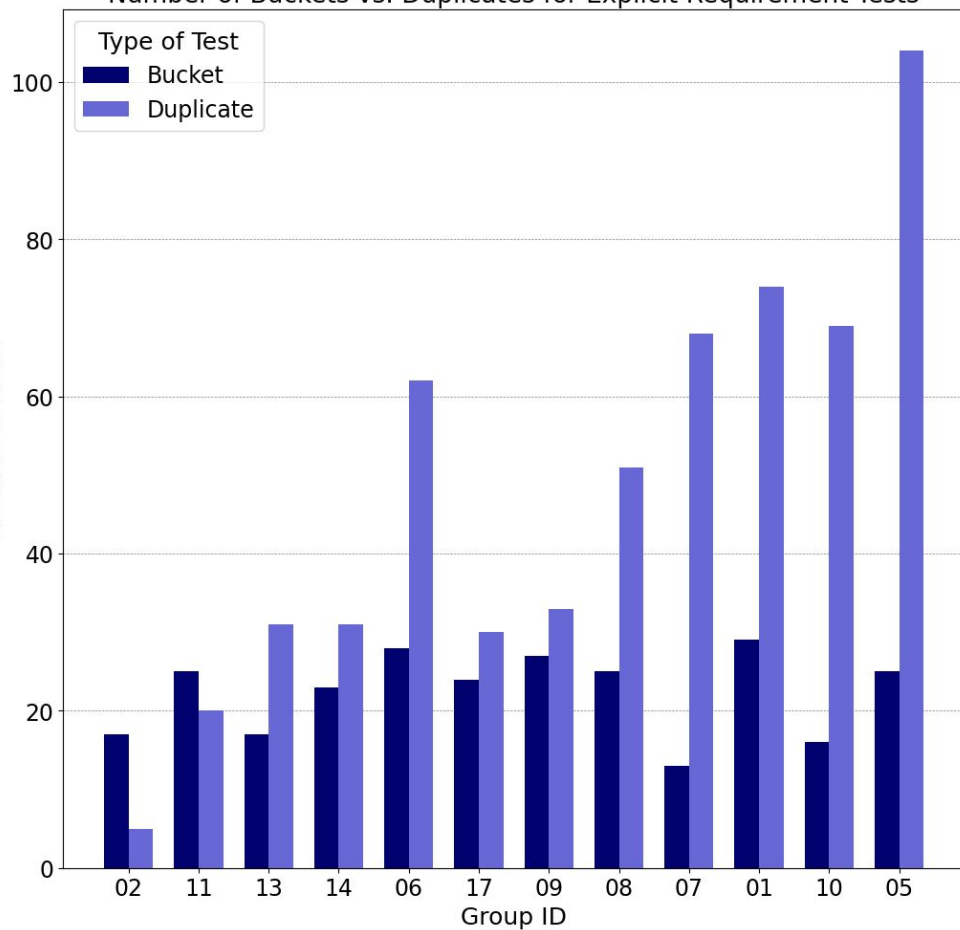
Bucket Tests vs. Duplicate Tests by Group

Number of Buckets vs. Duplicates for Happy Path Tests

Number of Buckets vs. Duplicates for Invalid Input Test

Number of Buckets vs. Duplicates for Explicit Requirement Tests

Number of Buckets vs. Duplicates for Implicit Requirement Tests
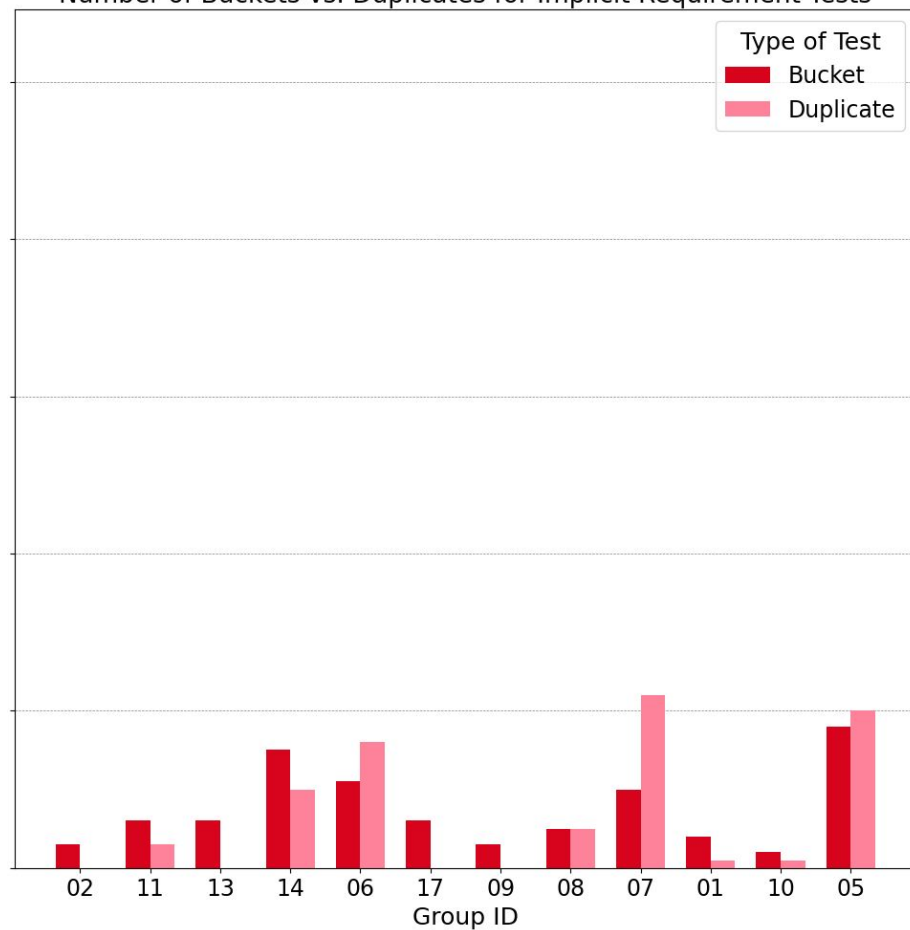
Discussion

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

- Happy path tests
- Invalid input tests
- Explicit requirement tests
- Implicit requirement tests

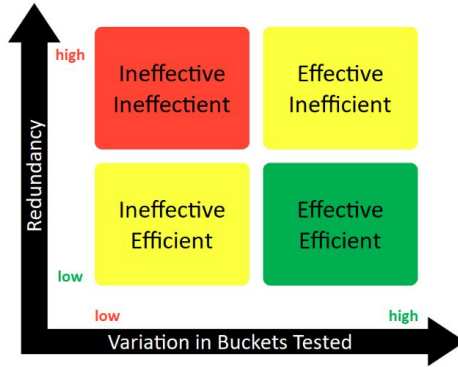What types of tests do students test **ineffectively**?

- Implicit requirement tests

Future Work

- Evaluating the excluded test suites manually
- Provide feedback to students using the system
- Investigate test suite bug detection

Test Suite Effectiveness and Efficiency

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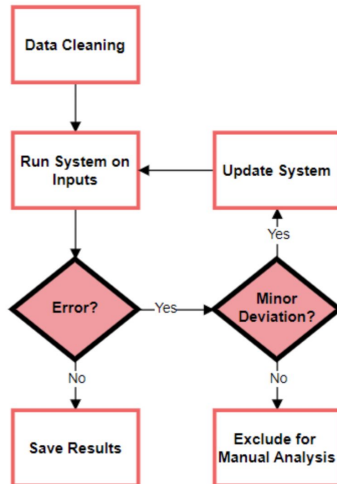- Inefficiency
  - High test redundancy

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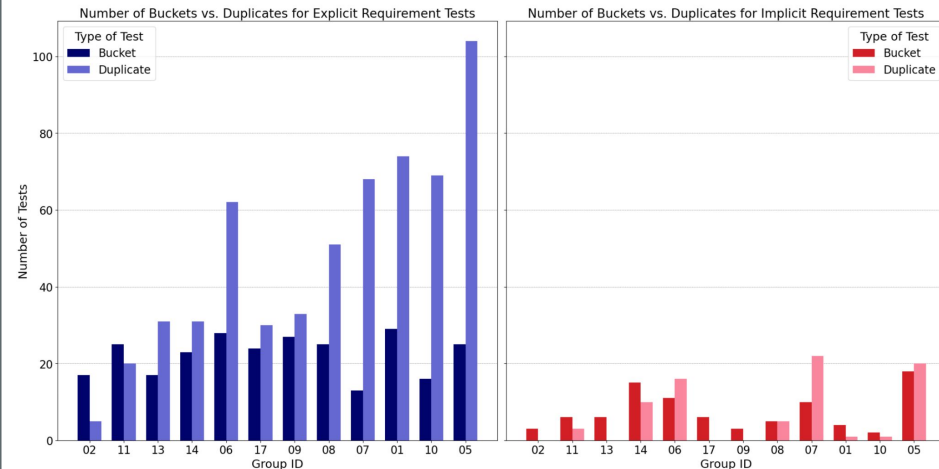- Effective Efficient (green, low redundancy, high variation)

Research Questions

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

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

System

Flowchart: Data Cleaning → Run System on Inputs → Error? — Yes → Minor Deviation? — Yes → Update System (back to Run System on Inputs); Minor Deviation? — No → Exclude for Manual Analysis; Error? — No → Save Results

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)