

A System for Automatically Grading Graphs in an Educational Setting

by

Daniel A. Martelly

B.S. EECS, Massachusetts Institute of Technology (2014)

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Masters of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Author
Department of Electrical Engineering and Computer Science
May 20, 2016

Certified by.....
Adam J. Hartz
Lecturer
Thesis Supervisor

Accepted by
Dr. Christopher J. Terman
Chairman, Masters of Engineering Thesis Committee

A System for Automatically Grading Graphs in an Educational Setting

by

Daniel A. Martelly

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2016, in partial fulfillment of the
requirements for the degree of
Masters of Engineering in Electrical Engineering and Computer Science

Abstract

Online educational tools which automatically grade student answers are typically restricted to structured input such as multiple choice, numbers, or structured text. This thesis describes the design and implementation of a tool which can automatically grade the unstructured graphical input of 2D graphs. Student answers are captured on a web interface which allows freeform drawings on a grid similar to drawing on graph paper. Automatic grading is accomplished by running a series of independent tests on a student's drawing. Using six criteria, a large variety of drawings can be graded, and the result of automatically grading a criterion often matches that of a human grader. In addition, to make this system accessible to teachers, an interface for choosing criteria has been designed and tested for use by teachers with little to no programming experience.

Thesis Supervisor: Adam J. Hartz
Title: Lecturer

Acknowledgments

I would like to thank Mr. Hartz for his constant guidance and support. He would always find the time to discuss ideas with me, help me plan my next steps, and was patient with me even while I was constantly late with deadlines.

This project would not have the same relevance without the students and teachers who tested the tools I created and gave me their feedback. I would like to thank all the students and teachers who contributed their time, as well as Rob Miller and Michael Keohane for their help in getting approval for human subject testing.

Contents

1	Introduction	15
1.1	Motivation	15
1.2	Outline	17
2	System Overview	19
2.1	Automatic Grading	19
2.2	Student Input	19
2.3	Creating Questions	21
3	Grading Methods	23
3.1	Design	23
3.2	Grading Algorithms	24
3.2.1	Common Approaches	24
3.2.2	Individual Criterion Algorithms	25
3.3	Example	27
4	Student User Interface	29
4.1	Design	29
4.2	Student Feedback	31
5	Teacher User Interface	33
5.1	Design	33
5.1.1	Display Options	35
5.1.2	Criteria Options	35

5.1.3	Draw Testing	39
6	Results	41
6.1	Grading Algorithm Performance	41
6.1.1	Procedure	41
6.1.2	Performance	42
6.2	Teacher User Interface Testing	45
6.2.1	Procedure	46
6.2.2	Teacher Results	46
7	Discussion	53
7.1	Results	53
7.1.1	Grading Performance	53
7.1.2	Teacher Tests	54
7.2	Alternative Approaches	55
7.2.1	Grading Algorithms	55
7.2.2	Student Interface	56
7.3	Future Work	57
7.4	Conclusion	58
A	Grading Performance Details	59
A.1	Circuit Questions Asked to Students	59
A.1.1	Power Hungry	59
A.1.2	Mixing Thevenins	62
A.1.3	Potentiometer Loading	64
A.1.4	Cats and Tart Pops	66
A.2	Interface Used for Recording Teacher Criteria Judgements	67
A.3	Survey	68
A.4	Graphs on which the algorithms and humans disagreed	69
A.4.1	Function Followed false negatives	70
A.4.2	Is Function false positive	72

B	Experimental Setup for Teachers	73
B.1	Procedure	73
B.2	Experiment Overview Web Page	74
B.3	Warm up Questions Web Page	76
B.4	Tasks for Testing Web Page	76
B.4.1	Rectangular Pen	77
B.4.2	Sunday Driver	77
B.5	Survey	78
C	Observations and Feedback from Teachers	79
C.1	Written Comments	79
C.2	Student Interface	80
C.2.1	Positive Feedback	80
C.2.2	Negative Feedback	80
C.2.3	Cosmetic Fixes	80
C.3	Teacher Interface	81
C.3.1	Positive Feedback	81
C.3.2	Negative Feedback	81
C.3.3	Cosmetic Fixes	82
C.3.4	Individual Results	83

List of Figures

2-1	System Diagram	20
3-1	Flow Chart of Grading Procedure	24
4-1	This student interface allows for any kind of drawing such as the parabola, C shape, and triangle displayed here.	30
5-1	Teacher interface with Criteria Options and Draw Testing menus selected	34
5-2	Teacher interface with Display Options and a preview of the student interface showing	34
5-3	Teacher interface with Criteria Options and criteria visualizations shown	35
5-4	Domain Used Visualization	36
5-5	Is a Function Visualization	36
5-6	Function Visualization	37
5-7	Derivative Visualization	37
5-8	Critical Points Visualization where the error margin for the origin point is 5 pixels and the error margins of the other two points set to 10 pixels.	37
5-9	Monotonicity Visualization	38
5-10	Example visualization when multiple criteria are in use	38
5-11	Selecting Draw Testing shows a student interface which will automatically be graded using the chosen criteria	40
6-1	Number of Domain Used criteria that human graders agreed on . . .	42
6-2	Number of Is a Function criteria that human graders agreed on . . .	43
6-3	Number of Critical Points criteria that human graders agreed on . . .	43

6-4	Number of Function Followed criteria that human graders agreed on .	44
6-5	Number of Monotonicity criteria that human graders agreed on . . .	45
6-6	47
6-7	47
6-8	48
6-9	48
7-1	A parabola drawn with the Flot interface	56
7-2	A C shape and a triangle drawn with the Flot interface	56
A-1	Answer for power consumed by 500Ω resistor	60
A-2	Answer for power consumed by resistor R	60
A-3	Answer for total power consumed by resistors	60
A-4	Answer for power consumed by 500Ω resistor	61
A-5	Answer for power consumed by resistor R	61
A-6	Answer for total power consumed by resistors	62
A-7	Answer IV curve for circuit 1	62
A-8	Answer IV curve for circuit 2	63
A-9	Answer IV curve for circuit 3	63
A-10	Reference configuration and graph	64
A-11	Circuit configuration 1 and answer graph	64
A-12	Circuit configuration 2 and answer graph	65
A-13	Circuit configuration 3 and answer graph	65
A-14	Circuit configuration 4 and answer graph	65
A-15	Circuit configuration 5 and answer graph	66
A-16	Answer for graphing V_{spray} versus V_{bb}	67
A-17	The web interface used to record teachers' judgements on criteria. . .	68

List of Tables

6.1	Human/Computer agreement for Domain Used Criterion	42
6.2	Human/Computer agreement for Is a Function Criterion	43
6.3	Human/Computer agreement for Critical Points Criterion	44
6.4	Human/Computer agreement for Function Followed Criterion	44
6.5	Human/Computer agreement for Monotonicity Criterion	45
6.6	Human/Computer agreement for all Criteria combined	45

Chapter 1

Introduction

This paper discusses a project for use in online courses, which automatically grades 2D graphical inputs. The paper describes the algorithms used for automatically grading graphs as well as the graphical user interfaces (GUIs) created for teachers and students to input data into the system.

With the completion of the first version of this project, teachers can decide how students' responses will be graded and what kind of feedback the students will receive. Teachers can accomplish this by selecting different kinds of criteria and customizing parameters according to the question. Students can view a question, draw their response, and submit their answer to a server. The server will automatically grade their answer and send them feedback as necessary so that they may correct any mistakes they have made.

1.1 Motivation

A survey in 2013 by the National Center for Education Statistics showed that 25% of students in postsecondary institutions are taking at least one online course [4]. Coursera, edX, and Udacity are examples of online educational initiatives that have affected millions of students [10]. For example, with the supervision of only a handful of professors, MIT edX's 6.002x has taught a college level class to completion to thousands of students in an interactive way in a single semester [6]. Even in a traditional

classroom where students have direct contact with the teacher, teachers are starting to take advantage of online tools to help their students. Courses at MIT such as 6.01 (Introduction to EECS via Robot Sensing, Software and Control), 18.03 (Differential Equations), and 8.02 (Physics II) have homework questions which are entered online and automatically graded.

Currently, most questions on online courses are limited to highly structured input such as multiple choice, numerical answers, formulas, and small pieces of code. Some well known platforms such as LON-CAPA [2], edX [6], and CAT-SOOP [8] are expanding the range of online questions by adding interactive graphs, mechanics simulators, and circuit simulators. However, online tools are still far away from being able to ask the variety of questions a teacher can ask in a traditional classroom (such as mathematical derivations, proofs, and graphs). This project adds a new kind of freeform input for graphing to online courses. Having input be as freeform as possible more closely mimics a traditional paper question. Drawing graphs is an important skill used in many fields such as engineering, history, and economics. Anecdotal evidence from teachers and researchers suggests that students benefit from analyzing and synthesizing appropriate graphs for a concept that is being learned [7, 9].

One particular way that online tools can complement teachers is by providing instantaneous feedback to students. There is still a lot of disagreement among researchers on the best ways to provide feedback, but a meta-analysis by V.J. Shute provides some evidence that retention of procedural knowledge such as mathematics and programming is improved when students are given immediate feedback [11]. The same meta-analysis references several papers which show that learning is more effective when feedback is specific rather than simply indicating if the answer is right or wrong. For this reason, it was important when working on this project to make it possible to give specific feedback.

1.2 Outline

Chapter 2 gives an overview of the parts of the project and how they fit together. Chapter 3 discusses the approach to automatic grading of graphs. Chapter 4 describes how student input is collected. Chapter 5 describes the interface for creating graph questions. Chapter 6 describes the results of testing the performance of the grading algorithms and the results of testing the teacher interface on teachers. Finally, chapter 7 discusses the results and proposes future work based on the results.

Chapter 2

System Overview

This project can be separated into three major parts: how graphs are automatically graded, how student input is captured, and how teachers create graphing questions. Figure 2-1 shows how the three parts interact.

2.1 Automatic Grading

Grading takes place on a server which receives student's drawings and sends back a numerical grade and textual feedback. Submissions are graded by running the drawing through a series of independent tests. Each test is called a criterion throughout this paper. The results of grading from each criterion are aggregated and sent back to the student. At the time of writing, there are six criteria to choose from for grading which are explained in detail in chapter 3. Each criterion can be run multiple times with differing parameters when grading a drawing.

2.2 Student Input

A student's input to a graphing question is captured on a web interface which is meant to emulate drawing on graph paper. The web interface, which is discussed in detail in chapter 4, allows freeform drawings on coordinate axes using four basic

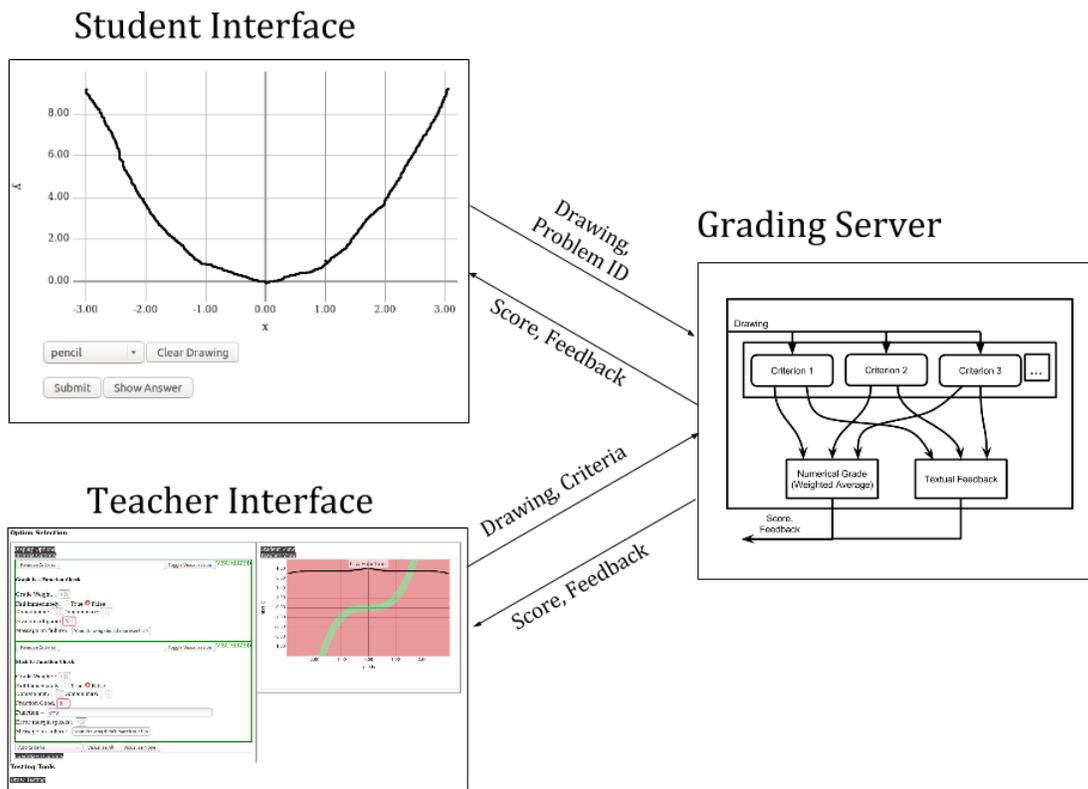


Figure 2-1: System Diagram

tools including a pencil and an eraser. When a student has drawn an answer, he can submit his drawing to the server to be graded and receive feedback.

2.3 Creating Questions

To create a question, a teacher must select which criteria he wants to use for a question, as well as the associated parameters for each criterion. Teachers do this through a web interface which allows them to select criteria and parameters, visualize the chosen criteria, and test their choices. Teachers are able to customize not only the metrics by which a graph is graded, but also the size and limits of the axes that are provided to the student. The details are discussed in detail in chapter 5

Chapter 3

Grading Methods

This chapter will describe the details of the grading algorithms.

3.1 Design

A drawing is sent to the server as a string in JSON format, which encodes information about which pixels were colored black, as well as information about the axes. The points on the graph are not saved directly as points on the axes, but the points relative to the axes can be calculated based on the location of the pixels. Criteria for a question are specified in a string or text file which is loaded by the grader. The grading algorithm for each criterion is applied to the drawing independently from the other criteria (as shown in figure 3-1). When the algorithms for each criteria have been applied to the drawing, the resulting numerical grades and textual feedback are combined together. When combining the results, the grader refers to a **weight** parameter and a **failFast** parameter for each criterion. The **weight** parameter is used for calculating a weighted average of all the numerical grades. **failFast** is a boolean parameter that when true, will cause the final grade for the drawing to be 0% if that individual criterion gets a grade of 0%.

A criterion only returns textual feedback when its individual grade is below 100%. By default, this feedback consists of a generic message describing which criteria were not satisfied. However, teachers can change the feedback for each criterion to provide

more specific feedback.

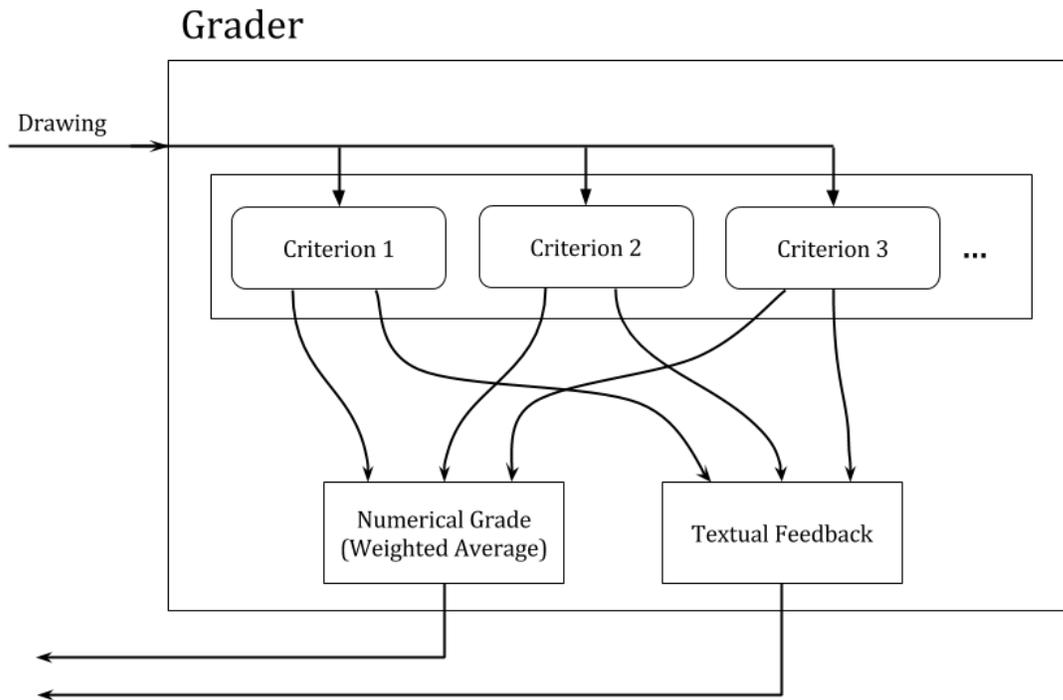


Figure 3-1: Flow Chart of Grading Procedure

3.2 Grading Algorithms

Each type of criterion has its own method of grading; however, there are some shared approaches to dealing with noise, which will be discussed before explaining each individual criterion type in more detail.

3.2.1 Common Approaches

Pass fraction Many of the criteria check each point drawn by the student independently from other points. In this case, the system counts the number of points which fulfill a criterion and the total number of points in the drawing. As long as the fraction of correct points over total points is above a certain threshold, the student still gets full credit for the criterion.

Error margin Some criteria check that a value, such as a point or the slope, is close enough to a desired value to be considered correct. The error margin is meant to account for difficulty in drawing precisely. The error margin is measured in pixels for points or in terms of an angle for the slope, because these relate more directly to people's ability to draw than an error in x , y , or the slope.

3.2.2 Individual Criterion Algorithms

In total, there are six criteria which are currently integrated into the system. Each criterion will be described below including details about what parameters each criterion accepts and the algorithm used for grading.

Domain Used This criterion checks that the student has drawn over a sufficient number of x values within a given domain. Teachers can choose the size of the domain that should be checked and what fraction of the domain should be covered by the drawing. By default, this criterion is satisfied when the drawing covers at least 80% of the possible x values.

Is a Function This criterion checks that the submitted drawing represents a function. A function is defined to be a mapping where any x value has a maximum of one associated y value. By default, this criterion is satisfied when 95% of the drawn points do not share an x value with another point.

Critical Points This criterion checks that the drawing passes through certain critical points on the x , y coordinate system within a chosen error margin. By default, this criterion is satisfied when at least one point in the drawing is at most 10 pixels away from the critical point.

Function Followed This criterion checks that the drawing is close enough to a specified answer function. To determine whether a drawn point is close enough to the answer, an error margin measured in pixels is used, which by default is set to 10 pixels. The algorithm first creates a list of x , y points which corresponds to the

answer function and converts them all to pixel coordinates. Then each drawn pixel coordinate is checked to see if it is within the error margin to one of the known answer pixel coordinates. By default, this criterion is satisfied if at least 80% of the drawn points are close enough to the answer function.

Derivative Matched This criterion checks that the slope of the drawing matches the slope described by a function within an error margin. By default the error margin is set to an angle of 20 degrees. To calculate the slope, the drawn function is smoothed lightly using a Gaussian blur, and then passed into an algorithm designed by Chartrand which is used for finding derivatives in noisy data [5, 12]. From the calculated derivatives, the angle of the slope at each point in the drawing is found. The angle of the slope calculated from the drawing is compared to the angle of the slope calculated based on the derivative function provided as a parameter. If the angle from the drawing is within the error margin of the answer angle, the slope is considered correct. By default, this criterion is satisfied if 80% of the drawn slopes match. The criterion assumes that the drawing represents a function and may behave unexpectedly otherwise.

Monotonicity This criterion checks either that the drawing either is a monotonically increasing function or checks that the drawing is a monotonically decreasing function. To determine whether a function is monotonically increasing, two checks are done. The first check looks at a sliding window of 40 values and makes sure that the biggest decrease within the window never exceeds 5 pixels. The choice of using 40 values for the sliding window and limiting to a 5 pixel decrease was determined trial and error with drawings done by the author of this paper when using a touchpad. The second check iterates from left to right and ensures that the y values never go too far below the maximum y value seen so far. How far the drawing can go below the current global maximum is determined by an error margin set to a default of 10 pixels. To check if a function is monotonically decreasing, the algorithm is run the same way except that the y values are flipped over the x axis prior to the two

checks. A user determines whether to check for a monotonically increasing function vs a monotonically decreasing function by setting a `trend` parameter.

3.3 Example

This section will walk through a possible choice in criteria to grade a simple parabola, $y = x^2$. Different people may have different opinions about how to grade a parabola. In this case the author thought that the drawing should have the approximate shape of the parabola and that the drawing should clearly include the global minimum, $(0, 0)$. The following choice in criteria accomplish this:

- **Domain Used:** Using this criterion ensures that the student has drawn over a sufficient amount of the domain to warrant further grading. The `failFast` parameter could be set to `True` so that a student earns 0% if they have not drawn a complete graph. The feedback on failure for this criterion could be, “Please draw over more of the domain of the graph.”
- **Is a Function:** Using this criterion ensures that the student has drawn a function. The `failFast` parameter could be set to `True` so that a student doesn’t receive confusing feedback from the next criterion, which works most reliably when the drawing represents a function. The feedback on failure for this criterion could be, “Make sure your drawing represents a function (i.e. a maximum of one y value for every x value).”
- **Function Followed:** The drawing could be compared against a function representing $y = x^2$. The default error margin of 10 pixels and default pass fraction of 80% could be used to account for human error in drawing. The feedback on failure for this criterion could be, “Your answer was incorrect, please try again.”
- **Critical Points:** The maximum or minimum of a parabola is could be considered an important feature. To require more precision at the global minimum of the parabola, the point $(0, 0)$ could be passed in as a critical point with a reduced

error margin of 5 pixels ensuring that the drawing came close to the specified value. The feedback on failure for this criterion can be, “Think about where the minimum or maximum of this function is and try again.”

Chapter 4

Student User Interface

The purpose of the student user interface is to gather the intended drawing of the students while allowing the input to be as freeform as possible. To make this tool accessible to students on a wide range of machines, the interface was developed and has been tested on modern web browsers such as Mozilla Firefox and Google Chrome.

4.1 Design

The student interface is meant to emulate the freedom of drawing on paper. Screenshots of the interface and some example drawings made using a touchpad can be seen in figure 4-1. The interface consists of a rectangular drawing area with labeled axes, a selection of tools, a submit button, and a button for showing the answer. The underlying representation is an unordered list of all pixels the student has drawn, as well as information about the axes to help determine the conversion to an (x, y) coordinate system. The points where a student has drawn are represented as black 3 x 3 squares drawn on top of the axes so that they may clearly be seen. Besides the points being drawn, a recording of the student's mouse strokes, mouse state, and tool choice are saved. Currently, only information about drawn points is used during grading.

Four drawing tools are made available to students: a pencil tool, eraser tool, line tool, and a smooth pencil tool. The pencil tool places points on the drawing area

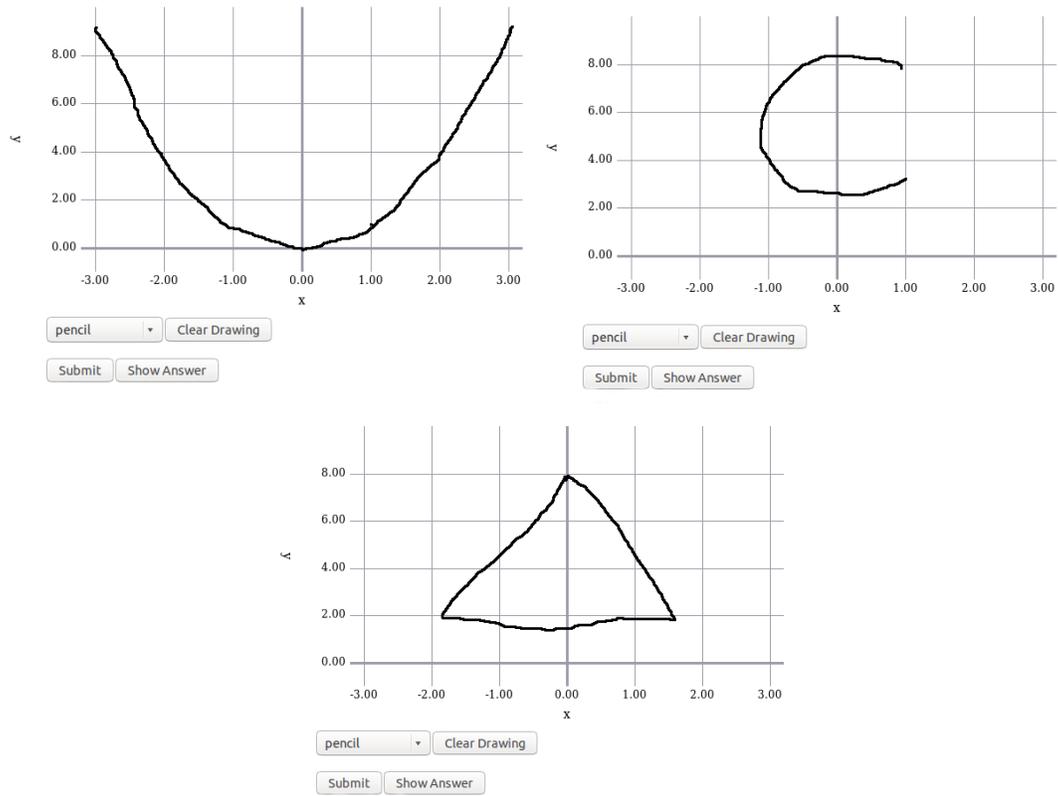


Figure 4-1: This student interface allows for any kind of drawing such as the parabola, C shape, and triangle displayed here.

wherever the pencil is clicked and dragged. The eraser tool erases all the points from a circular area. The line tool draws a straight line starting from where the student presses the mouse button and ending where they release the mouse button. Finally, the smooth pencil tool helps remove excess noise by doing a 2D Gaussian blur on a curve drawn in a single mouse stroke.

The interface automatically draws and labels axes. Since the drawing area is transparent except for where a student has drawn, the axes are generally visible. Teachers select the minimum and maximum of both the x and y axes as well as the step size between axis lines. Teachers also set the axis labels. In addition, teachers may optionally give students the option of setting the minimum, maximum, and step size of the x axis, y axis, or both axes.

When a student submits his drawing, it is converted to a JSON string to be sent to the grading server. After grading, the server sends back a grade and a list of strings indicating textual feedback, as described in chapter 3. The grade and textual feedback are shown below the submit button.

4.2 Student Feedback

As part of an experiment on grading algorithm performance, students were asked to draw on an earlier version of the interface in which only the pencil and eraser tools were available. Overall, based on survey responses as shown in appendix section A.3, students appear to have enjoyed using the interface, but some students complained about the difficulty of drawing with a trackpad. In response to several comments in regards to the difficulty of drawing straight lines and drawing smooth curves, the line tool and smooth pencil tool were added to the interface.

Chapter 5

Teacher User Interface

This section describes a web interface created to help teachers teachers select parameters for a graphing problem and visualize how the grading procedure will happen. The motivation for creating the interface was to make the this project accessible to teachers with little to no programming experience.

5.1 Design

When creating a question, teachers choose how the graphing area will be displayed to students, as well as how the students' drawings will be graded. The grading criteria for a particular problem is stored as formatted text which is possible for teachers to edit directly. However, it may be difficult for teachers to describe what criteria they want in this format and to predict the result so a web interface was created to help. In the interface and the write up below, the choices of how the graphing area will be displayed are called "Display Options" and the choices of which criteria will be used for grading are called "Criteria Options".

As a teacher changes the parameters in the Display Options and Criteria Options, visualizations of their choices update in real time. While they are choosing their criteria, teachers can test their choices as if they were a student by submitting a drawing using the example student interface at the bottom of the web page. Screenshots of the teacher interface can be found in figures 5-1, 5-2, and 5-3.

Option Selection

The screenshot displays the 'Criteria Options' panel on the left and a 'Student View' preview on the right. The 'Criteria Options' panel includes a 'Remove Criteria' button, a 'Toggle Visualization' button (set to 'VISUALIZED'), and a 'Stick to Function Check' section. Below this, there are input fields for 'Grade Weight' (set to 1), 'Fail Immediately' (radio buttons for 'True' and 'False', with 'False' selected), 'Domain min' and 'Domain max' (empty), 'Fraction Good' (set to 0.5), 'Function = |abs(x)|' (text input), 'Error margin (pixels)' (set to 10), and 'Message on failure' (set to 'Your drawing didn't match'). At the bottom of the panel are 'Add Criteria' and 'Load/Save Options' buttons, and a dropdown menu for 'Visualize All' and 'Visualize None'. The 'Student View' preview shows a coordinate plane with a red shaded area and a green V-shaped line. The x-axis is labeled 'xaxis' and the y-axis is labeled 'yaxis', both ranging from -4.00 to 4.00.

Testing Tools

The screenshot shows the 'Draw Testing' tool interface. It features a blank coordinate plane with x and y axes ranging from -4.00 to 4.00. Below the graph, there are buttons for 'pencil', 'Clear Drawing', 'Submit', and 'Show Answer'.

Figure 5-1: Teacher interface with Criteria Options and Draw Testing menus selected

Option Selection

The screenshot displays the 'Display Options' panel on the left and a 'Student View' preview on the right. The 'Display Options' panel includes sections for 'X axis' and 'Y axis'. For the 'X axis', there are radio buttons for 'Linear scale' (selected) and 'Log scale', and radio buttons for 'Axis set by teacher' (selected) and 'Axis set by student'. Below these are input fields for 'Min: -3', 'Max: 3', and 'Step: 1'. The 'Y axis' section has similar settings with 'Min: -5', 'Max: 5', and 'Step: 1'. Both sections have 'Dimension in pixels' input fields (500 px for X and 300 px for Y). At the bottom of the panel are 'Criteria Options' and 'Load/Save Options' buttons. The 'Student View' preview shows a blank coordinate plane with x and y axes ranging from -4.00 to 4.00. Below the graph are buttons for 'pencil', 'Clear Drawing', 'Submit', and 'Show Answer', and a 'feedback in here' link. The 'Teacher View' label is visible at the bottom of the preview area.

Testing Tools

Draw Testing

Figure 5-2: Teacher interface with Display Options and a preview of the student interface showing

Option Selection

The screenshot displays the teacher interface for setting criteria. It is divided into two main sections: 'Graph is a Function Check' and 'Stick to Function Check'. Each section has a 'Remove Criteria' button and a 'Toggle Visualization' button labeled 'VISUALIZED'. The 'Graph is a Function Check' section includes a 'Grade Weight' of 1, 'Fail Immediately' set to False, 'Domain min' and 'Domain max' fields, a 'Fraction of good' of 0.3, and a 'Message on failure' that reads 'Your drawing should represent a fi'. The 'Stick to Function Check' section includes a 'Grade Weight' of 1, 'Fail Immediately' set to False, 'Domain min' and 'Domain max' fields, a 'Fraction Good' of 0.3, a 'Function' input field containing 'x**3', an 'Error margin (pixels)' of 1, and a 'Message on failure' that reads 'Your drawing didn't match our fun'. At the bottom, there are buttons for 'Add Criteria', 'Visualize All', and 'Visualize None', along with a 'Load/Save Options' link.

On the right side, the 'Student View Teacher View' shows a graph titled 'Is A Function'. The graph has an x-axis from -2.00 to 2.00 and a y-axis from -4.00 to 4.00. A green curve is plotted, and a black line is drawn above it. The background of the graph is red.

Testing Tools

Draw Testing

Figure 5-3: Teacher interface with Criteria Options and criteria visualizations shown

5.1.1 Display Options

By filling out the text boxes in the Display Options section, teachers specify how the student interface will look like when presented to students. Teachers specify the width and height in pixels of the input area for the student interface, as well as the axis labels. In addition, teachers determine the minimum and maximum displayed values for each axis and the step size between graph lines. Finally, teachers can decide whether a student should have control over the minimum and maximum values of each axis. As teachers change each Display Options parameter, a preview of the student interface on the right side of the web page is updated in real time as shown in figure 5-2.

5.1.2 Criteria Options

The Criteria Options section allows teachers to add and remove criterion to the grading scheme and select the parameters for each criterion. Each criterion is encapsulated in its own box to visually separate it from the other criteria as shown in figure 5-3. Each box contains the name of the criterion in bold text, a small explanation of what

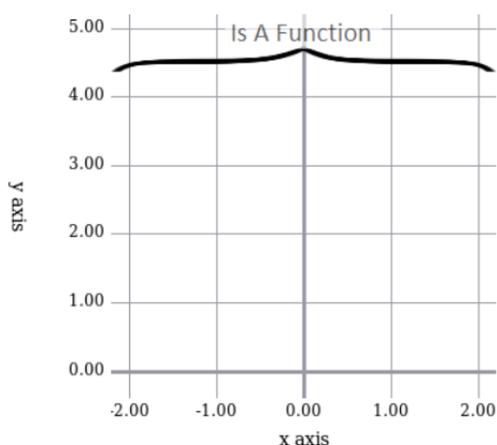
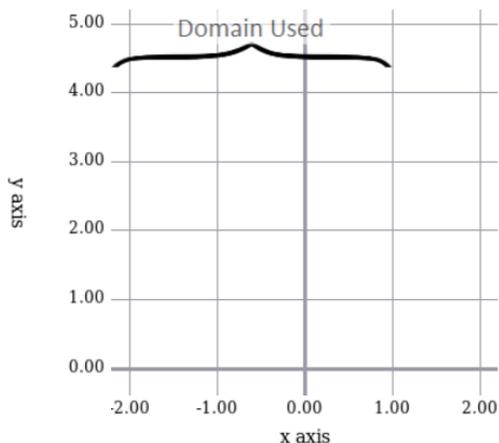


Figure 5-4: Domain Used Visualization Figure 5-5: Is a Function Visualization

the criterion checks for, and several fields for specifying each parameter of the criteria. If an invalid choice for a parameter has been made, the border of the box takes on a red color, and red text appears near the top of the box explaining that invalid inputs have been found. If criterion parameters are valid, the box border takes on a green color when the criterion is being visualized actively in the Teacher View. Otherwise (i.e. the criterion has valid inputs but is not being visualized) the box border is grey. To change whether a criterion is being actively visualized, a button labeled “Toggle Visualization” is located at the top of a criterion box which toggles visualization.

Each criterion has its own way of being visualized on the graph. These visualizations are all updated in real time as the teacher changes parameters. The visualizations also update automatically if the Display Options are changed.

Domain Used This criterion draws a bracket over the domain of x values which a student’s drawing should cover. (See figure 5-4.)

Is a Function This criterion draws a bracket over the area of the graph in which a student should draw a function. (See figure 5-5.)

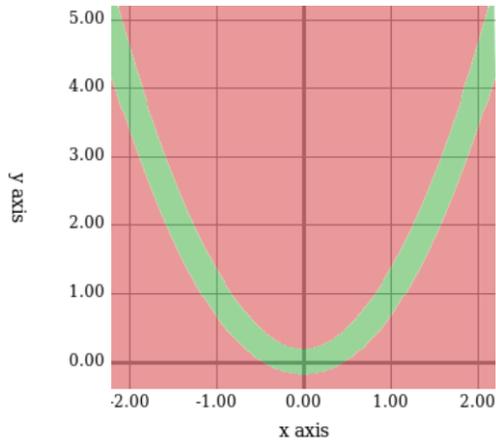


Figure 5-6: Function Visualization

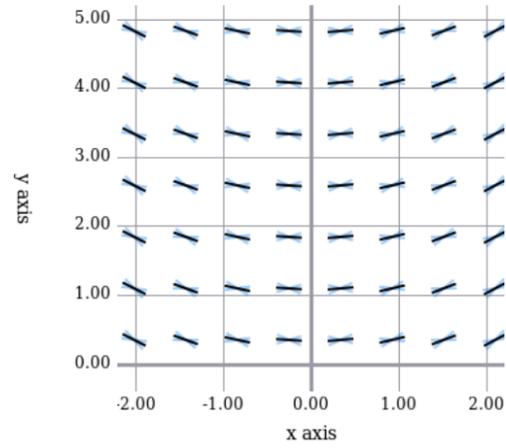


Figure 5-7: Derivative Visualization

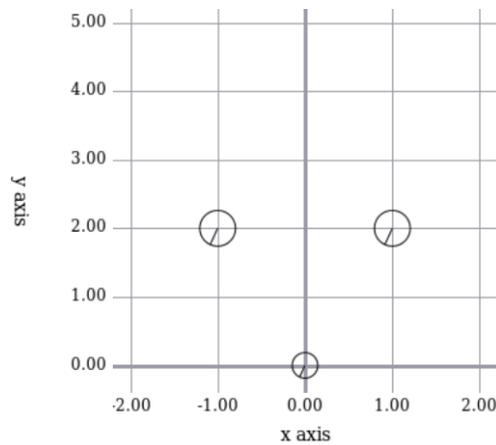


Figure 5-8: Critical Points Visualization where the error margin for the origin point is 5 pixels and the error margins of the other two points set to 10 pixels.

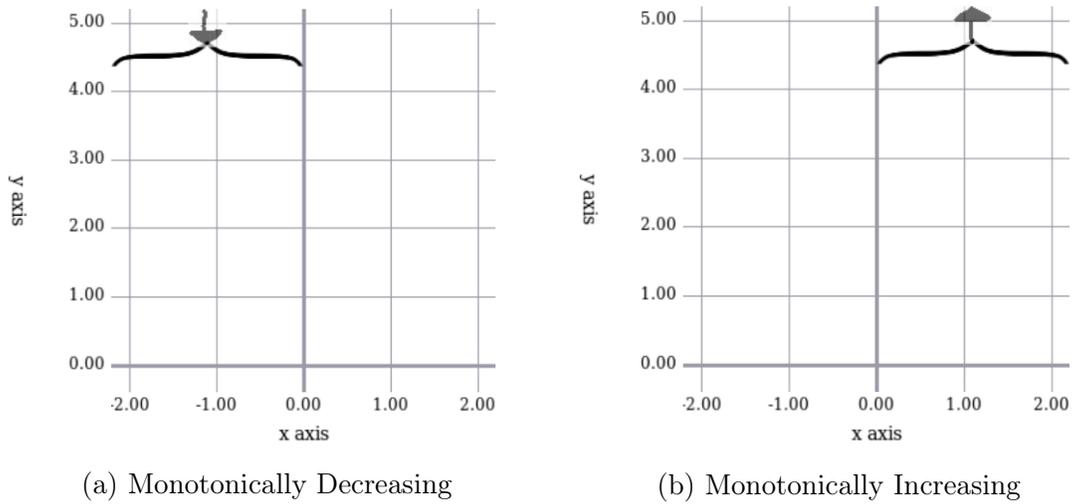


Figure 5-9: Monotonicity Visualization

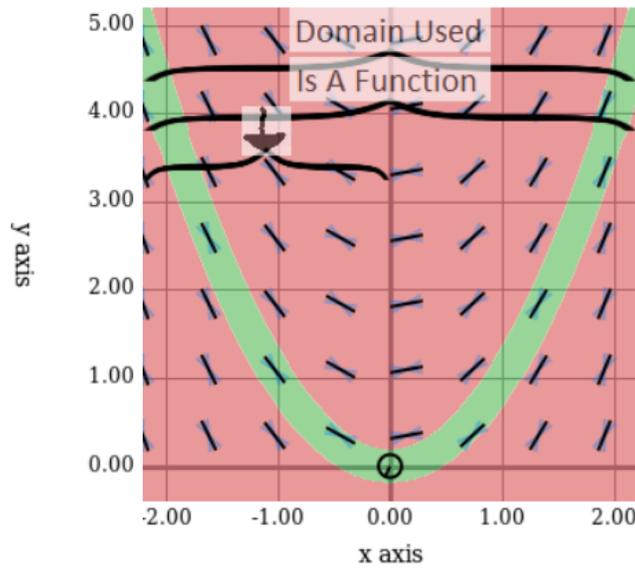


Figure 5-10: Example visualization when multiple criteria are in use

Critical Points This criterion draws circles centered around the point that a student's drawing should pass through. The radius of the circle is determined by the error margin chosen for the criterion which defaults to 10 pixels. (See figure 5-8.)

Function Followed This criterion draws green in all the places a student is expected to draw and red in all the places that a student should not draw. The size of the green area can be changed by altering the error margin, which describes how close the student's drawing has to be to the answer function. (See figure 5-6.)

Derivative Matched This criterion draws slope lines over the region of the graph in which the slope must match the provided derivative function. (See figure 5-7.)

Monotonicity This criterion draws a bracket over the area of the graph which must be monotonic. The icon on top of the bracket changes depending on the desired trend for the criterion. If the criterion is testing that a function is monotonically increasing, an arrow pointing upwards is shown. If the criterion is testing that a function is monotonically decreasing, an arrow pointing downwards is shown. (See figure 5-9.)

5.1.3 Draw Testing

At the bottom of the teacher interface there is a space designated for testing the criteria as shown in figure 5-11. Teachers can draw on an input coordinate system, submit their drawing, and receive the grade and feedback a student would receive for that drawing with their chosen criteria. As the teacher changes the Display Options, their drawing will be scaled and translated accordingly. However, since grading often takes more than a second, the grade and feedback associated with their drawing is not updated automatically with a change in Criteria Options or Display Options. Instead, a teacher must resubmit their drawing to see the effect on their grade.

Option Selection

[Display Options](#)
[Criteria Options](#)
[Load/Save Options](#)

[Student View](#)
[Teacher View](#)

Testing Tools

Draw Testing

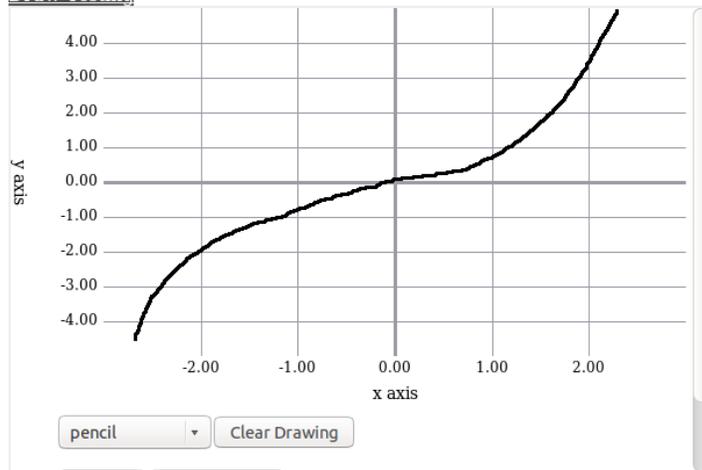


Figure 5-11: Selecting Draw Testing shows a student interface which will automatically be graded using the chosen criteria

Chapter 6

Results

6.1 Grading Algorithm Performance

Ideally, the grades awarded by the algorithm should match those given by human graders for the same sets of criteria. To test this, results from the system were compared against results from humans.

6.1.1 Procedure

Teachers were asked to make judgments on several student drawings about whether they fulfilled individual criteria. For each criterion, teachers gave a grade of pass or fail; for details on how these data were collected please refer to appendix A.2. Then, the grading algorithm was run on the same criteria for the same drawings. In total 6 teachers made their own judgments on specific criteria for 48 drawings from 35 students. This resulted in 237 judgments spread across five different criteria: Domain Used, Is a Function, Critical Points, Function Followed, and Monotonicity. Unfortunately, no data was collected for the Derivative Matched criterion.

The results from the human graders were thought of as the correct answer for assigning grades. Ideally, the grading algorithm would agree with the human graders. In cases where the human graders do not all agree, establishing what the correct grade should be is not possible. To avoid throwing out too many data points, only

judgments where at least 5 out of 6 humans, $> 80\%$, agreed with each other were compared to the algorithm. From the original total of 237 judgments per teacher, only 191 were used to compare to the grading algorithm.

The drawings collected came from students of an introductory course at MIT, 6.01 - Introduction to EECS via Robot Sensing, Software and Control. In 6.01, students use an online system to submit homework problems. As part of an optional exercise in the fall 2015 semester, students were given 4 circuit theory questions consisting of a total of 15 graphing questions using the student interface and the grading algorithms described in this paper (see appendix A.1 for a list of the questions used). 35 students attempted at least one of the optional graph questions, and in total, 48 drawings were collected and shown to teachers to be graded.

6.1.2 Performance

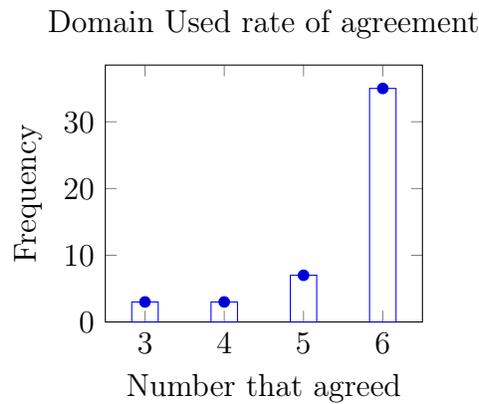


Figure 6-1: Number of Domain Used criteria that human graders agreed on

		Human		% agreement
		Pass	Fail	
Computer	Pass	40	0	100%
	Fail	0	2	

Table 6.1: Human/Computer agreement for Domain Used Criterion

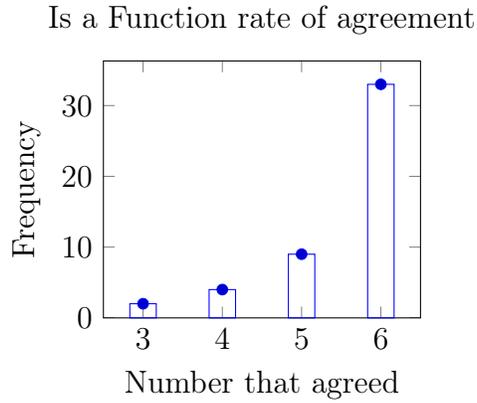


Figure 6-2: Number of Is a Function criteria that human graders agreed on

		Human		% agreement
		Pass	Fail	
Computer	Pass	40	1	98%
	Fail	0	1	

Table 6.2: Human/Computer agreement for Is a Function Criterion

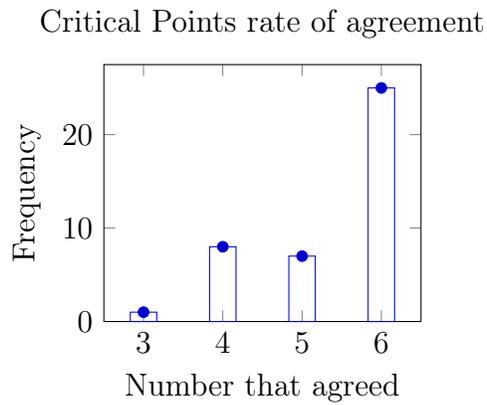


Figure 6-3: Number of Critical Points criteria that human graders agreed on

		Human		% agreement
		Pass	Fail	
Computer	Pass	23	0	100%
	Fail	0	10	

Table 6.3: Human/Computer agreement for Critical Points Criterion

Function Followed rate of agreement

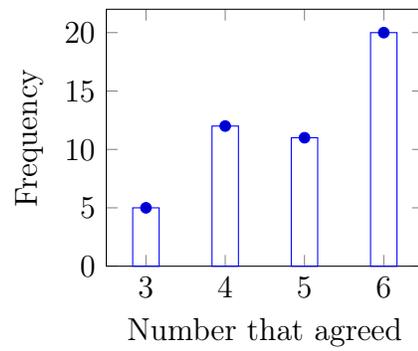


Figure 6-4: Number of Function Followed criteria that human graders agreed on

		Human		% agreement
		Pass	Fail	
Computer	Pass	11	0	77%
	Fail	7	13	

Table 6.4: Human/Computer agreement for Function Followed Criterion

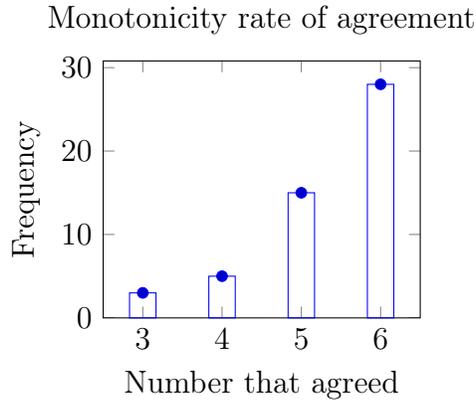


Figure 6-5: Number of Monotonicity criteria that human graders agreed on

		Human		% agreement
		Pass	Fail	
Computer	Pass	33	0	100%
	Fail	0	13	

Table 6.5: Human/Computer agreement for Monotonicity Criterion

		Human		% agreement
		Pass	Fail	
Computer	Pass	147	1	96%
	Fail	7	36	

Table 6.6: Human/Computer agreement for all Criteria combined

6.2 Teacher User Interface Testing

The motivation behind making the teacher interface was to make it easy for teachers to create graphing questions and harness the capabilities of the provided criteria. The following test tries to evaluate the success in accomplishing these goals.

6.2.1 Procedure

Participants were acquainted with the student interface by solving several sample problems using the student interface. Next, participants were asking to use the teacher interface described in chapter 5 to create two graphing questions chosen by the experimenter. For all the participants, this experiment was the first time they had seen the student interface and teacher interface. Once participants were satisfied with their choices, their selection were sent and saved to a server.

During testing, an attempt was made by the experimenter to give as little help as possible, and to let participants discover the interface for themselves. However, guidance was given as necessary when participants seemed particularly confused. In addition, the experimenter would point out features that went unnoticed for a while in order to get feedback on all features.

The data collected on participants included a video recording of them using the interface, their final submitted selections on the options selected for grading, and data from a survey given at the end of the experiment.

Full details of the procedure can be found in appendix B.

6.2.2 Teacher Results

Below are the results of testing on five middle school and high school math teachers.

Survey Results

Participants were asked the following survey questions. The results of the survey follow in figures 6-6, 6-7, 6-8.

- How easy was it to find the options you were looking for?
- How confident are you that the question you created would grade students fairly?
- In which settings do you think this kind of tool would help students learn?
- How would you describe the amount of your programming experience?

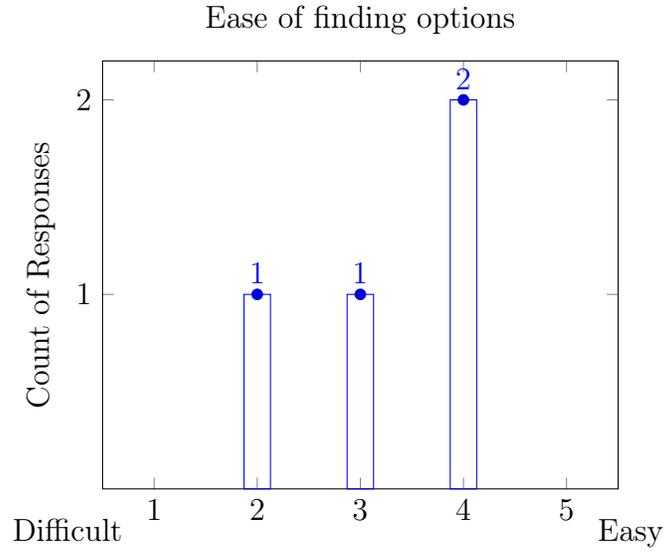


Figure 6-6

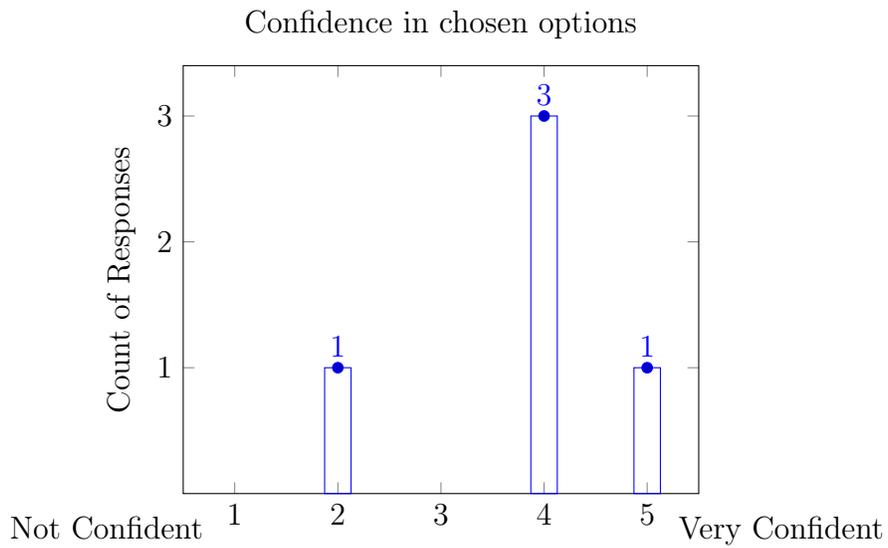


Figure 6-7

In which setting will this help students learn?

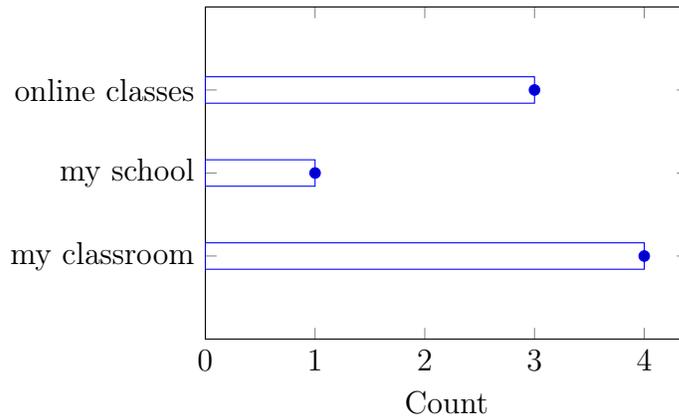


Figure 6-8

Programming experience

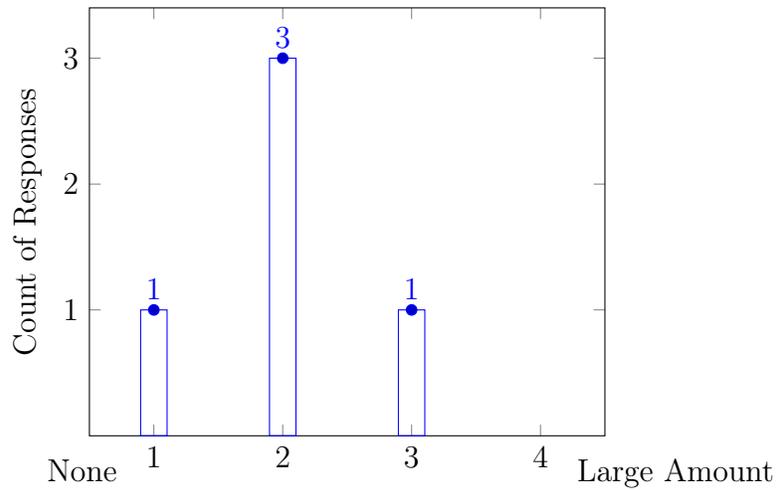


Figure 6-9

Important Observations and Feedback

This section examines some observations that were deemed most important by the researcher. A more complete listing of observations and feedback, including written comments, small bug fixes, and cosmetic changes, can be found in appendix C.

Overall Impressions Overall, participants were excited about the prospect of using this tool for automatic graph grading in their classrooms as they felt it was a

good complement for other online graphing tools they were using. Figure 6-8 shows that most teachers believe that the tool would help students learn in their own classroom and for online classes. In fact, one teacher wrote, “I thought this project was executed very well and with the slightest bit of training, could be effectively used by most teachers in my high school math department, or other departments that make use of graphing.” In particular, teachers appreciated that students had to draw in their answers in a freeform way, which differs greatly from other online tools they were familiar with. Another teacher wrote, “Once you explained the options to me and showed me around, it all made a lot of sense. I do think it requires explanation, though, so that teachers are clear on what students see, what teachers see, and what each of the criteria options mean.” The sentiment expressed in that quote was common for other participants who initially were intimidated by the interface. However, after a brief interaction with the interface and some explanations by the researcher, most participants said they felt much more comfortable with the interface.

The Criteria Model Participants had various levels of difficulty in forming a mental model of how the criteria worked. Three of the participants were able to guess how the criteria worked by experimenting with the inputs. However, the other two participants did not interact with the interface as they tried to understand the underlying model. They asked the researcher questions about how the criteria worked and only then were willing to experiment with some of the inputs. Their questions strongly suggested that they were trying to figure out a mental model for the system.

Participants complained that some of the word choices were confusing and intimidating which contributed to a reluctance to experiment with the inputs in Criteria Options. Several of the participants agreed that Criteria Options should be renamed Rubric since this was a word that is more familiar with teachers and consistent with the criteria model of grading. Another unfamiliar word was “monotonic”. Although there was additional text which could help describe the purpose of the different inputs, it could only be accessed by hovering over the inputs. This was difficult for the participants to discover, and, because participants had to wait for the help text to

appear, it was not used often, even after the participant were aware of its existence.

Another problem that presented itself was that the grade received from the student interface was difficult to interpret. In particular, it was difficult to understand what positive points were received since no textual feedback is received for getting criteria correct.

Visualizations The different ways of visualizing the criteria quickly became comfortable for all the participants, though some visualizations were more intuitive than others. The use of red and green to visualize the required and forbidden zones was immediately clear to all participants and it was common for participants to express their appreciation of the feature. The critical point visualization was also immediately clear to participants. The slope line and the bracket visualizations were more difficult to understand but generally made sense after an explanation.

Draw Testing Most teachers did not test their chosen criteria in the Draw Testing section at the bottom of the teacher interface. Teachers were usually reminded at least once during the experiment that testing was a possible action.

Two teachers consistently tested their choice in criteria in the Draw Testing section before they were satisfied to submit. Both teachers only tested the criteria by drawing the correct answer and never purposely drew wrong answers. In one example, a participant tested the default criteria—checking if the domain has been used and whether the drawing is a function—got 100% and concluded that they were done with the experimental task. The researcher asked the participant to draw a straight horizontal line, an incorrect answer, to see what grade they would receive. After receiving 100% for this answer as well since it passes the default criteria, the participant realized that he/she still had more work to do.

Criteria Selected Teachers had different ideas about which criteria to choose for each problem. The most popular criterion to use besides the two default criteria—checking that the domain has been used and that the drawing is a function—were the critical point criterion and the function followed criterion. Some teachers only used

one or the other, others would use both together. Many teachers did not think to use the monotonic or derivative criterion unless it was brought up by the experimenter.

Chapter 7

Discussion

7.1 Results

7.1.1 Grading Performance

Of the criterion types tested, 4 of them (domain used, is a function, critical points, and monotonicity) had an agreement greater than 98% with the human graders. This shows that the outcome expected of the algorithms by a human matches the name and description of the criterion.

The last criterion tested, the Function Followed criterion, had 77% agreement with human graders. Although it did not perform as well as the other criteria, it is encouraging that in all the cases where it did not agree with the human graders, the human grader gave a passing grade and the algorithm gave a failing grade. In other words, the algorithm is consistently too strict about what constitutes a correct answer. Relaxing the error margin and the pass fraction of the criterion may improve the agreement of the Function Followed criterion. It is worth noting that even the human graders had a tendency not to agree and over 35% of the data points had to be thrown out for this reason. It seems to be that deciding whether something is a function is highly variant depending on the person and the function being graded, so it is less of a surprise that it is difficult for the algorithm to agree as well. To see all the cases where the algorithm disagreed with the humans, refer to appendix A.4.

7.1.2 Teacher Tests

The experiments with teachers revealed that the criteria model is something that teachers can understand and manipulate intelligently. Subjects were excited about using the tool and in particular, they were excited about the freeform nature of student's input, which is not available on other online graphing teaching tools.

Although some teachers were able to guess how the criteria model worked, others struggled in forming a mental model so that they could intelligently choose criteria. Better word choice on menu items could help with forming mental models. For example the word "rubric" is familiar to teachers and conveys the job of Criteria Options. In addition, teachers were not familiar with the word "monotonic," so the words "increasing" and "decreasing" could be used instead. Finally, the teacher interface should have help text be more accessible (for example, by having a question mark at the end of each input which displays the help text when clicked).

It is also important that teachers are encouraged to make use of the testing functionality. During the experiment, many teachers did not test their criteria; those who did only tested for the correct answer. Some of the teachers may have changed their selected criteria based on testing. For example, one teacher who said they were confident in their answers on the survey only had one criterion for checking a parabola. The criterion used checked for three critical points on the parabola: the two x-intercepts and the maximum. Any function which passes through those three points, including sinusoids, cubics, and other functions, would receive 100%. Perhaps this is what the teacher intended, but the author believes that the teacher would have added more criteria if they had tested their selection.

The numerical grade presented to students can be misleading if not chosen correctly. For example, one teacher submitted a drawing which received 67%, which suggested to them that they were close to the right answer. However, this was not the case, as the drawing had only passed the domain used criterion and is a function criterion which are not usually indicative of a student's understanding. To avoid this problem altogether, it is worth considering replacing the numerical grade with receiv-

ing pass or fail. Students will still receive textual feedback, which can help them make progress to the correct answer but they won't be sent possibly confusing signals from the numerical grade. Alternatively, the weight of the Domain Used criterion and Is a Function criterion for the weighted average could be reduced.

7.2 Alternative Approaches

7.2.1 Grading Algorithms

In an earlier iteration, there was a grading algorithm which used curve fitting algorithms to help grade drawings. The teacher would have to define a function with different coefficients which the curve fitting algorithm would try to fit for. For example, a sinusoid function might be defined as $y = a \sin(bx + c) + d$. A curve fitting algorithm would then find the best coefficients and the R^2 value for a drawing. If the R^2 value is too large, that means that the student most likely did not draw a function with the correct class of shape. Otherwise, the coefficients can be compared to the answer coefficients and a grade assigned based on how close the coefficients are to each other.

Strengths and Limitations

This kind of criterion could be a powerful addition since it is able to recognize shapes regardless of scale and offset. There are good opportunities for feedback which could depend on which coefficients of the curve fitted drawing do not match the answer coefficients. In addition, multiple coefficients could be accepted as an answer if accuracy is not as important. This criterion was not used due to a lack of time and because of its complexity.

The limitations of this grading algorithm are strongly related to the limitations of the curve fitting algorithm used. For example, currently, curve fitting algorithms only really work for continuous functions which limits the situations in which this grading algorithm can be used. In addition, it may be difficult to explain how to use this kind of criterion to teachers due to its complexity.

7.2.2 Student Interface

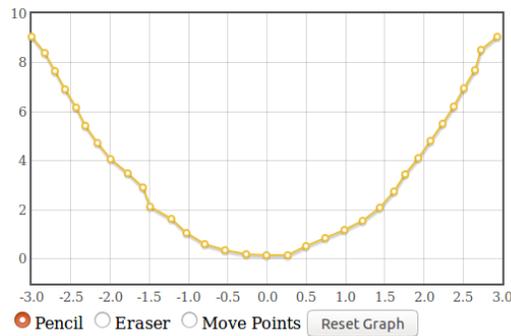


Figure 7-1: A parabola drawn with the Flot interface

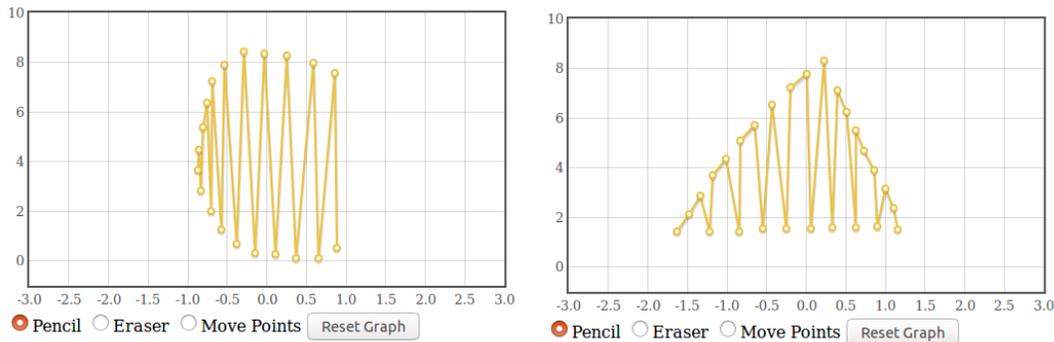


Figure 7-2: A C shape and a triangle drawn with the Flot interface

An earlier version of the student interface plotted individual points on the coordinates and drew lines between points using a JavaScript graphing library called Flot [1]. The interface forced all drawings to be a function which was demonstrated by drawing lines between points which are closest in the x direction. The interface included a pencil tool which added a new point when clicked (and added several points when clicked

and dragged), an eraser tool which removed points, and a move point tool which moved previously placed points.

Strengths and Limitations

Although some students expressed that they enjoyed using the Flot-based interface, this design was abandoned due to student feedback suggesting that the interface was unintuitive and hard to use for some users. A common concern voiced by students was that it was hard to know how precise the drawings should be. In an effort to increase the precision of their drawings, some students repeatedly clicked with the pencil tool to add more intermediate points and improve the smoothness of their drawing. One common positive piece of feedback was that it was easy to make linear graphs since it only required placing two easily movable points on the graph. Although the latest student interface has a line tool, it is not possible to manipulate a line that has already been placed like the Flot-based interface can.

Besides problems during the user experience, the Flot-based interface limits drawings to be functions where there is one y value for every x value. If a shape which is not meant to be a function is drawn, the interface makes it into a function giving the unintuitive results as shown in figure 7-2. The latest student interface does not suffer from this problem since it captures drawing input without extra manipulation giving the results shown in figure 4-1.

7.3 Future Work

During the teacher interface tests, less than half of the teachers tested their criteria selection. Those teachers who tested their criteria only tested them with the right answer. To encourage teachers to test their criteria and to help them test it effectively, a future iteration of the teacher interface could automatically generate samples of correct and incorrect drawings based on the criteria chosen. Automatically generating correct answers may show the teacher situations in which their criteria are

too permissive, and automatically generating incorrect answers may show the teacher situations in which their criteria are too restrictive.

More work could be done in determining what feedback to give students for different criteria. At the moment, a list of all the feedback messages of the failed criteria are sent back to the student as feedback. This could be overwhelming for a student and it may be more instructive to only send one feedback message at a time. Enabling teachers to choose what feedback to give based on the combination of criteria which passed or failed would be a worthwhile addition.

With the six criteria presented in this project, many kinds of functions can be graded, however, there are many other graphical problem domains which can be explored. For example, new criteria could be developed for grading geometry problems, for grading chemistry phase diagrams, or for drawing molecular structures. Some of the work mentioned in section 7.2.1 could be polished and improved for this use.

7.4 Conclusion

This tool adds the ability to ask automatically graded unstructured graphical questions online. The student and teacher subjects tested during the course of this project showed excitement for the interfaces and the grading tools. Initial tests show that the grading algorithms are able to grade drawings given criteria in a similar way to a human grader. Although limited in the graphical questions it can grade, the results of testing suggests that online homework systems can integrate graphical questions.

Appendix A

Grading Performance Details

A.1 Circuit Questions Asked to Students

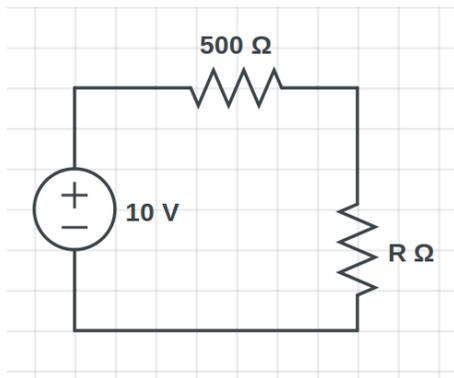
Students were asked to solve the questions shown below. The answers have been included.

A.1.1 Power Hungry

Calculating Power

The power consumed by an electrical component can be calculated by $P = IV$.

Series Power



Practice : What would be the power consumed by the 500Ω resistor if $R = 500\Omega$?

Answer : $.05W$

Find the equation for the power consumed by each resistor (500Ω and R) as you increase the resistor R . Graph the results for each resistor below as well as the total power consumed.

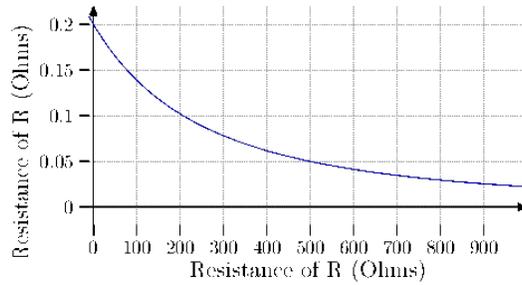


Figure A-1: Answer for power consumed by 500Ω resistor

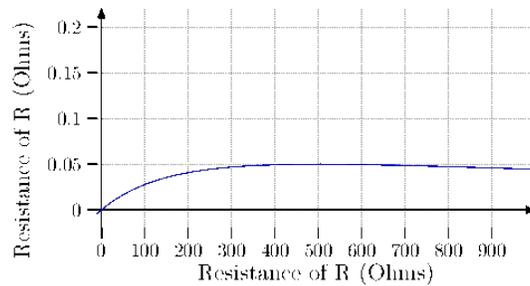


Figure A-2: Answer for power consumed by resistor R

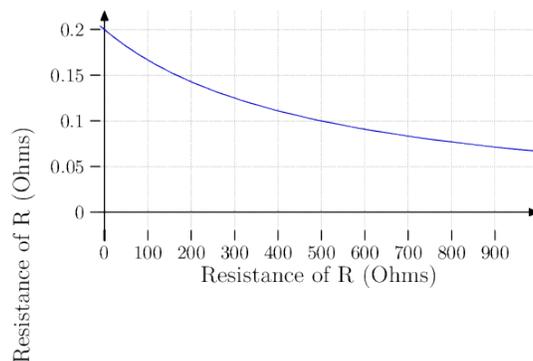
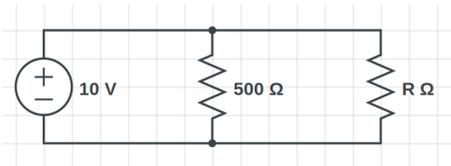


Figure A-3: Answer for total power consumed by resistors

Check Yourself : If you had a voltage source with a limited amount of energy, which circuit would run longer? A circuit with $R = 1k\Omega$ or $R = 1\Omega$? What about if

you were trying to make a heating element for a hot plate? Which circuit would give you more heat?

Parallel Power



Let's look at another circuit. Just like before, find the equation of power consumption for each resistor and graph the results below. Ignore what happens when $R = 0$.

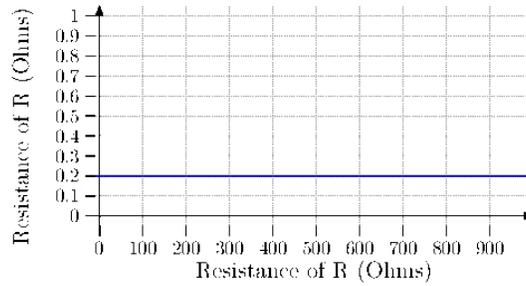


Figure A-4: Answer for power consumed by 500Ω resistor

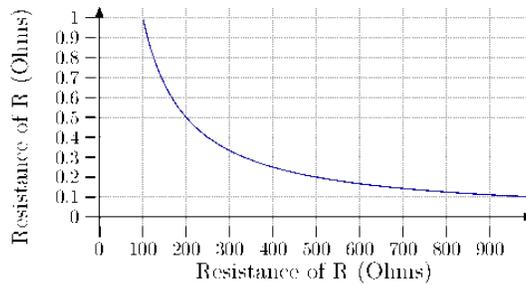


Figure A-5: Answer for power consumed by resistor R

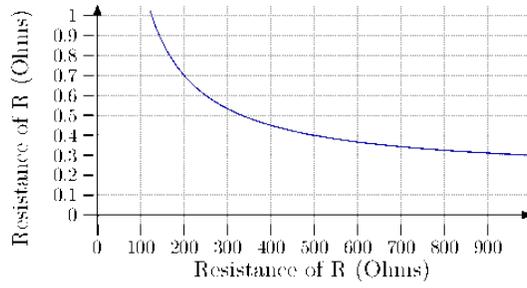
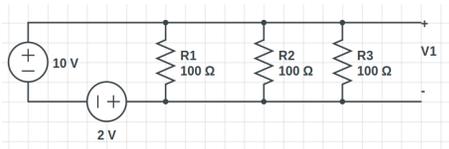


Figure A-6: Answer for total power consumed by resistors

Check yourself : Imagine you had some 500Ω light bulbs and placed them in the series and parallel circuits above? How would the brightness of the bulb change as you varied R ? How do you think your home is wired? You can think of the appliances in your home as resistors of different values.

A.1.2 Mixing Thevenins

Circuit 1



Draw the IV curve for the circuit which describes the relationship between I and V at the terminals.

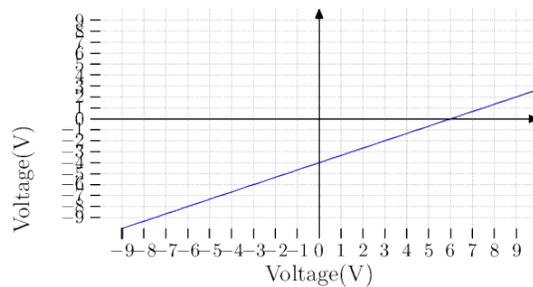
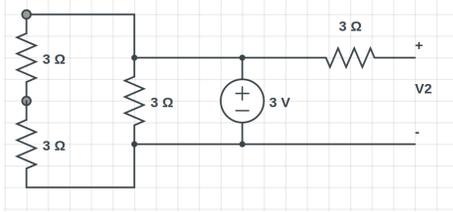


Figure A-7: Answer IV curve for circuit 1

Circuit 2



Draw the IV curve for the above circuit where the voltage source has a voltage of $3V$.

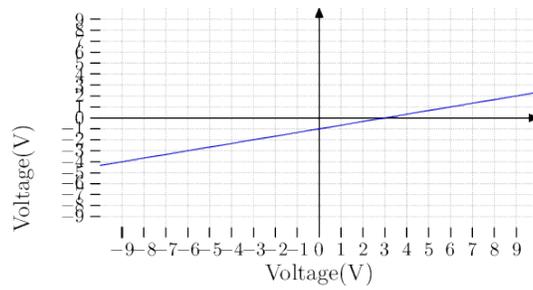


Figure A-8: Answer *IV* curve for circuit 2

Mixing

Let's connect the negative terminal of circuit 1 to the positive terminal of circuit 2 and look at the circuit from the unconnected terminals.

How would the IV curve of this circuit look? Think about how the slopes, x-intercepts, and y-intercepts of the first 2 circuits should interact.

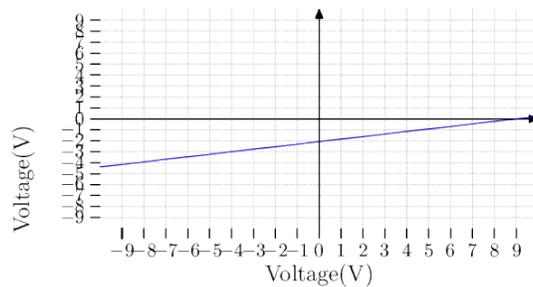


Figure A-9: Answer *IV* curve for circuit 3

A.1.3 Potentiometer Loading

Reference

In lab, we connected up a potentiometer as a voltage divider. Consider the following circuit where $R_p = 10k\Omega$. In this configuration, V_o and α are related as shown in the graph below.

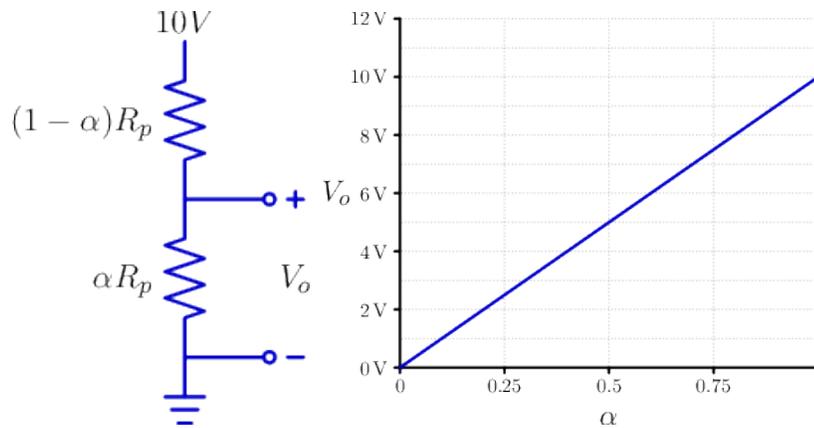


Figure A-10: Reference configuration and graph

For each of the following configurations, sketch the relationship between V_o and α in that circuit.

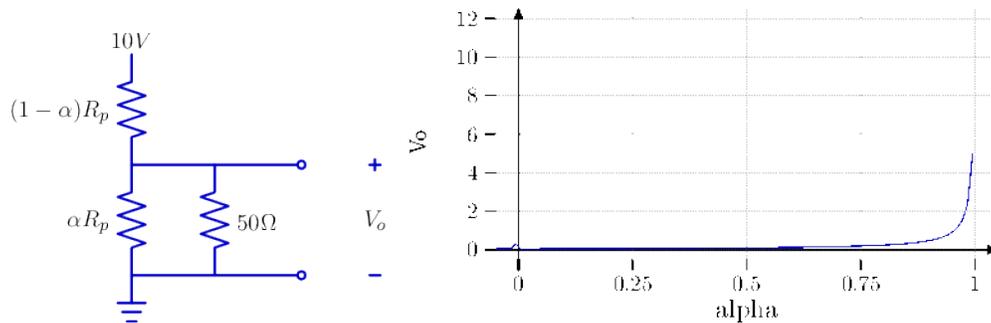


Figure A-11: Circuit configuration 1 and answer graph

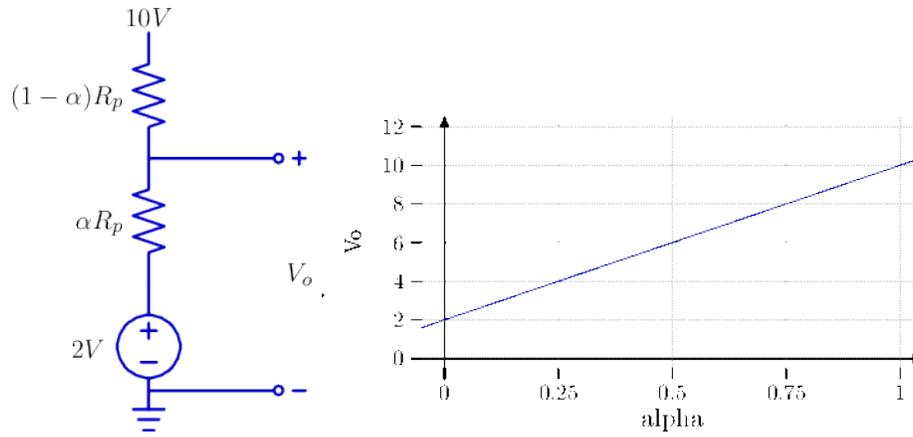


Figure A-12: Circuit configuration 2 and answer graph

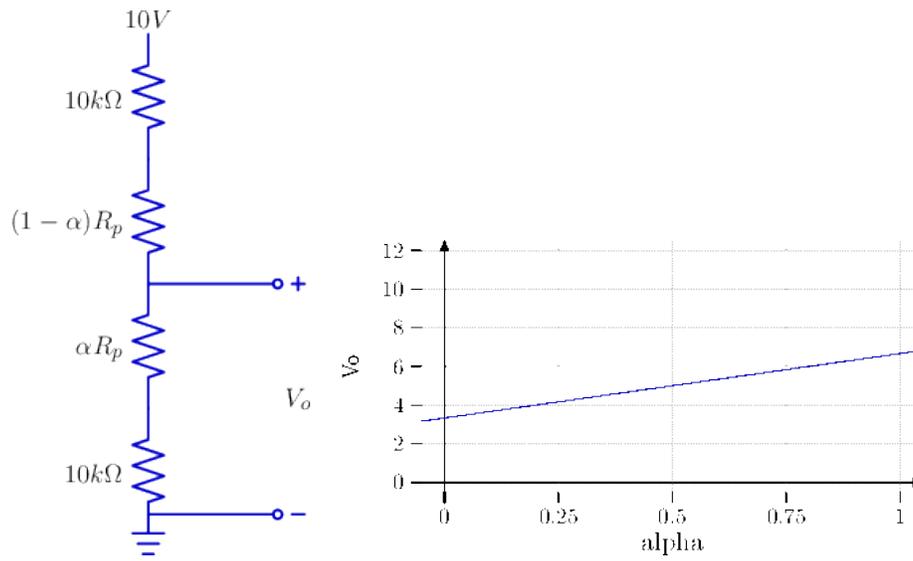


Figure A-13: Circuit configuration 3 and answer graph

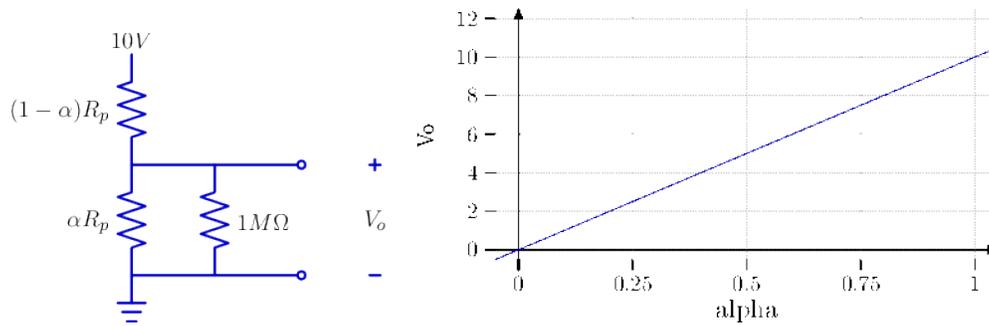


Figure A-14: Circuit configuration 4 and answer graph

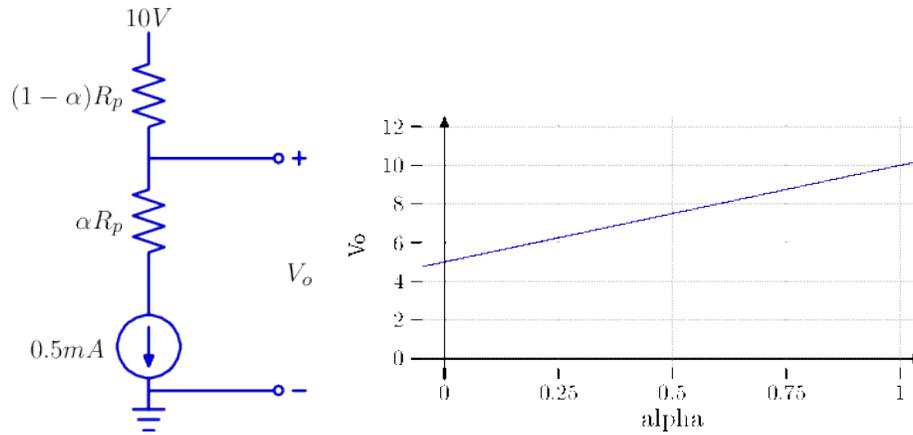


Figure A-15: Circuit configuration 5 and answer graph

A.1.4 Cats and Tart Pops

Sensing

Your cat, Nyan, keeps finding a way into your box of Tart Pops leaving you with nothing to eat for breakfast, lunch or dinner. You know Nyan doesn't like being sprayed with water and you're able to figure out a way to automatically spray water near your Tart Pops whenever the device has a $10V$ drop across it. The next thing you do is build a break beam sensor like so:

The resistor labeled R is a photo-resistor. Whenever Nyan passes between the light and photo-resistor (on her way to Tart Pops) the resistance goes down to 300Ω . Otherwise it stays above $2k\Omega$.

What is the voltage V_{bb} when Nyan is present?

Answer $V_{cat} = 2.5V$

What is the voltage V_{bb} when Nyan is not present?

Answer $V_{no\ cat} = 8V$

Amplifying

In order to spray Nyan, you amplify the signal (remember you need $10V$ for the spray to work) through a differential amplifier like the one below:

What is the equation of V_{spray} in terms of V_{bb} , V_{off} , R_1 , and R_2 ?

Let $\frac{R_2}{R_1} = k$, $V_{off} = V_o$, and $V_{bb} = V_b$.

Answer $V_{spray} = V_o + (V_o - V_b \cdot k$

Find the values of k and V_{off} which cause spray when Nyan is in the way and sets the value of V_{spray} to $0V$ in other situations.

Answer $V_{off} = 5.2$

Answer $k = 1.86$

Presenting

Graph what V_{spray} would look like if you increased V_{bb} from $0V$ to $10V$.

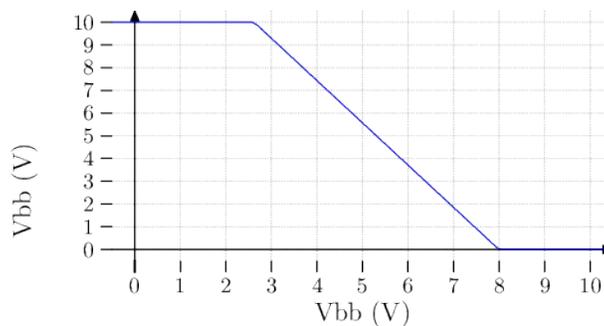


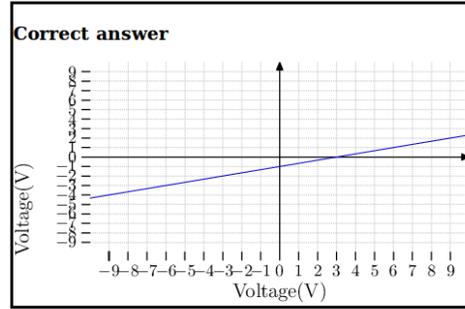
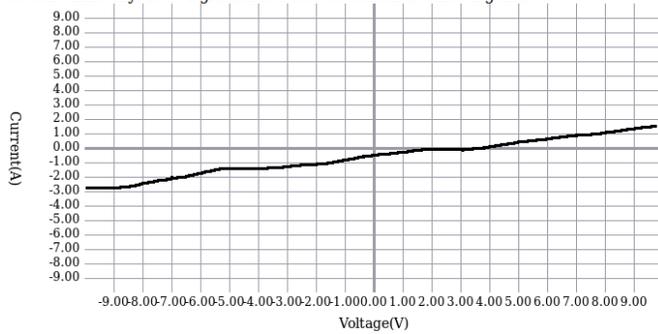
Figure A-16: Answer for graphing V_{spray} versus V_{bb}

A.2 Interface Used for Recording Teacher Criteria Judgements

The web interface showed a student's drawing in the top left corner, a graph of the answer in the top left corner, and the criteria they were asked to judge at the bottom. Teachers would click on a check box if they believed that the drawing fulfilled that particular criterion.

The teachers were not given many details about which parameters were used. Instead they were asked the following questions associated with each criterion:

You're grading Washington on question thevenins.q000001.
 You are martely. You've graded 0. Don't be scared to enter a grade.



1. Use of Domain: Does it cover a sufficient amount of x values?
2. Is a Function: Does it have a maximum of one y value for every x value?
3. Function: Does it match the function (as shown in answer picture)?
4. Monotonicity: is this a monotonically INCREASING (doesn't decrease) function?
5. Points: Does it pass through these points? [(3, 0), (0, -1)]

Submit(Spacebar)

[Go to next student](#)

[Go to next question](#)

Figure A-17: The web interface used to record teachers' judgements on criteria.

- Domain Used: Does it cover a sufficient amount of x values?
- Is a Function: Does it have a maximum of one y value for every x value?
- Important Points: Does it pass through these points? [Relevant points would appear here]
- Function Followed: Does it match the function (as shown in the answer picture)?
- Monotonically Increasing: Is this a monotonically INCREASING (doesn't decrease) function?
- Monotonically Decreasing: Is this a monotonically DECREASING (doesn't increase) function?

A.3 Survey

Students were asked for their comments on a survey. The questions asked are below as well as some students' responses.

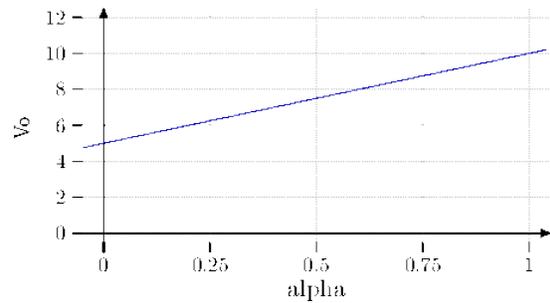
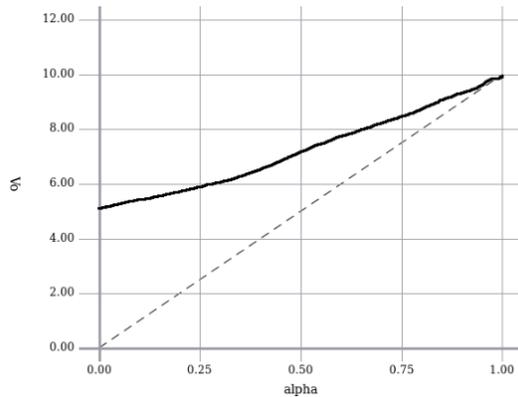
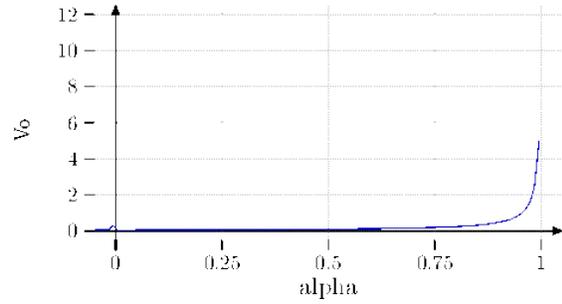
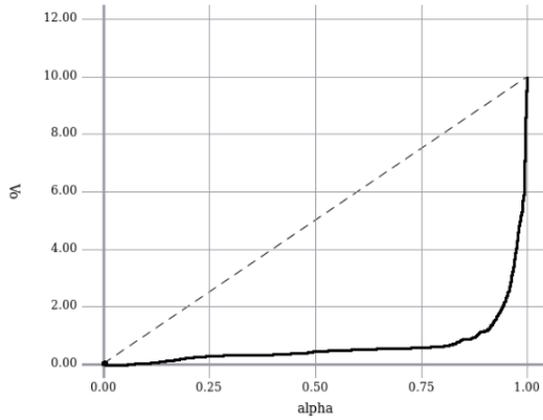
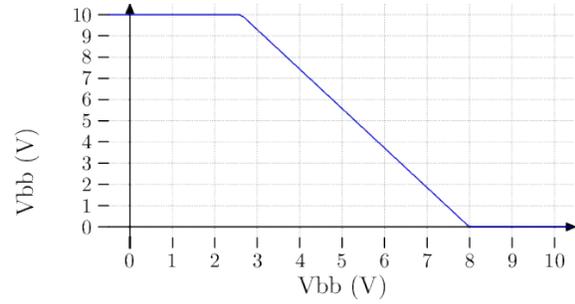
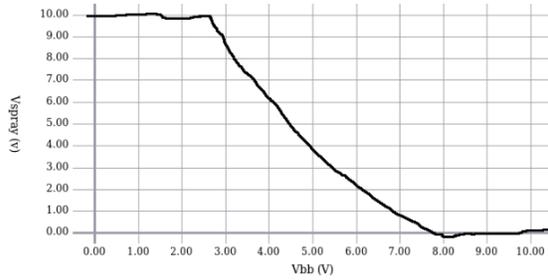
- Did you have technical problems drawing graphs? If so, can you describe what went wrong and what browser and computer you were using?
 - “Its hard to draw on a laptop trackpad”
 - “No”
- What did you like about the graph drawing problems?
 - “They were interesting”
 - “That it mostly works!”
- What did you not like about the graph drawing problems?
 - “They were good. Drawing was just difficult”
 - “Drawing straight lines is hard”
- Additional comments:
 - “I was impressed by the grading system.”
 - “In one of the Thevenin problems I got correct Vth but wrong Rth, and I still got 100%.”
 - (paraphrased) A student expressed interest in having drawing being easier and suggested tools which could help with drawing lines and drawing smoother curves.

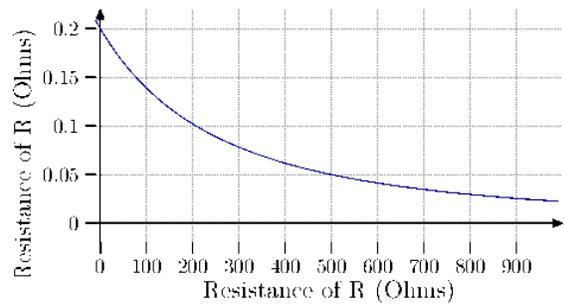
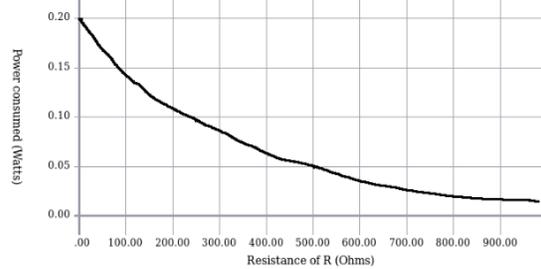
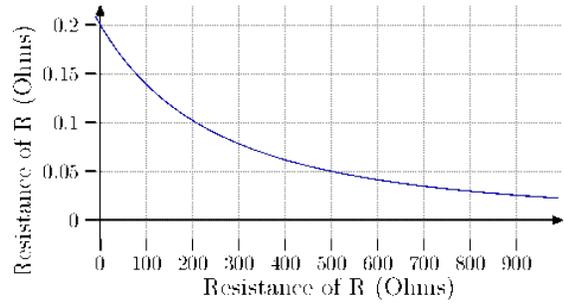
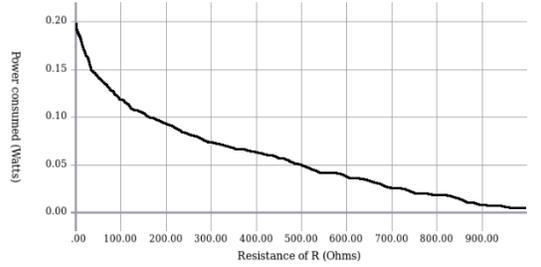
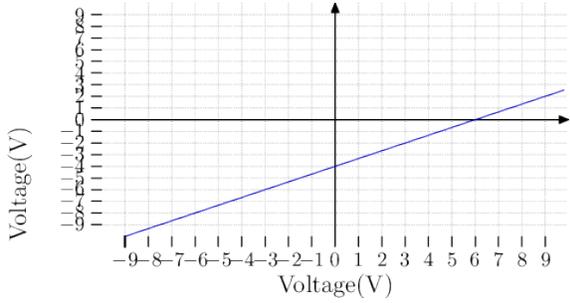
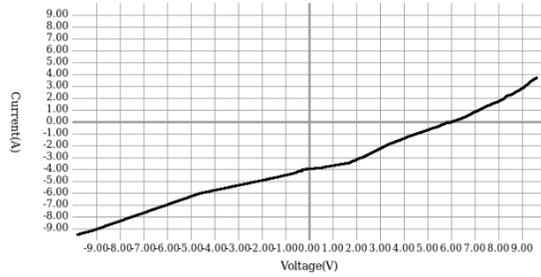
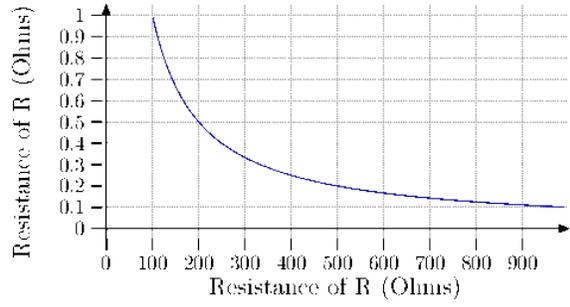
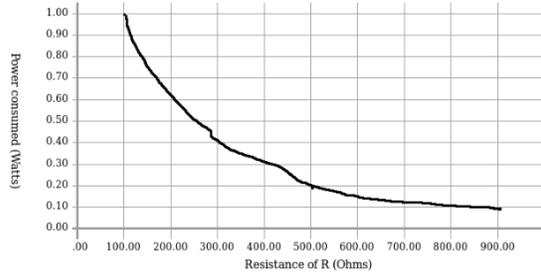
A.4 Graphs on which the algorithms and humans disagreed

Below are all the graphs where the grading algorithm disagreed with the majority of human graders.

A.4.1 Function Followed false negatives

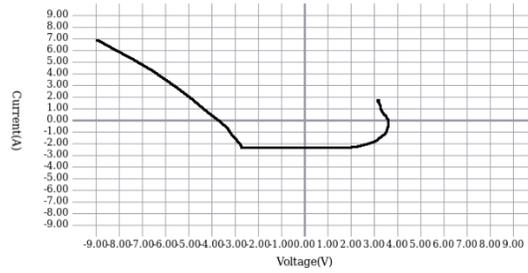
The following graphs all received a failing grade for the Function Followed criterion by the grading algorithm but received a passing grade for the same criterion by the majority of humans graders. In the following pictures, the student's drawing is on the left and the answer function it was compared to is on the right.





A.4.2 Is Function false positive

The following drawing received a passing grade by the grading algorithm for the Is Function criterion and a failing grade for the same criterion by the majority of human graders.



Appendix B

Experimental Setup for Teachers

B.1 Procedure

1. A consent form is sent to interested participants along with a one page writeup explaining the basic features of the interface (B.2). The teacher can read through these on their own time before meeting the experimenter.
2. The teacher and experimenter meet in person or via Skype and a screen capture video recording software is started. There is also an associated audio recording. The teacher is alerted to the fact that they're being recorded.
3. The teacher is provided some graphing questions as described in section B.3 to try on the student interface. They can do as many or as few questions as they like.
4. Once the teacher feels comfortable with the student interface, they are sent a link which contains some preselected graphing questions which they are supposed to select criteria for using the teacher interface. The problems are shown in section B.4.
5. At this point, the teacher is sent the link for the teacher interface. For all the teachers, this was the first time they had seen the teacher interface.

6. Teachers selected the display and criteria for the graphing questions and submitted their choices.
7. After 30 minutes or after finishing the tasks, the teacher was sent a survey with the questions shown in section B.5.

B.2 Experiment Overview Web Page

Teachers were presented with the following introductory text.

Your role in the project

There's a lot of great online tools for students to practice math problems on the computer. However, most tools usually boil down to multiple choice or entering some numbers or text and checking if the answer matches exactly. I decided to work on adding a new kind of problem which requires drawing graphs for an answer. A while ago my tool required an understand of coding to set up a graphing question which is not very teacher friendly. But, I think that it's just as important for students to be able to use the tool as well as their teachers. For that reason, it's important to me that you don't need a degree in computer science to make use of my research.

Please don't hesitate to give me feedback, both bad and good, as you're testing the interface. If something is confusing or unintuitive, let me know! I can probably redesign it to make it better but I need to know there's something wrong in the first place. If you like something, let me know too so that I don't remove it. Finally, I would like to thank you for taking the time to help out with the project.

Features of the Interface

Here's a quick rundown of what to expect from the interface.

Option Selection

On the left of the web page, you should see something that says "Display Options" and something that says "Criteria Options". Display Options determine what the students will see when they first see the problem. Criteria Options determines how a student's submissions will be graded.

Visuals

On the right of the web page, you should see something that says "Student View" and something that says "Teacher View". Student View shows you an example of what a student would see when they first saw the problem. You will not be able to draw on the Student View. Teacher View is meant to be a visualization of what you've selected in the Criteria Options.

Testing

If you would like to personally test out your criteria, there should be a button at the bottom that says "Draw Testing". In here, you can draw on a sample graph and submit your answer. After a delay (usually less than 10 seconds), you should get a grade and possibly some textual feedback.

The basics of Python Syntax for math

In certain criteria, you will be asked to input a function or a derivative in terms of x . The text you input should be legal in a programming language called Python. It looks mostly like math so don't be frightened if you've never programmed before. Below are some examples and some common pitfalls to avoid.

Common pitfalls

1. Whenever you want something to multiply together, you always need an asterisk. $5(5)$ is invalid. $5*(5)$ is valid.

- Exponentiation happens with a double asterisk (**). The caret symbol (^) does something completely different.
- Brackets are not allowed in mathematical expressions in Python. Use parentheses instead.

Examples

Math Notation	Python Syntax
x^2	<code>x**2</code>
$4(5 - x)$	<code>4*(5-x)</code>
$\sin(x)$	<code>sin(x)</code>
\sqrt{x}	<code>x**0.5</code> or <code>sqrt(x)</code>
$1 * 2 + 3 - 4/5$	<code>1*2+3-4/5</code>

B.3 Warm up Questions Web Page

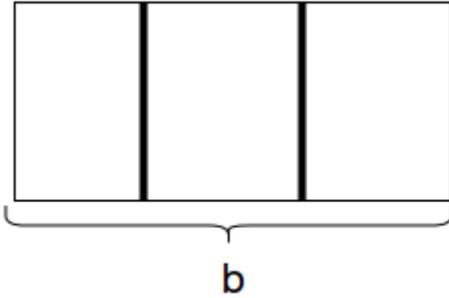
Teachers were asked to solve as many of the following questions on the student interface as they liked.

- Draw a monotonically decreasing function
- Plot $(x - 4)/2 + 1$
- Plot $x^3/3 - 5$
- Plot the derivative of $x^3/3 - 5$

B.4 Tasks for Testing Web Page

Teachers were asked to create the following questions using the teacher interface.

B.4.1 Rectangular Pen



For the following question, refer to the picture above. Imagine a rectangular pen split into 3 sections which uses exactly 50 meters of fencing. Draw how the total area of the pen changes as b ranges from 0m to 50m.

B.4.2 Sunday Driver

Jerry goes out for a 30 minute drive. Unfortunately, there's something wrong with his car and it can only accelerate at a constant rate of 1km/hour/minute. His brake time is not affected and is not necessarily constant.

This is how Jerry's drive goes

1. Jerry starts at rest.
2. For the first 5 minutes, Jerry accelerates. After that his acceleration drops to 0.
3. At minute 15, he sees a stop sign and begins applying the brakes.
4. At minute 16, his car has stopped.
5. At minute 17, he accelerates for the next 3 minutes.
6. At minute 28, he's close to his destination and slowly applies his brakes until the car stops at minute 30

Sketch the velocity of the car over the 30 minute period.

B.5 Survey

Teachers were asked to respond to the following questions on a survey sent after the experiment.

1. Your name
2. How easy was it to find the options you were looking for? (Scale from 1 to 5)
3. How confident are you that the question you created would grade students fairly? (Scale from 1 to 5)
4. In which settings do you think this kind of tool would help students learn? (In my classroom, In my school, In online classes, None-It would not help students learn)
5. How would you describe the amount of your programming experience?
6. General comments

Appendix C

Observations and Feedback from Teachers

C.1 Written Comments

These are the comments written by participants in the survey at the end of the experiment.

- “Once you explained the options to me and showed me around, it all made a lot of sense. I do think it requires explanation, though, so that teachers are clear on what students see, what teachers see, and what each of the criteria options mean. I don’t teach higher level math so it may make total sense to a Calc teacher. I found the verbiage a bit difficult to understand... but, again, it could just be because I teach middle school math. Great color feature and I also love the margin of error allowance which allows for unsteady hands as they draw the graphs. It would be truly amazing if a computer program could mark graphing. The closest thing I’ve seen is IXL.com. That’s probably your closest competition in terms of ease of interface use. Students get immediate feedback with IXL also. Good luck with this! It’ll be great once it’s refined.”
- “I think that the tool shows a lot of promise and if you were going to continue working on it could turn into something usable pretty rapidly”

- “Create a feedback tool for students when you release it into a classroom setting. This way they can reflect on any errors - or even explain why they think they are still correct! I’m looking forward to seeing the progress of the tool!”
- “I thought this project was executed very well and with the slightest bit of training, could be effectively used by most teachers in my high school math department, or other departments that make use of graphing. Has the potential to really shine on a touchscreen or interactive whiteboard.”

C.2 Student Interface

C.2.1 Positive Feedback

- Teachers appreciated the fact that the tool allows for students to enter arbitrary drawings and that it complements well with some other graphing tools they use.

C.2.2 Negative Feedback

- Teachers found the numerical grade received could be confusing and misleading. They were unsure what they received points for and felt that the grade did not accurately reflect how much understanding the drawing demonstrated.

C.2.3 Cosmetic Fixes

- The axes lines at 0 could be differentiated better from other axes lines.
- The eraser tool should show area that will be erased.
- Unnecessary zeros after the decimal point on axis number labels were distracting.

C.3 Teacher Interface

C.3.1 Positive Feedback

- Red/Green visualization made sense.
- Critical point visualization made sense.
- Bracket visualization was initially confusing but made sense after an explanation.
- After the 30 minute conversation, teachers were generally excited about using the interface in their class and felt that they would be capable of understanding and using the tool with minimal additional training.

C.3.2 Negative Feedback

- Teachers didn't think to test their own criteria.
- One teacher tested criteria that were too permissive and thought they had figured it out since they got 100%. After prompting they realized that an incorrect drawing also passed and began adding more criteria.
- Most teachers did not test their criteria choices. Those who did test their choices, only tested it with a correct answer.
- The derivative criterion was confusing for teachers. The idea of checking the slope or derivative was not clear until the experimenter explained how to use it.
- "Fraction Filled/Good" parameters for determining the pass fraction of some criteria is hard for teachers to understand.
- The "Criteria Options" model was very unclear for 2 out of 5 of the teachers. Until the 2 teachers were given an explanation, they could not be convinced to click on and explore the teacher interface. The explanation provided described

how a submitted drawing passes through each criteria independently and the weighted average of the grades is used.

- Teachers often weren't aware of all the choices in criteria that they had available. They would choose the first one that looked like it had potential and not look at the rest.

C.3.3 Cosmetic Fixes

- When axis is to be set by the student, teachers want to be able to set the initial minimum and maximum axis values.
- The selected minimum and maximum values for graph axes do not actually display as they are slightly off screen. The display should add a buffer zone so that minimum and maximum values are actually included.
- Error checking on parameters needs to be improved. For example, one teacher accidentally typed in a negative grade weight which results in the undesired behavior of possibly having negative scores.
- The scroll bar on "Criteria Options" is hidden by certain browsers and requires hovering over where it should be to find it. Making the scroll bar visible at all times will make it clearer that the "Criteria Options" section is scrollable.
- Some of the choices in wording was intimidating. For example "Rubric" would have been more familiar to teachers than "Criteria Options" and still convey the same idea.
- The "Add Criteria" button at the bottom of the "Criteria Options" is hard to find the first time.
- The help text which gives an explanation for a parameter is hard to find because it requires hovering over the parameter.
- Pixels is not a familiar word for many teachers.

- When changing criteria after doing a drawing test, there should be an indication that the grade shown is no longer a valid grade for the current criteria.

C.3.4 Individual Results

Teacher 1

Rectangular Pen Options X-axis goes from 0 to 25 with a step of 5 and a label of “b”. Y-axis goes from 0 to 80 with a step of 10 and a label of “area”. Dimensions in pixels of graph area not changed from default.

Two default criteria (Domain Used and Is a Function) and Function Followed Criterion. Function followed function is set to $(25*x-x**2)/2$ which is the correct answer. The error margin was set to 6 pixels. All 3 criteria examine the entire domain.

Sunday Driver Options X-axis goes from 0 to 30 with a step of 2 and a label of “minutes”. Y-axis goes from 0 to 6 with a step size of 1 and a label of “velocity (km/hr)”. Dimensions in pixels of graph area were not changed from the default.

The two default criteria (Domain Used and Is a Function) were included as well as critical points, monotonicity, and derivative criteria. The critical points checked that each bend in the graph, i.e. where the slope changes, was at the right location. The monotonicity and derivative criteria both checked that the initial slope in the first 5 minutes of the graph was monotonic and at the right slope.

Survey Response Easiness of finding options: 4 (easy)

Confidence in selection: 4 (confident)

Amount of programming experience: 2 (small amount of programming experience)

Teacher 2

Rectangular Pen Options X-axis goes from -2 to 25 with a step of 10 and a label of “xaxis”. Y-axis goes from 0 to 100 with a step of 10 and a label of “yaxis”. Dimensions in pixels of graph area not changed from default.

Two default criteria (Domain Used and Is a Function), Critical Point Criterion, and Function Followed Criterion. The Critical Point Criterion checks the point (0, 1) with an error margin of 11 pixels. Function Followed function was set to $((25*x - x**2))/2$ and error margin was set to 20. The criteria examine most of the domain.

Sunday Driver Options No response was submitted due to running out of time.

Survey Response Easiness of finding options: 2 (difficult)

Confidence in selection: 4 (confident)

Amount of programming experience: 2 (small amount of programming experience)

Teacher 3

Rectangular Pen Options X-axis goes from -1 to 30 with a step of 5 and a label of “b-value”. Y-axis goes from 0 to 100 with a step of 5 and a label of “area”. Dimensions of axes were changed to 500 x 500 (instead of default of 500 x 300).

Used 3 Critical Point criteria. Two were used for the x-intercepts at (0, 0) and (25, 0) with an error margin of 10 pixels. The last one was used for the maximum of the parabola at (12.5, 78.125) with an error margin of 25 pixels.

Sunday Driver Options X-axis goes from 0 to 30 with a step of 2 and the default label of “xaxis”. Y-axis goes from 0 to 20 with a step size of 2 and the default label of “yaxis”. Dimensions in pixels of graph area were not changed from the default.

The two default criteria (Domain Used and Is a Function) were included as well as critical points, monotonicity, and derivative criteria. The critical points checked that each bend in the graph, i.e. where the slope changes, was at the right location. The monotonicity criteria checked that the places with non zero slope were monotonic and the derivative criteria made sure those same places had the right slope.

Survey Response Easiness of finding options: 3 (neither easy nor difficult)

Confidence in selection: 4 (confident)

Amount of programming experience: 3 (moderate amount of programming experience)

Teacher 4

Rectangular Pen Options X-axis goes from 0 to 30 with a step of 5 and a label of “b”. Y-axis goes from 0 to 80 with a step of 5 and a label of “Area”. Dimensions in pixels of graph area not changed from default.

Two default criteria (Domain Used and Is a Function) and Function Followed Criterion. Function followed function is set to $-0.5x^2 + 12.5x$ which is the correct answer. The error margin was set to 10 pixels. The criteria examine most of the domain.

Sunday Driver Options X-axis goes from 0 to 30 with a step of 2 and a label of “Minutes”. Y-axis goes from 0 to 6 with a step size of 1 and the a label of “Velocity”. Dimensions in pixels of graph area were not changed from the default.

The two default criteria (Domain Used and Is a Function) were included as well as several instances of the function followed criterion. Each function followed criterion checked a linear part of the graph and did not overlap with each other.

Survey Response Easiness of finding options: 4 (easy)

Confidence in selection: 5 (very confident)

Amount of programming experience: 2 (small amount of programming experience)

Teacher 5

Rectangular Pen Options No response was submitted due to running out of time.

Sunday Driver Options No response was submitted due to running out of time.

Survey Response Easiness of finding options: 4 (pretty easy)

Confidence in selection: 2 (low confidence)

Amount of programming experience: 1 (no programming experience)

Bibliography

- [1] Flot. <http://www.flotcharts.org>. Accessed: January 15, 2014.
- [2] The learning online network with capa. <http://www.lon-capa.org>. Accessed: May 15, 2014.
- [3] Skulpt. <http://www.skulpt.org>.
- [4] How many students take distance learning courses at the postsecondary level? <https://nces.id.gov/fastfacts/display.asp?id=80>, 2013.
- [5] Rick Chartrand. Numerical differentiation of noisy, nonsmooth data. *ISRN Applied Mathematics*, 2011.
- [6] P.F. Mitros K.K. Afridi G.J. Sussman C.J. Terman J.K. White L. Fischer and A. Agarwal. Teaching electronic circuits online: Lessons from mitx's 6.002x on edx. Technical report, MIT, May 2013.
- [7] O.Zaslavsky G.Leinhardt and M.K.Stein. Functions, graphs, and graphing: Tasks, learning, and teaching. *Review Of Educational Research*, 60, 1990.
- [8] Adam J. Hartz. Cat-soop: A tool for automatic collection and assessment of homework exercises. Master's thesis, Massachusetts Institute of Technology, June 2012.
- [9] Sanjoy Mahajan. Observations on teaching first-year physics, December 2005.
- [10] Dhawal Shah. Moocs in 2014: Breaking down the numbers, 2014.
- [11] Valerie J. Shute. Focus on formative feedback. *Review Of Education Research*, 78, March 2008.
- [12] Simone Sturniolo. `tvregdiff.py`. math.lanl.gov/~rick/Software/TVDiff/tvregdiff.py.