



Undergraduate Honors Theses

2020-12-05

Creating Statistical Testing Tools in Digital Testing Environments: Eliminating Confirmation Bias and Improving Pedagogy in the Humanities Testing Lab at BYU

Robert Hatch
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/studentpub_uht



Part of the [Teacher Education and Professional Development Commons](#)

BYU ScholarsArchive Citation

Hatch, Robert, "Creating Statistical Testing Tools in Digital Testing Environments: Eliminating Confirmation Bias and Improving Pedagogy in the Humanities Testing Lab at BYU" (2020). *Undergraduate Honors Theses*. 165.

https://scholarsarchive.byu.edu/studentpub_uht/165

This Honors Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Honors Thesis

CREATING STATISTICAL TESTING TOOLS IN DIGITAL TESTING ENVIRONMENTS:
ELIMINATING CONFIRMATION BIAS AND IMPROVING PEDAGOGY
IN THE HUMANITIES TESTING LAB AT BYU

by

Robert Hatch

Submitted to Brigham Young University in partial fulfillment
of graduation requirements for University Honors

Humanities Department
Brigham Young University
December 2020

Advisor: Devin Asay

Honors Coordinators: Gregory Stallings / Rex Nielson

ABSTRACT

CREATING STATISTICAL TESTING TOOLS IN DIGITAL TESTING ENVIRONMENTS:
ELIMINATING CONFIRMATION BIAS AND IMPROVING PEDAGOGY
IN THE HUMANITIES TESTING LAB AT BYU

Robert Hatch

Humanities Department

Bachelor of Spanish Teaching

The Humanities Testing Lab at BYU is a computer testing environment for humanities students and professors. Statistics are useful in a pedagogical sense because they present unbiased information about the progress of the class. The Humanities Testing Lab did not previously have a statistical tool, making it the perfect place to monitor the change in teacher perspective when a statistical tool was introduced. In an effort to better the BYU Humanities Testing Lab, an ideal statistical tool was researched, created, and delivered to participating professors. Their interactions with the tool were measured through a set of before and after surveys.

Acknowledgements

I would like to thank Russel Hansen for his involvement since the beginning of this project from brainstorming to programming. I would also like to thank the rest of my committee. Devin Asay was critical in setting up the interactions with the focus group. Gregory Stallings, who started this adventure with us, and Rex Nielson who brought us across the finish line. I am giving a big shout out to the professors who volunteered to test the final product and for Casey Griffiths who provided inspiration by sharing his process. My wife deserves thanks of course for listening to me blather until I made sense. A final thanks to the folks in the Honors Office for answering each and every one of my questions.

Table of Contents

Title.....i
Abstract.....ii
Acknowledgments.....iii
Table of Contents.....iv
List of Tables and Figures.....v

I.Introduction.....1
II.Research.....3
III.Design.....17
IV.Results.....23
Works Consulted.....25

Appendix A: Coding.....26
Appendix B: Surveys and Responses.....33
Appendix C: Extended Tables.....35

List of Tables and Figures

Research

Fig. 1.1 Canvas Quiz Summary

Fig. 1.2 Canvas Question Analysis

Table 1 Canvas Export

Fig 2.1 Learning Suite Exam Summary

Fig 2.2 Learning Suite Question Detail

Fig 3.1 Blooket Overview

Fig 3.2 Blooket Question Detail

Design

Fig 4.1 HLAB Statistical Tool Blank

Fig 4.2 HLAB Statistical Tool Information

Fig 4.3 HLAB Statistical Tool Uploads

Fig 4.4 HLAB Statistical Tool Filled

Fig 4.5 HLAB Statistical Tool Exam Overview

Fig 4.6 HLAB Statistical Tool Details

Appendix C

Table 2 Extended Canvas Export

I Introduction

One of the most controversial topics in the pedagogical world is the effects of standardized testing on the time spent in class. Under current laws, schools receive funding based on the scores their students receive from standardized tests such as the ACT or the SAT. The thinking behind this system is that each student in the nation should be at the same educational level and have a common baseline of skills upon high school graduation. While this is a noble goal, the controversy comes from the teachers asked to effectuate it. Getting numbers to agree with each other across the nation is easy, but it is becoming ever more evident that getting people to agree with each other is not. Each teacher has their own methods, goals, and benchmarks. What more, each student has their own expectations for a class. Many teachers feel they are being asked to replace time spent on what they believe to be important with time spent on what someone else believes to be important. Others believe that this form of learning isn't long lasting for students. Often students will ask "Will this be on the test?" If the teacher answers yes, the student pays attention out of obligation. If the teacher answers no, the student pays no attention at all. This plight is too large to address in just one project, but there is an aspect of it that can be improved regardless of the situation.

Despite the uproar, tests have an important place in each teacher's classroom. The relationship between student and test was originally designed to be helpful to everyone. The original purpose of testing a student was to ensure they had mastered a skill and could perform it without constant supervision and aid. A recent improvement in testing environments is the inclusion of computers. The Humanities Testing Lab at BYU has

Creating Statistical Testing Tools

options for tests to include video and audio clips, and for students to submit audio clip responses. This is especially useful for language learning and evaluation.

Another benefit of using computers is the elimination of the bias found in the subconscious. When conducting research there is an issue known as confirmation bias. Researchers who suffer from confirmation bias only accept results they expect while coming up with reasons to reject results they didn't expect. Teachers unknowingly suffer from both confirmation bias and regular subconscious bias.

There are many things that stand in the way of accurately measuring and portraying the progress of a group of students. Computers have come to be very helpful in combatting the issues associated with testing, both in the way tests are presented to students and in the way results are presented to teachers. Perhaps one of the most useful tools a computer can provide is an aggregation of student data. Whether the grades come from one test or the overall course, a lack of accurate statistical data inhibits learning because the teacher doesn't know that their students don't understand the subject matter. A computer, however, can present averages, trends, outliers and other information without error. We simply have to ask.

In an effort to better the BYU Humanities Testing Lab, a statistical tool was researched, created, and delivered to participating professors. Their interactions with the tool were measured through a set of before and after surveys. The following findings will present the process of the creation and implementation of the tool as well as a rundown of professor responses.

II Research

The research for this tool was experiential. Many statistical tools were used and some of their best suited features were selected to be integrated into the HLAB (Humanities Testing Lab) Tool. In this section, each researched tool and selected feature will be explained.

Statistical Tool 1: Canvas

The first, and probably most well-known statistical tool, is the suite of statistics that come integrated with Canvas. Canvas is an online classroom environment which presents assignments and grades to students and can also administer tests and grade them. It is a favorite of teachers in virtual classroom environments. Besides giving options to moderate tests and grade assignments digitally, it also presents detailed statistics of each test. (Fig.1.1)

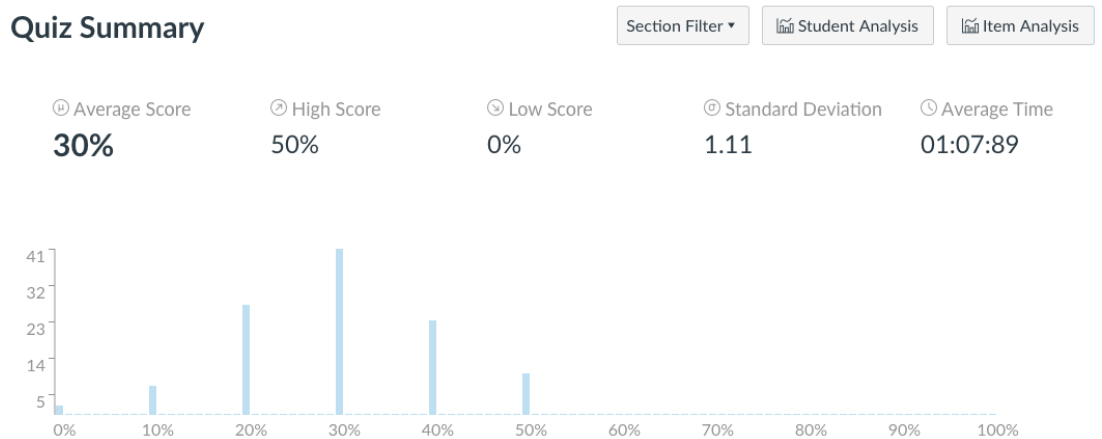


Fig. 1.1 Canvas Quiz Summary – Seen here are the test results for a pretest. On the left are the number of students who achieved each score. On the bottom are the scores. Also seen are the average score, as well as the highest and lowest scores, and the average time.

Creating Statistical Testing Tools

As you move down the page each question is displayed individually with its answers. (Fig. 1.2) Clicking on an answer will give you a list of students who selected that answer.

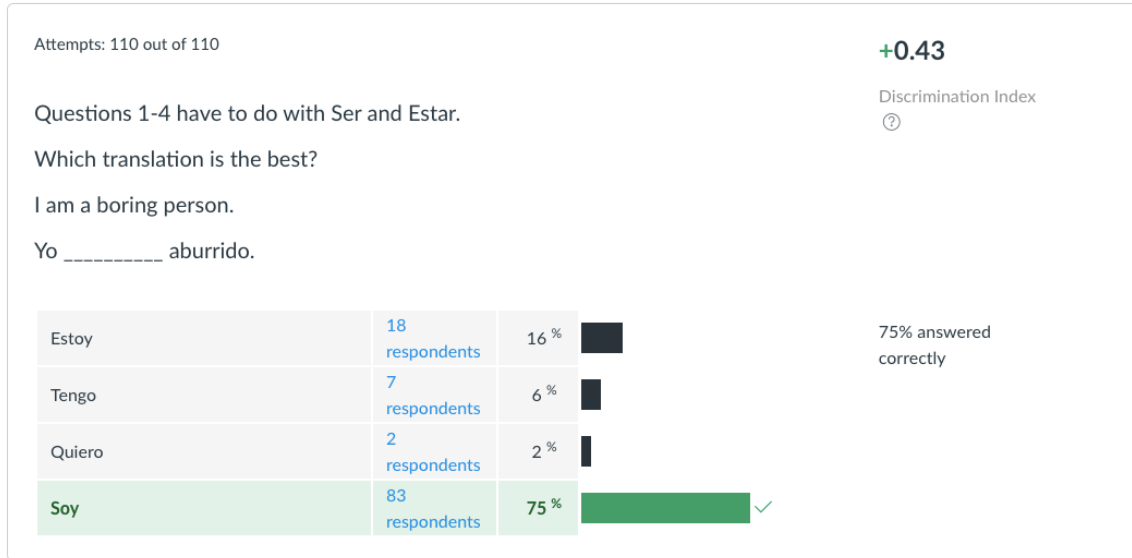


Fig 1.2 Canvas Question Analysis – Here the question is displayed along with the possible answers. Clicking on the respondents will give a list of students who selected that answer. The correct answer is in green.

This particular tool also gives the option to export results into an Excel spreadsheet in several different ways. (Table 1) Most of the numbers here are for the computer to use later as it does its calculations, but the correct and wrong column are useful for creating graphs directly from the spreadsheet.

Question Id	Answered	Top	Middle	Bottom	Quiz	Correct	Wrong
1514145	110	30	70	10	6	83	27
1514147	110	30	70	10	6	53	57
1514150	110	30	70	10	6	46	64
1514151	109	30	70	9	6	90	19
1514208	109	30	69	10	6	31	78
1514209	110	30	70	10	6	18	92

Table 1.1 Canvas Export – By default this creates an Excel Spreadsheet full of data. Most of the data is used by the computer when calculating, but many columns are useful. A more complex version of this chart can be found in Appendix C.

Creating Statistical Testing Tools

After researching this specific function it became apparent that useful statistical tools require much more data than is shown to the user. This extra data is sometimes referred to as metadata.

Canvas divides students into 3 skill levels shown here as top, middle, and bottom. The reason for this will be discussed later.

Statistical Tool 2: Learning Suite/Testing Center

BYU's website for students to access information on their courses is called Learning Suite. It is the official site for grades to be posted, and though it is still rough around the edges it has many uses. It is connected to the non-digital Testing Center, letting students see their scores instantly. It also compiles statistics for professors. This is how most BYU professors administer their tests and receive their statistics. Students fill in scantron papers which are afterwards scanned by a computer to create data for their results. Professors have access to the following statistical report. (Fig 2.1)

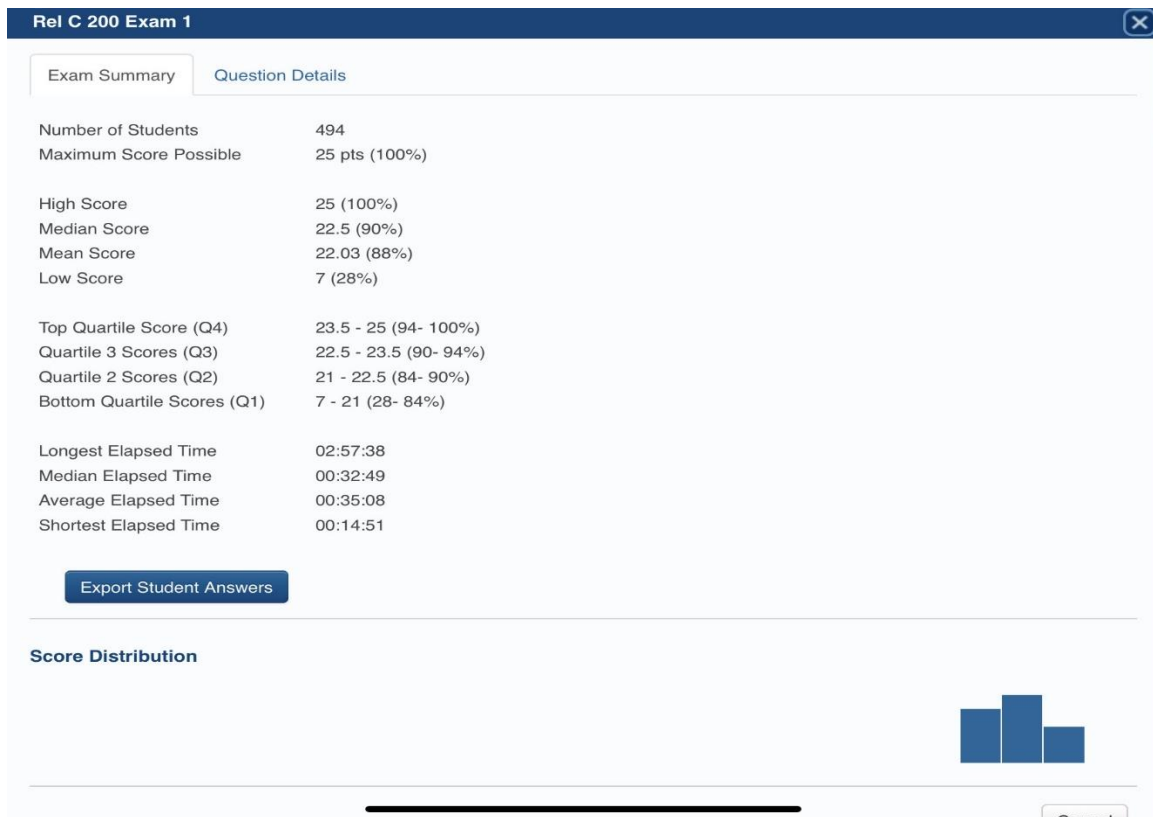


Fig. 2.1 Learning Suite Exam Summary – Statistics are presented along with a basic graph. Hovering over the bars of the graph will present more exact numbers on the website.

The main differences we see here are the inclusion of average time, shortest time, and median score. It should also be noted that here students are divided into 4 skill levels.

Creating Statistical Testing Tools

An individual question analysis is also available. (Fig. 2.2)

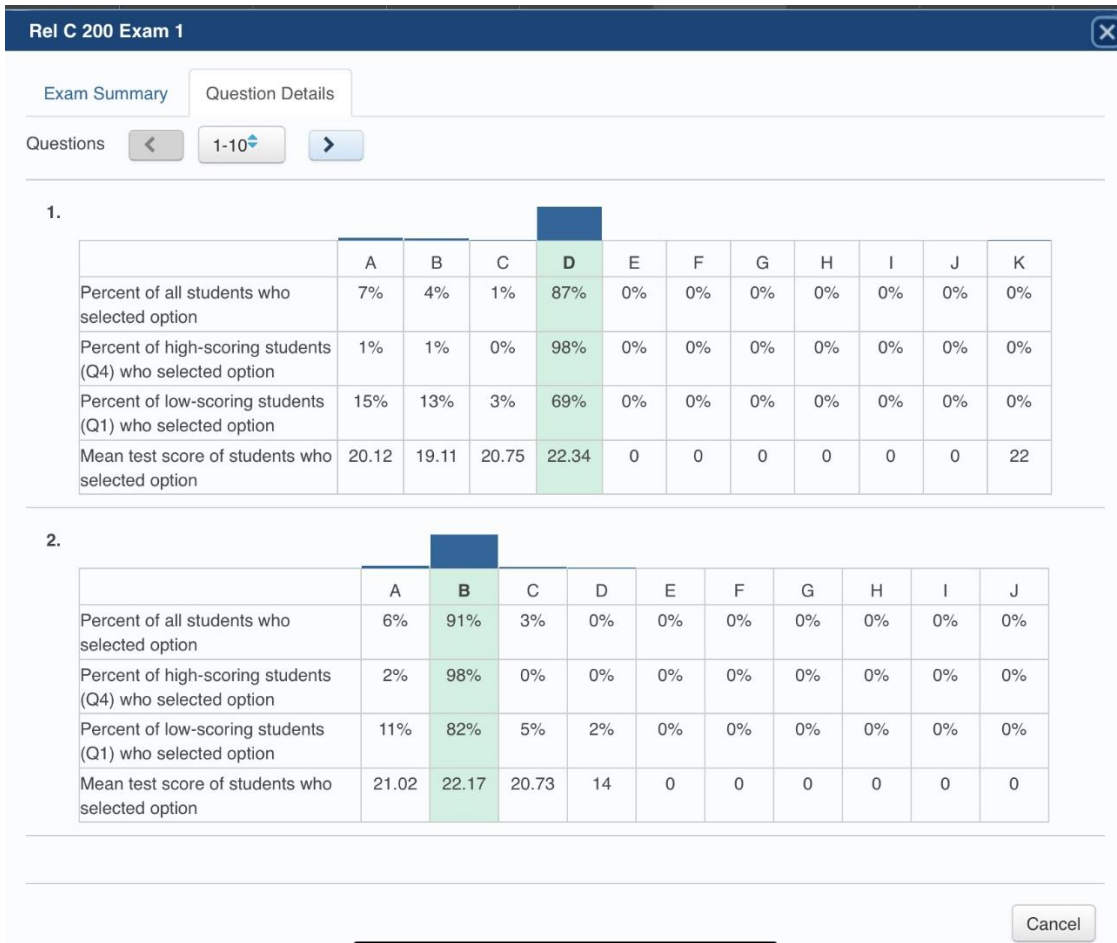


Fig 2.2 Learning Suite Question Detail – Green Columns indicate correct answers. Note that the final row is not a percentage.

Learning Suite also has the option to export data. It is not available here because it contains student names and is protected under FERPA. This statistical tool was something we wanted to emulate as it was what most professors on campus interacted with.

Other Statistical Tools

Other tools that were presented with formative assessments also have statistics. Examples include Kahoot, Quizlet, Gimkit, and Blooket. These statistics aren't recorded and disappear quickly, but they let the teacher know how the class understands as a whole. (Fig 3.1)

Specifically, it is common for these quiz games to present students with different questions in different orders. All questions come from the same set and are related, but not all are presented to each student.



Fig 3.1 Blooket Overview – The statistic here is the average score of every question asked to every student. Not every student had the same questions.

Creating Statistical Testing Tools

Even for temporary assessments like these, individual terms are analyzed. There is an option here to view responses in the order of most incorrect. (Fig. 3.2) Also presented on this page are student scores. On this particular website, students must also make strategic (ex. Power-ups, Multipliers) choices in a game-like setting, converting their scores into “money”. Students only have access to other student’s money scores, while teachers have access to all scores. Note once again, the option to export. This option costs extra.

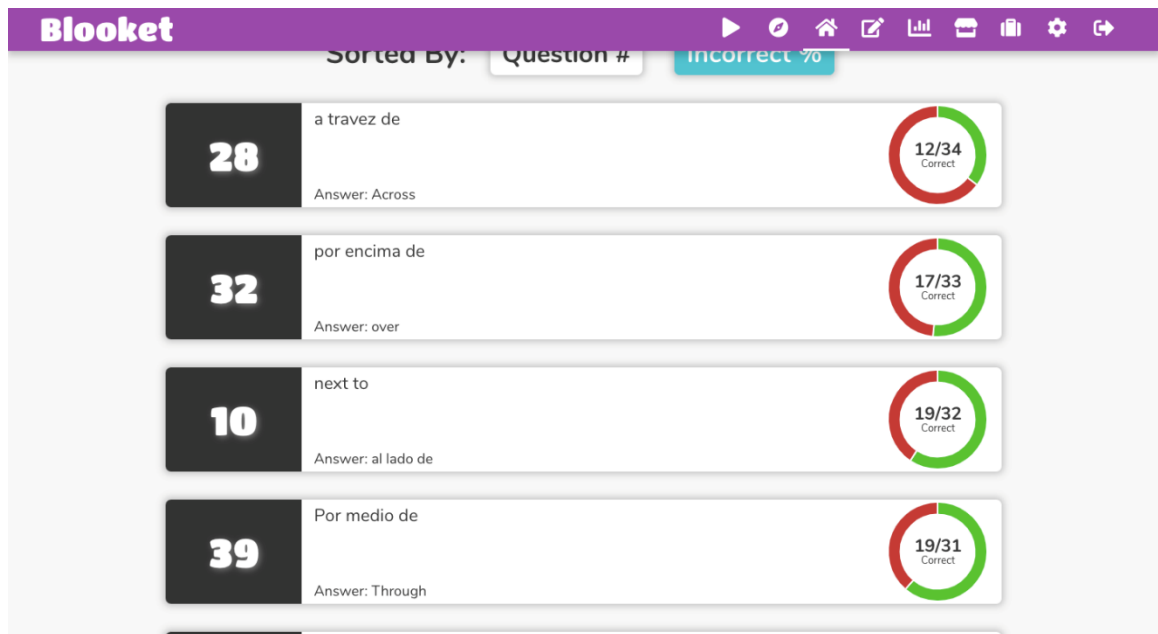


Fig. 3.2 Blooket Question Detail – Here presented are the terms most incorrectly scored by students. Not all students were presented with all questions.

The major insight gleaned from these quiz game websites was that most of them offered advanced statistics for a monthly fee. Without statistics, these were just games, and not very useful for teachers in the long run.

Creating Statistical Testing Tools

Selected Features

From the above tools, and conversations with involved professors, the following features were selected. They are listed here in order of appearance in the created tool. They also have definitions and additional information as to why they were chosen, or where they can be used in classroom planning. An important statistic to remember is 85%. It is considered a mastery skill level.

Exam Overview

These statistics were chosen from among those researched as the most relevant to BYU Professors using the Humanities Testing Lab. Not every statistic is important in every set of test results, but the indications that might come up are very helpful.

Highest Score – Displays the highest percentage score achieved by any student. If the score is 100%, then it is possible for a student to understand everything on the test. If no one scored 100% something on the test was difficult to understand, whether it be the subject matter or the wording of a specific question.

Average Score- Displays the average percentage score achieved by all students. If the average score is above 85%, most of the students understand most of the content. This is perhaps the most important statistic presented after any test.

Median Score- Displays the median score of all students. Whatever this score is, 50% of students are above it, and the other 50% are below it. On occasion, the average score will be above Mastery, but the Median score will be below mastery due to

Creating Statistical Testing Tools

outliers. In this case, the average score would be misleading when considering the success of all students.

Lowest Score- Displays the lowest percentage score achieved by any student.

Excluding scores of 0%, this is a good statistic to know how far away from understanding the subject matter your struggling students are.

Longest Time- Displays the longest time of any student in minutes. This statistic isn't usually the most prominent. It is useful however when you want to know if your tests are "rigorous". Most tests have a time limit. The closer any of the time statistics are to the time limit, the more challenging the test is.

Average Time- Displays the average time of all students in minutes. If the average time is close to the limit, consider extending the time limit. If there is a lot of excess time, consider shortening the time limit

Shortest Time- Displays the shortest time of any student in minutes. If the shortest time is very short, the test may be too easy or too short, depending on your goals. Keep in mind, times will vary more than scores, because even if students understand, they express understanding in different ways.

Most Missed Question- Displays the question(s) with the most incorrect answers.

Whichever question or questions have the highest number of incorrect responses indicates a subject that should be retaught. It is also an indicator that the question might be worded poorly, or the answer key may be incorrect.

Creating Statistical Testing Tools

All Correct- Displays questions that every student answered correctly. These questions may be too easy or entirely unnecessary. It is an indication of well-known subject matter.

Majority Missed- Displays questions that more than 50% of the class responded incorrectly on. Another indicator that subject matter needs more attention, or there is a formatting error.

Mastery Students- Displays the number of students who scored higher than 85% on the test. If this number is above 70% it is an indicator that there has been enough time spent on the subject matter of the test.

Non-Mastery Students- Displays the number of students who scored lower than 85% on the test. If this number is very high, consider spending more time on the subject matter of the test.

Question Detail

Splitting question results into different skill groups was a method that came up in the research. It is useful to know if a question was missed on part of the teacher or part of the student. Usually, if a student passed the test, they put in more effort. If they put in effort and missed a specific question, that specific question deserves some attention on the part of the teacher.

All Students- Displays the percentage of students who answered the selected question correctly, as well as the percentage of students who answered incorrectly. For each individual question this is an indicator of how well students understood the subject matter.

Mastery Students- Displays the percentage of students who scored higher than 85% who answered the selected question correctly, as well as the percentage of those students who answered incorrectly. Students with higher scores likely put in more effort to study or perhaps simply were in class more often. If most high effort students answered a question correctly, everything is how it should be. If most high effort students answered incorrectly it is an indicator that the subject should be revisited, or the question is poorly formatted.

Non-Mastery Students- Displays the percentage of students who scored lower than 85% who answered the selected question correctly, as well as the percentage of those students who answered incorrectly. Students who got lower scores were likely exposed to less subject matter. If a specific question received high scores from low effort students, the subject matter has been covered sufficiently.

It is best to use all of these statistics together instead of separately.

Also available is the option to view the overall statistics for just mastery students and just non-mastery students in order to amplify the effects of skill division.

III Design

This is a screenshot of the opening screen of the finished application. (Fig. 4.1)

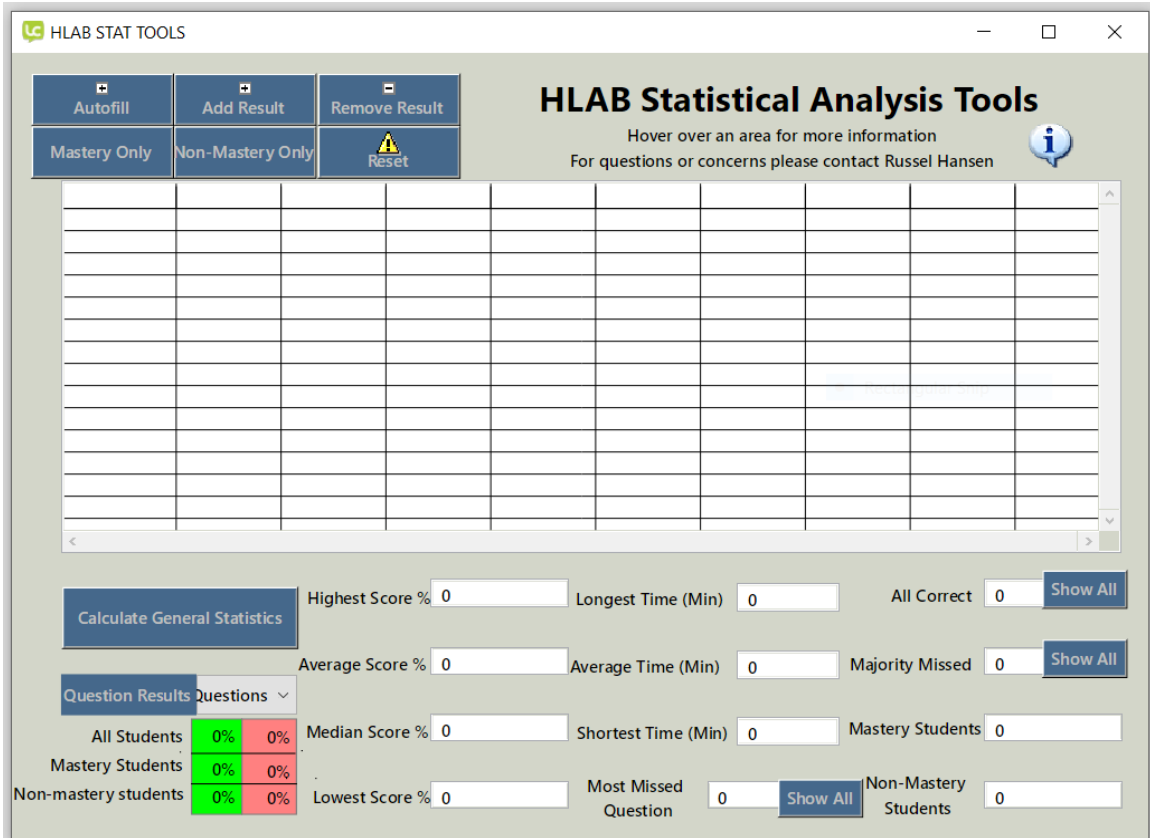


Fig 4.1 HLAB Tool Overview – The opening screen of the statistical tool. Everything is zeroed to ensure previous test results don't affect later test results.

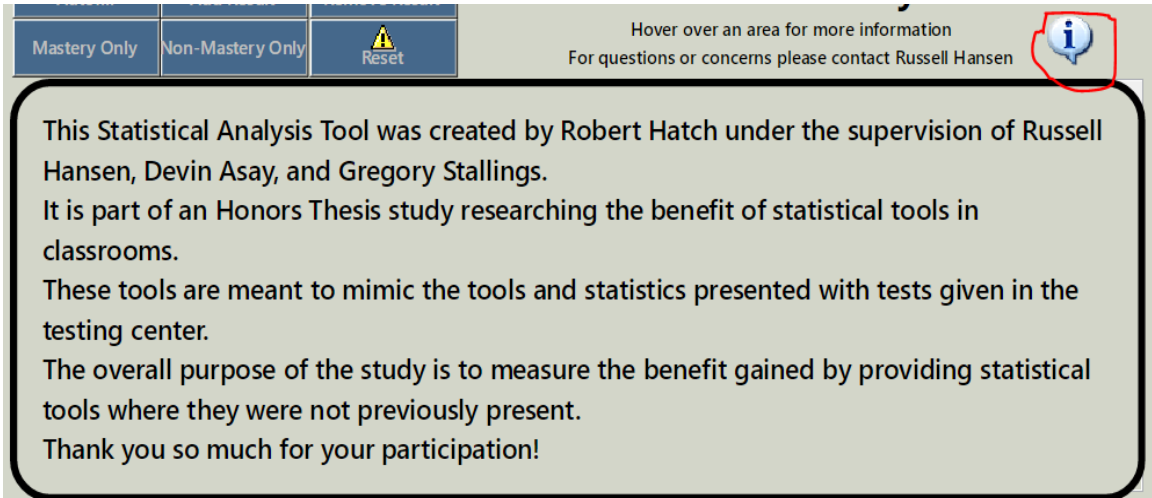
This application was created using the LiveCode programming language. While it is not as commonly used as other programming languages, it is the program used by the Humanities Testing Lab and is taught in a beginner and an advanced course at BYU as part of the Digital Humanities coursework.


Following are the functions of each of the buttons you can interact with in the application. Some of their coding can be found in Appendix A.


Information

The following will be displayed when the information bubble is clicked.

Fig. 4.2 HLAB Statistical Tool information – A brief informational section.



Mastery Only Non-Mastery Only  Reset

Hover over an area for more information
For questions or concerns please contact Russell Hansen 

This Statistical Analysis Tool was created by Robert Hatch under the supervision of Russell Hansen, Devin Asay, and Gregory Stallings.
It is part of an Honors Thesis study researching the benefit of statistical tools in classrooms.
These tools are meant to mimic the tools and statistics presented with tests given in the testing center.
The overall purpose of the study is to measure the benefit gained by providing statistical tools where they were not previously present.
Thank you so much for your participation!

Creating Statistical Testing Tools

Upload Area

Here is the main upload area.

(Fig. 4.2) Each button can be used to change which specific responses are being analyzed.

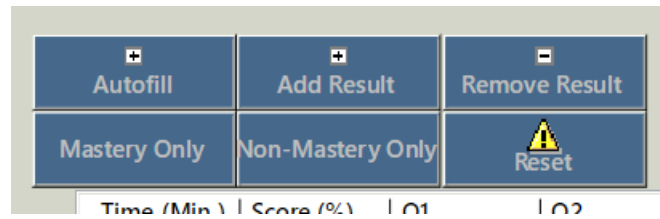


Fig. 4.3 HLAB Statistical Tool Uploads - This is the main user interface for the professors. Here results can be uploaded and edited as needed.

Autofill – Opens a screen to browse for a folder. The selected folder is then parsed and displayed, removing names of students. The statistics are automatically calculated and displayed from this information.

Add Result – Opens a screen to browse for a single file. The selected file is then parsed and displayed. The statistics are automatically updated and displayed from this information. Useful if a student took the test later than the others, or their results are found in a different place. It can also be used to duplicate results to see how many students need to pass in order to raise statistics to a desired goal.

Remove Result – Deletes one set of responses. If there is an error or an outlier this can help get more accurate statistics overall.

Mastery Only– Temporarily hides the results of students who did not score higher than 85%. Statistics are updated and displayed. This option is useful for investigating the statistics of a higher achievement skill group.

Creating Statistical Testing Tools

Non-Mastery Only– Temporarily hides the results of students who scored higher than 85%. Statistics are updated and displayed. This options is useful for investigating the statistics of a lower achievement skill group.

Reset – Deletes all results and zeroes out statistics so a new set can be analyzed.

After choosing a set of responses the screen will appear like this. (Fig. 4.3) Like before, there is a spreadsheet with metadata for the computer. It is displayed here so that users can check for errors. Also seen are overall statistics and a question detail for question 1.

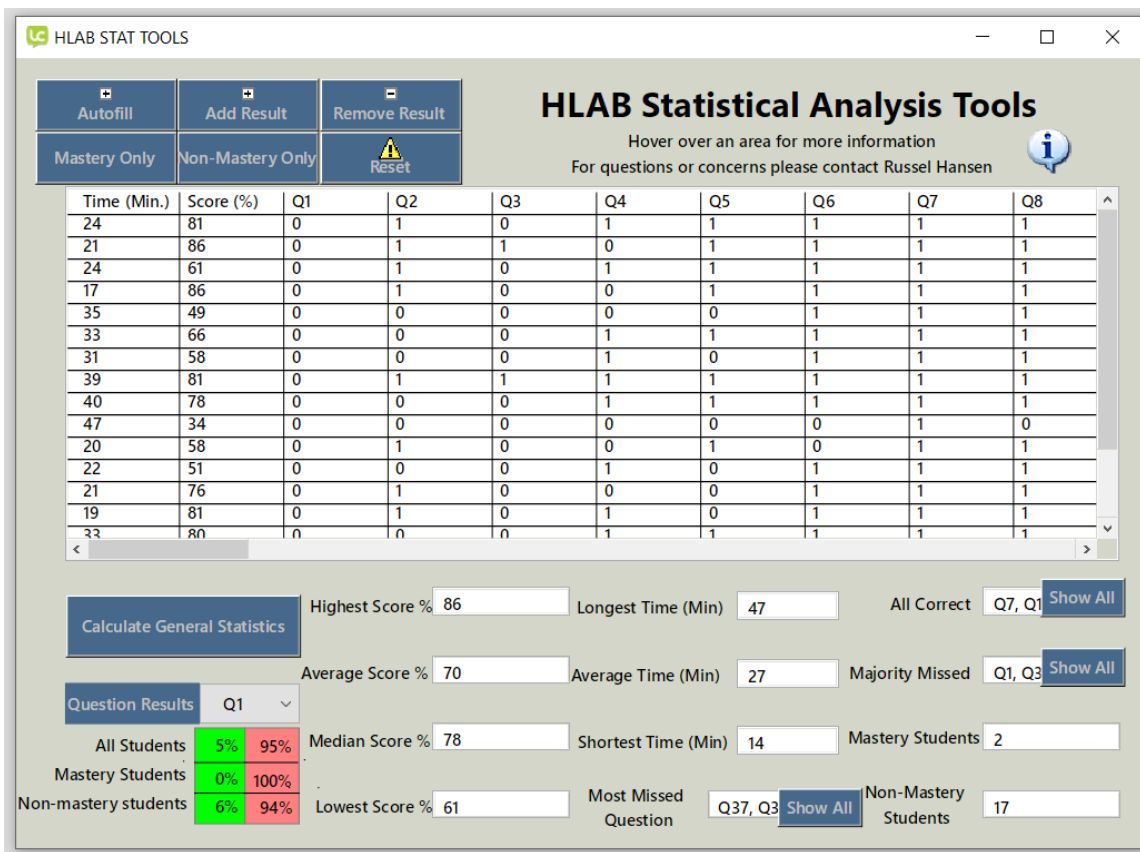


Fig. 4.4 HLAB Statistical Tool Filled – This is what the screen displays after information has been uploaded. Statistics are automatically calculated and displayed upon upload.

Exam Overview

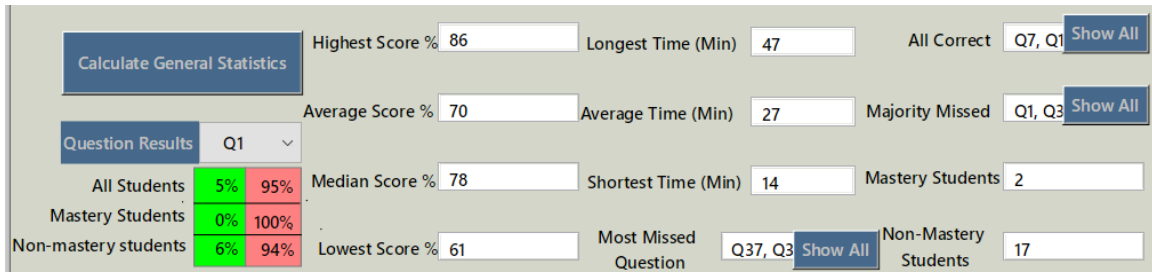


Fig. 4.5 HLAB Statistical Tool Exam Overview – Here the statistics are presented for the test scores shown above. The buttons labeled “show all” present a complete list of the questions found in that area in a small box.

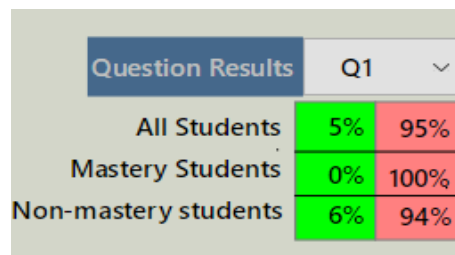
Each of the statistics has been explained above. Here professors can view statistics for their entire class, or classes, depending on the data they entered.

Calculate General Statistics – Individually calculates the statistics shown above and places the numbers in their respective field. Also indicated are the units for each statistic such as percentages or minutes.

Show All – Each of these buttons expands the list of the corresponding statistic. A small box appears with the complete list of questions, and the button changes to display “Hide” so that the list can be collapsed.

Question Detail

The final piece of the tool is the Question Detail. Here, test questions can be analyzed one by one. Results are divided by skill level as well as given overall, for reasons mentioned above. Clicking on the dropdown menu gives a list of all the questions on the test. Selecting one replaces the statistics seen here with the statistics for that question. Statistics are always shown in rounded percentages.



The screenshot shows a table titled "Question Results" with a dropdown menu set to "Q1". The table displays performance data for three student groups: All Students, Mastery Students, and Non-mastery students. Each row shows the percentage of correct answers in green and incorrect answers in red.

	Correct	Incorrect
All Students	5%	95%
Mastery Students	0%	100%
Non-mastery students	6%	94%

Fig 4.6 HLAB Statistical Tool Question Detail – Here, Question 1 of the exam is analyzed by skill level. Green indicates correct answers in percentage and red indicates incorrect answers.

IV. Results

It should be noted that these results are inconclusive. Due to COVID lowering the amount of social interaction the entire Spanish and Portuguese department, as well as many other teachers that were planning on using the Humanities Testing Lab, moved to other resources such as Canvas. Due to the reduced sample size, these results might not be accurate.

Professors who stayed in contact with the Humanities Testing Lab were contacted and asked to participate in two anonymous surveys. The surveys are identical. One was administered before the statistical tool was introduced. The second was administered after the tool had been experienced. In hypothesis, teacher perspectives of and experience with statistical tools would increase. A full script of the survey and data collected can be found in Appendix B. A brief summary of comments and findings will be presented here.

Findings

Based on the results of the survey, teachers who used the application had an increased desire for more statistical tool use in the classroom, and an increased confidence in their pedagogy skills and the way their tests impacted the classroom. This confirms the general hypothesis of the use of statistical tools, but once again, findings are inconclusive.

Future Applications

Based on feedback from the participants, there are many areas of improvement for both the application of the tool and the computer application itself.

The computer program can be improved in the following ways according to study participants:

- Additional Features Wanted:
 - A printable overview
 - The ability to manually input grades for essay type questions
- Unnecessary Features Found:
 - Adding one result at a time
 - Removing one result at a time

The study can be improved in the following ways:

- Increase the sample size of the participation group
- Compare class scores across several semesters as the tool is used

Works Consulted

Blooket, www.blooket.com/. Accessed 20 Oct. 2020

Canvas. canvas.instructure.com/. Accessed 15 Jun. 2020

“Learning Games: Make Learning Awesome!” *Kahoot!*, kahoot.com/. Accessed 12 Jun. 2020

Learning Tools & Flashcards, for Free | Quizlet. quizlet.com/. Accessed 12 Jun. 2020

learningsuite.byu.edu/. Accessed 15 Jun. 2020

Appendix A: Coding

Statistical Calculations: (LiveCode Programming language is Copyrighted)

On mouseup

set itemdelimiter to tab

//Get highest score

put 0 into HScore

repeat with b = 2 to the number of lines in field "results"

if item 2 of line b of field "results" > HScore **then**

put item 2 of line b of field "results" into HScore

end if

end repeat

put HScore into field "Highest Score"

//Get lowest score besides 0

put 100 into LScore

repeat with c = 2 to the number of lines in field "results"

if item 2 of line c of field "results" < LScore **then**

put item 2 of line c of field "results" into LScore

end if

end repeat

put LScore into field "Lowest score"

if Item 1 of line 2 of field "results" = empty **then**

put 0 into field "Lowest Score"

end if

//Get average score

put 0 into AvScore

repeat with s = 2 to the number of lines in field "results"

add item 2 of line s of field "results" to AvScore

end repeat

if the number of lines in field "results" > 2 **then**

Divide AvScore by (number of lines in field "results" - 1)

Put the round of AvScore into field "Average Score"

else

put 0 into field "average Score"

end if

//Get median score

put item 2 of line 2 of field "results" & comma into Mscore

repeat with d = 3 to the number of lines in field "results"

put item 2 of line d of field "results" into MCheck

Creating Statistical Testing Tools

```
        if Mcheck > line -1 of Mscore then
            put Mcheck & comma after Mscore
        end if
        if Mcheck < line 1 of Mscore then
            put Mcheck & comma before Mscore
        end if
    end repeat
    put median(Mscore) into field "Median Score"

//Get longest time
    put 0 into Htime
    repeat with e = 2 to the number of lines in field "results"
        if item 1 of line e of field "results" > Htime then
            put item 1 of line e of field "results" into Htime
        end if
    end repeat
    put Htime into field "Longest time"

//Get average elapsed time excluding 0's
    put 0 into AvTime
    repeat with s = 2 to the number of lines in field "results"
        add item 1 of line s of field "results" to AvTime
    end repeat
    if the number of lines in field "results" > 2 then
        Divide AvTime by (number of lines in field "results" - 1)
        Put the round of AvTime into field "Average time"
    else
        put 0 into field "average time"
    end if

//Get shortest time besides 0
    put item 1 of line 2 of field "results" into Ltime
    repeat with f = 3 to the number of lines in field "results"
        if item 1 of line f of field "results" is not 0 then
            if item 1 of line f of field "results" < Ltime then
                put item 1 of line f of field "results" into Ltime
            end if
        end if
    end repeat
    put Ltime into field "Shortest time"

if item 1 of line 2 of field "results" = empty then
    put 0 into field "Shortest time"
end if

put empty into field "All Correct"
```

```
put empty into field "majority"
put the number of lines in field "results" into Lcheck
repeat with j = 3 to the number of items in line 1 of field "results"
  put 0 into allCount
  repeat with i = 2 to the number of lines in field "results"
    if item j of line i of field "results" is 1 then
      add 1 to allCount
    end if
  end repeat
end repeat

//Get most missed question(s)
if allCount = Lcheck then
  put comma & " " & item j of line 1 of field "results" after field "most missed"
end if

if allCount < Lcheck then
  put item j of line 1 of field "results" into field "Most Missed"
  put allCount into Lcheck
end if

//List questions that all students answered correctly
if allCount = (the number of lines in field "results" - 1) then
  put item j of line 1 of field "results" & comma & " " after field "All
Correct"
end if

if item 1 of line 2 of field "results" = empty then
  put 0 into field "Most Missed"
  put 0 into field "All Correct"
end if

//list Questions that were missed by the majority of test takers
if allCount < (the number of lines in field "Results" - 1) / 2 then
  put item j of line 1 of field "results" & comma & " " after field "Majority"
end if

if item 1 of line 2 of field "results" = empty then
  put 0 into field "majority"
end if
end repeat

if line 2 of field "results" = empty then
  put 0 into field "most missed"
  put 0 into field "All correct"
  put 0 into field "majority"
end if
```

Creating Statistical Testing Tools

delete character -2 of field "all correct"

delete character -2 of field "majority"

//Calculate and show the percentage of test takers who achieved mastery = 85%

repeat with k = 2 to the number of lines in field "results"

if item 2 of line k of field "results" > 85 **then**

add 1 to passCount

end if

put passCount into field "Mastery"

//Calculate the percentage of students who failed

put (the number of lines in field "results" -1 - passCount) into field "Non Mastery"
end repeat

if item 1 of line 2 of field "results" = empty **then**

put 0 into field "mastery"

put 0 into field "non mastery"

end if

if line 2 of field "results" = empty **then**

set the label of button "questions" to empty

put 0 & "%" into field "allgood"

put 0 & "%" into field "allbad"

put 0 & "%" into field "masterygood"

put 0 & "%" into field "masterybad"

put 0 & "%" into field "NMgood"

put 0 & "%" into field "NMbad"

else

put line 1 of the label of button "questions" into QReset

send mouseup to button "questions"

end if

end mouseup

Appendix B: Surveys and Responses

Survey: (Formatting changes have been made.)

Questions one through five will ask whether you agree or disagree with a statement on a scale of Completely Disagree to Completely Agree. Please circle your level of agreement.

1. I feel comfortable creating and grading exams.

Completely Disagree Disagree Neutral Agree Completely Agree

2. The Humanities testing lab is in some ways better than the standard testing building.

Completely Disagree Disagree Neutral Agree Completely Agree

3. I think students prefer being tested on computers.

Completely Disagree Disagree Neutral Agree Completely Agree

4. I would like more tools to help me assess students learning.

Completely Disagree Disagree Neutral Agree Completely Agree

5. I feel very confident in my pedagogical skills.

Completely Disagree Disagree Neutral Agree Completely Agree

Questions 6-8 are short answer questions. Please respond in 2-4 sentences.

6. What was the last average test score of your classes? Was it higher or lower than you expected?

7. What do you feel is the best possible relationship between students, teachers, and tests?

8. How do test scores affect what you teach, if at all?

9. Are you familiar with any grading tools or other forms of digital analysis? What are your thoughts on these tools?

10. Please use the space provided below to record any remaining comments you have on the humanities testing lab or statistical tools and what role they play in your teaching.

Please check this box to show that you provide consent for your answers to be anonymously quoted as part of a research study: ____

Creating Statistical Testing Tools

General Results:

Survey 1

1. 4
2. 4
3. 2.5
4. 3.5
5. 3.5
6. Students Achieved Mastery Level or Higher
7. Measurement, Motivation
8. Item analysis, no effect
9. Unfamiliar
10. Time = Participation, Paperless is better

Survey 2

1. 4
2. 4
3. 3
4. 4
5. 5
6. Students Achieved Mastery Level or Higher
7. Measurement, Mastery
8. Item analysis
9. Useful for evaluating format and student performance
10. Bugs in program

Appendix C: Complete Canvas Table (With formatting changes)

Question Id	Question Title	Count	Top	Middle	Bottom	Quiz	Correct	Wrong
1514145	Questions 1-4	110	30	70	10	6	83	27
1514147	Questions 1-4	110	30	70	10	6	53	57
1514150	Questions 1-4	110	30	70	10	6	46	64
1514151	Questions 1-4	109	30	70	9	6	90	19
1514208	Questions 5-6	109	30	69	10	6	31	78
1514209	Questions 5-6	110	30	70	10	6	18	92

Correct Student Ratio	Wrong Student Ratio	Correct Top Student Count	Correct Middle Student Count	Correct Bottom Student Count
0.754545455	0.245454545	29	52	2
0.481818182	0.518181818	27	25	1
0.418181818	0.581818182	20	23	3
0.825688073	0.174311927	29	59	2
0.28440367	0.71559633	18	13	0
0.163636364	0.836363636	7	11	0

Variance	Standard Deviation	Difficulty Index	Alpha	Point Biserial of Correct	Point Biserial of Distractor 2	Point Biserial of Distractor 3	Point Biserial of Distractor 4
0.185207	0.430356	0.754545	0.090176	0.448263	-0.24672	-0.05076	-0.34034
0.249669	0.499669	0.481818	0.090176	0.573441		0.014177	-0.57875
0.243306	0.49326	0.418182	0.090176	0.37421	-0.264	-0.23242	
0.143927	0.379377	0.825688	0.090176	0.38699	-0.37268	-0.11515	0.005562
0.203518	0.45113	0.284404	0.090176	0.530239	-0.18217	0.000672	-0.34015
0.13686	0.369945	0.163636	0.090176	0.163791	0.101452	-0.00531	-0.25066