



All Theses and Dissertations

2016-09-01

Creating and Automatically Grading Annotated Questions

Alicia Crowder Wood
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Computer Sciences Commons](#)

BYU ScholarsArchive Citation

Wood, Alicia Crowder, "Creating and Automatically Grading Annotated Questions" (2016). *All Theses and Dissertations*. 6099.
<https://scholarsarchive.byu.edu/etd/6099>

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

Creating and Automatically Grading Annotated Questions

Alicia Crowder Wood

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Masters of Science

Dan R. Olsen, Chair
Michael D. Jones
Kevin D. Seppi

Department of Computer Science
Brigham Young University

Copyright © 2016 Alicia Crowder Wood
All Rights Reserved

ABSTRACT

Creating and Automatically Grading Annotated Questions

Alicia Crowder Wood
Department of Computer Science, BYU
Master of Science

We have created a question type that allows teachers to easily create questions, helps provide an intuitive user experience for students to take questions, and reduces the time it currently takes teachers to grade and provide feedback to students. This question type, or an “annotated” question, will allow teachers to test students’ knowledge in a particular subject area by having students “annotate” or mark text and video sources to answer questions. Through user testing we determined that overall the interface and the implemented system decrease the time it would take a teacher to grade annotated quiz questions. However, there are some limitations based on the way students answered text annotated questions that would require a rewrite of the user interface and system design to decrease the grading time even more for teachers.

Keywords: education, machine learning, user experience

TABLE OF CONTENTS

1. INTRODUCTION	1
2. PRIOR WORK	10
3. OVERVIEW OF TEXT AND VIDEO ANNOTATED QUESTIONS	17
Text Annotated Question	19
Video Annotated Question	23
4. USER EXPERIENCE ISSUES - TEACHER CREATE	28
How to create a new question	28
Round 1 of User Studies	33
Round 2 of User Studies	34
Round 3 of User Studies	35
Round 4 of User Studies	37
Other UI Issues	39
Evaluating Creation of Questions	41
5. USER EXPERIENCE ISSUES - STUDENT TAKE	41
How to Take a Question	41
Round 1 of User Studies	45
Round 2 of User Studies	47
Round 3 of User Studies	49
Round 4 of User Studies	51
Other UI Issues	53
Evaluating Taking of Questions	54

6. USER EXPERIENCE ISSUES - TEACHER GRADE	55
How to Grade a Question	55
Round 1 of User Studies	58
Round 2 of User Studies	59
Round 3 of User Studies	63
Other UI Issues	64
Major Unresolved Issues	65
Evaluating Grading of Questions	71
7. MACHINE LEARNING	72
Text annotation	72
Results	73
Video	83
8. IMPLEMENTATION ISSUES	85
MVC and Angular Issues	85
Format to Upload	87
Implementing Text Highlights	89
Implementing Video Highlights	90
Google App Engine Issues	91
9. SUMMARY	92
10. BIBLIOGRAPHY	95

1. Introduction

Being able to analyze a student's learning and do it quickly is a necessity for teachers. Currently there are few good ways of solving this problem. The current solutions most teachers use are multiple-choice and open-ended questions, but this is insufficient. We have created a new type of question that is more open-ended than multiple-choice questions and is more quickly gradable than open-ended questions. This new question type is called an annotated question – one that can be created easily by teachers, taken easily by students, and quickly graded by teachers.

Multiple-choice questions are a quick solution to the problem of making easy to grade questions, but do not always engage students in the course material as much as possible. Multiple-choice questions provide a finite set of possible answers, which instead of promoting learning, promotes eliminating bad answer choices and guessing one correct answer. This is completely opposite the goal of quizzes and tests, which is to evaluate how much and how well students are learning course materials. In addition, some questions in a multiple-choice form may not be as applicable in some subject areas that are more subjective or have more than one “correct” answer. For example, the question shown in Figure 1-1 is an example of an inadequate multiple choice question in a history class. This question is searching for one right response, but in fact one choice does not fully answer this question. Most people may think “Thomas Jefferson” is the correct answer, but in reality, all five of these men were appointed by the Continental Congress to draft the Declaration of Independence [6]. This question is misleading and seems forced to fit into a multiple-choice form. One might think that adding other response options such as “all of the above” or “none of the above” solves this issue, however based on studies done by Popham, this type of question is considered invalid [10]. For students taking a quiz question with these types of options, if they find two correct answers they do not need to know about the validity of the other two answers, and can pick the “all of the above” option and still be correct. Also, in some subject areas, a multiple choice question does

not always work the best, but teachers of large classes end up using multiple-choice tests because of their time constraints.

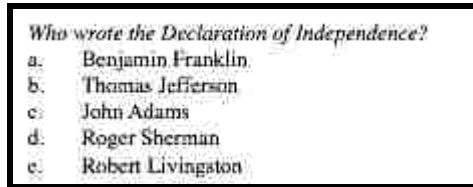


Figure 1-1. An example of a bad multiple choice question.

Open-ended questions are another option chosen by many teachers because they are more flexible and less restrictive than multiple-choice questions can sometimes be. The drawback is that open-ended questions take much longer for teachers to grade and take even longer to provide feedback when a student has answered a question incorrectly. For big classes with a large number of students, open-ended questions do not scale because they take too much time for teachers and TAs to grade and provide adequate feedback to all the students. This almost defeats the purpose of a more open-ended question – if a teacher chooses not to use multiple choice questions because of their restrictive nature, but then is not able to provide adequate feedback to students on open-ended questions, the students' learning is again hindered. Students can learn a lot by seeing what they did wrong on a test and being told the correct answer and how they may have misunderstood.

Also, not all types of questions are educationally stimulating as shown in Figure 1-2. Instead of grading recall or recognition questions, we have learned how to quickly grade and give feedback on analysis or application questions. An annotated question causes students to have to study or analyze a given the source and apply what they have learned by annotating the sources instead of merely recognizing or recalling an answer. On the question scale shown in Figure 1-2, an annotation question is at least better than a recall question, and it closer to a synthesis question.

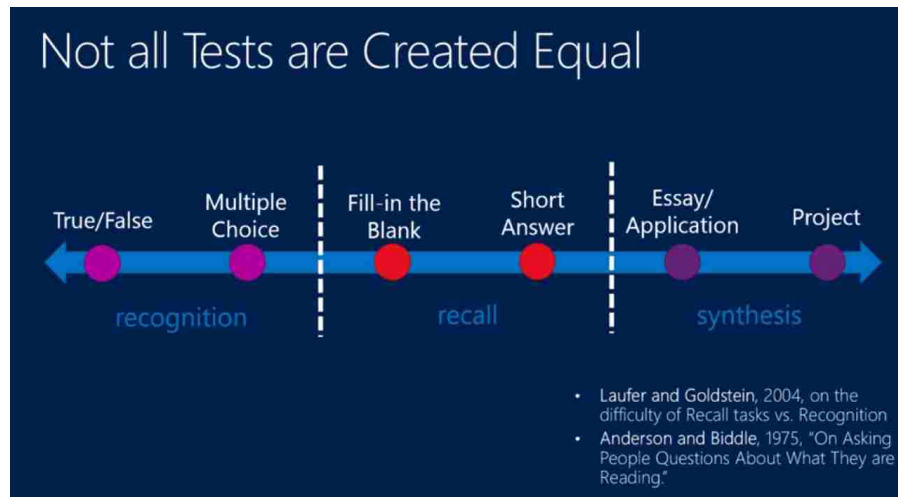


Figure 1-2. Six different types of questions and how they compare on the recognition, recall, synthesis scale.[5].

It will be beneficial to teachers and students to have some type of question that is more flexible than multiple choice, and applicable across more subjects, but still quickly gradable. Also, being able to provide useful feedback to students in a timely manner is crucial.

We need a good question that:

- Engages students with the course material
- Is applicable to many different subjects and classes.
- Can be graded quickly by teachers.
- Is able to give students good feedback.
- Is easy for teachers to create
- Is easy for students to take

For most teachers, having a solution that meets all these criteria would help save a lot of time, as well as allow them to use the time they already have more effectively. Teachers will be able to better prepare for classes by having better feedback of which questions were most commonly missed and the way that

students answered those questions incorrectly. Teachers and TAs will also be able to spend the time they would have spent grading, better contributing to student achievement in other ways – such as more one-on-one office hours, or time to be spent with smaller groups of struggling students [15].

Students will also be able to more quickly receive feedback and further understanding on concepts they may not have fully understood initially. This is especially critical in classes which build on a student's understanding of prior units or concepts. If the student does not fully understand the first concept after being tested on that knowledge, it sets them up for a higher rate of failure for the future concepts learned if they have not been provided with a solid base of understanding.

~~We believe~~ a possible solution to these issues is to create a new type of question along with an “automatic” grading system. This new question requires students to annotate sources such as a text document or a video. These questions are more open-ended than multiple choice but still allow teachers to quickly grade the responses. Annotating, which is marking or highlighting specific parts of a source can be applied to almost any subject area. A teacher can upload a document or video and ask students to annotate the source in a way that gives an answer to a question. This type of question uses an “automatic” grading system to minimize the time a teacher spends grading. This will decrease the number of responses a teacher has to grade per question and this training data will help auto grade the remaining questions through machine learning. We have created an intuitive and functional user interface that helps teachers more easily create these questions, helps students more easily take the quizzes, and help teachers more quickly and easily grade these questions.

One example of a multiple choice question that could be better constructed as a text annotated question is the multiple choice question shown in Figure 1-3 [16].

Given the above passage from To Kill a Mockingbird by Harper Lee, match the literary devices from the box below to its quote. Some devices may be used more than once, and some may not be used at all.

<i>a. diction</i>	<i>b. imagery</i>	<i>c. symbols</i>	<i>d. onomatopoeia</i>
<i>e. alliteration</i>	<i>ab. irony</i>	<i>ac. syntax</i>	<i>ad. Allusion</i>
<i>ae. simile</i>	<i>bc. Metaphor</i>	<i>bd. personification</i>	<i>be. repetition</i>

1. “Maycomb County had recently been told that it had nothing to fear but fear itself.”
2. “it was a tired old town”
3. “flicked flies,” “sweltering shade”
4. “Ladies...were like soft teacakes”
5. “streets turned to red slop; grass grew on the sidewalks”
6. “Men’s stiff collars wilted”
7. “slowly,” “ambled,” “shuffled”
8. “flicked”
9. “nowhere to go, nothing to buy,...nothing to see”
10. “Somehow it was hotter then: a black dog suffered on a summer’s day”

Figure 1-3. An example of a recall multiple choice question that would be better suited as an annotated question [16].

These questions help test simple recognition of literary devices given specific quotes from a reading passage. However, a teacher would more likely be interested in determining if a student can analyze a text, apply what they know about different literary devices, and determine which parts of the text demonstrate different devices. In order to create this analysis type question, we will create the following text annotated question.

QuizTeq Home | Example Course | Quiz 1 Answering 1 / 6 dummy1

TextAnnotatedQuestion (Q1) Points: 12 Q2 > Exit Quiz

In the passage below, highlight the following literary devices in the corresponding colors.
Not all devices may be used and some may be used more than once.

Maycomb was an old town, but it was a tired old town when I first knew it. In rainy weather the streets turned to red slop; grass grew on the sidewalks, the courthouse sagged in the square. Somehow, it was hotter then: a black dog suffered on a summer's day; bony mules hitched to Hoover carts **lurched** in the **lurching shade** of the live oaks on the square. Men's stiff collars wilted by nine in the morning. Ladies bathed before noon, after their three-o'clock naps, and by nightfall were like soft teacakes with frostings of sweat and sweet talcum.

People moved slowly then. They **ambled** across the square, **shuffled** in and out of the stores around it, took their time about everything. A day was twenty-four hours long but seemed longer. There was no hurry, for there was nowhere to go, nothing to buy and no money to buy it with, nothing to see outside the boundaries of Maycomb County. But it was a time of vague optimism for some of the people: Maycomb County had recently been told that it had nothing to fear but fear itself.













 diction	 imagery	 symbols
 onomatopoeia	 alliteration	 irony
 syntax	 allusion	 simile
 metaphor	 personification	 repetition

Figure 1-4. An example of an annotated question made from the multiple choice questions found in figure xx.

As seen in Figure 1-4, the same question found in Figure 1-3 can be made into a much better analysis type question using a text annotated question. Instead of giving the student multiple options to pick from, they are required to analyze the text and pick out which parts would be considered specific literary devices. Not only that, but students are required to read deeply to find out how many of each device is found in the text.

After the student answers the questions, if it were the multiple choice version, they would receive a list of the questions that were marked wrong without much extra feedback on why they were wrong or what were the correct answers. For an annotated question, there is more feedback provided to a student, and it is provided to the student much more quickly. As seen in Figure 1-5, the teacher can provide detailed feedback for each piece of the question. In the figure, the part of the passage highlighted in red, has a red “minus-sign” over it. This signifies that that highlight is “incorrect”. This piece of feedback is currently selected and on the right side there a dark bar selecting the feedback item: “-.5 Not a symbol...”. This is

the feedback associated with the red highlight. The teacher can select each highlight and attach specific feedback and points. This same feedback will then be automatically attached to any student response that incorrectly marks a similar part of the document.

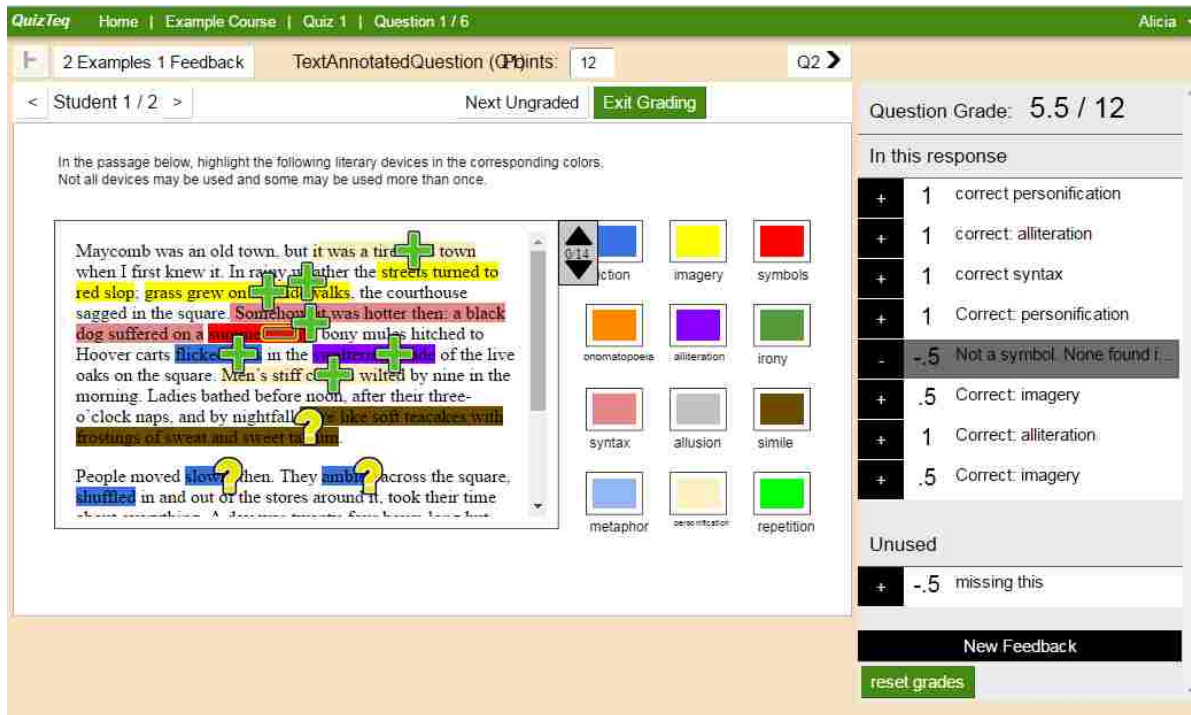


Figure 1-5. A screen shot of the teacher grading page for the question in Figure 4 that was answered by a student.

Another example of a question that would be better suited as an annotated question is the following:

Watch the clip from President Obama's inaugural address and list the different problems he plans to solve and how he wants to solve them.

Figure 1-6. An example open ended question to test a student's understanding of a video

This question reformed as a video annotated question as seen in Figure 1-7, would require the student to watch the same speech, answer the same questions, but, in the process would be able to guarantee a few improvements on the question from Figure 1-6.

First, a teacher would be able to grade the video annotated question much more quickly. A teacher could use this type of a quiz as a homework assignment, and before class the following day, know which students may not have understood a part of the video, or missed a critical concept discussed in the video. The teacher can take this information and prepare better for the next class period to be able to know which concepts were most commonly missed or misunderstood. By providing much quicker feedback, teachers would be able to provide small quizzes for students more regularly that give them better feedback on their progress every week or so instead of only a few times throughout the semester. If a student gets back the results to a short quiz the next day, instead of the next week, they will be able to notice what they did right or wrong, and remember better what the quiz was testing instead of receiving the feedback weeks after they took the quiz.

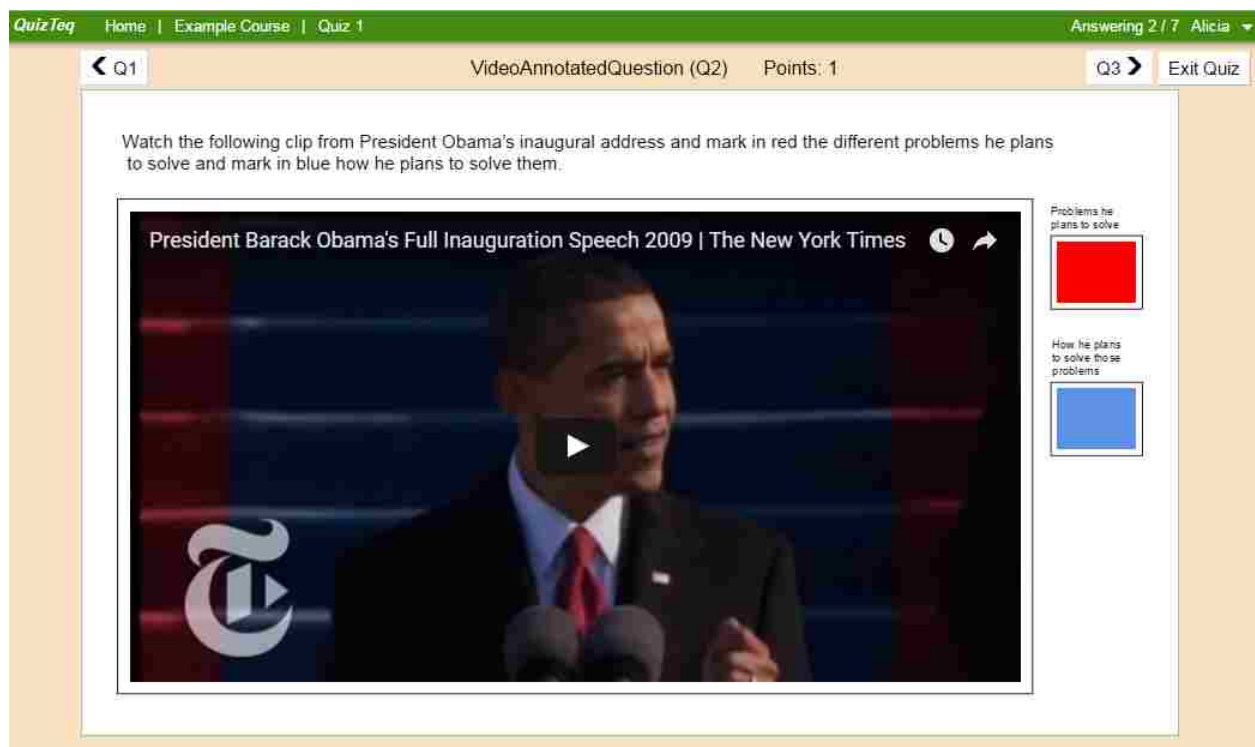


Figure 1-7. An example of how the previous open ended question could be converted to an annotated question

How have we measured success?

In order to determine if we have met our requirements, we performed multiple user studies with unique participants on each round of the studies. Before any implementation, we created PowerPoint mockups and performed usability testing with teachers and students on the initial user interface design. After multiple iterations with the PowerPoint mockups, we implemented the designs, and then began another round of user testing. To determine if we met our goal of providing an intuitive interface for teachers and students, we performed multiple rounds of user studies on each of the three main interfaces: teacher create, student take, and teacher grade. For each of these views, we continued to test the interface with users until there were consistently no major complaints about the interface and its functionality. At this point, we knew that we had an interface that most people could figure out and would be relatively easy for them to use.

To measure our success in terms of grading, we hand-graded approximately 5 text annotated questions and 5 video annotated question with approximately 50 students that answered each of the 10 question and then analyzed how much faster using the “automated” grading would be and how accurate it would be able grade the student responses in comparison to the gold standard hand-grading.

2. Prior Work

We used previous research in this field to find what others have done to solve similar problems to the problems we are trying to solve. The research used addresses one or more of the following requirements:

- Being able to quickly grade questions is an essential part of the question we have created for it to scale beyond a class of a few students.
- It is important to quickly provide students with feedback because this will help reduce time for teachers and facilitate better learning for students.
- The process of interactive machine learning is useful in constantly adding more training data, and in our case, teachers will constantly be grading and adding more training data.
- Finally, others have found certain steps to create the clearest and most intuitive user interface design, which principles we have tried to follow when designing our interface.

Quickly Grade

Being able to quickly grade is a crucial part of the system we have created. As mentioned above, some teachers currently use multiple-choice questions, and although they can be graded very quickly, multiple-choice questions may not be a very good test of knowledge [12]. Other methods employed by teachers are short answer, essay questions, projects, or labs. These questions are all a better test of knowledge than multiple-choice questions, but, the time to grade them is much longer, which may deter teachers from using these better questions.

“Divide and Correct” [3] was designed to grade and give feedback on a large number of short answer text responses. This application grades by initially grouping similar words together and then allowing teachers to mark each group as either correct, incorrect, or partially correct. It then allows teachers to

attach feedback to each group of answers that is then reused for each student who answered the question in that way. These ideas address many of the issues we have tried to solve, however, they simply grade short text answer questions. Short answer questions do not usually have much depth and we believe that having students interact more directly with course material may be more engaging than a short answer question would be.

Also, “Divide and Correct” does not refine cluster of answers based on grading that has already been done, it uses word similarity to create all the clusters initially. Allowing the teacher to grade as they go, and continually refine clusters helps teachers grade quickly then if all the different groups of answers were shown initially.

In 2013, Baumstark and Rudolph attempted to create a grading system where students could upload an assignment (in this case, assignments from computer network administration courses), which would then be graded remotely by the teacher and immediately provide students with a grade and feedback [1]. This new grading system was intended to grade a student’s ability to perform server or network configurations on a virtual machine (VM) and then submit that VM for grading. The size of the VM makes it impractical to submit that way, and the best alternative is to grade assignments in person. To solve this problem they provide a way to remotely access the students’ computer to be able to easily grade the responses. This method does not allow the teacher to automatically grade any questions, they have to grade each one by hand, and in our application the teacher does not have to hand grade each response individually.

In a study done in 2016, Wilcox found that having automated tools for grading software projects helped provide timely and accurate feedback to students so that they can understand and fix defects in their program [13]. In their approach they used an automation tool specific to grading programming

assignments based on a comparison of expected output with student output. The main issue with this approach is that they do not deal with any other types of questions besides programming questions and there is no clear way to extend their approach beyond programming. We hope to provide automatic grading for questions that are applicable to many different subject areas.

Another approach to solving the problem of spending too much time grading was experimented by Noguchi and Fujimura in 2015 [14]. They used the idea of peer grading to increase the speed of grading and the quality of feedback given to students. They used a schedule of having students submit an assignment one week and then students were required to complete the grading within the next week. They found that the peer grading does decrease the time teachers spend grading, and also found that students tend to give higher scores overall than teachers, because teachers may deflate scores because of poor quality and quantity of reports. Another potential reason for higher scores given by the students would be that the students may not know the complete correct answer as they are grading, which can lead to answers being incorrectly graded and not given appropriate feedback.

Provide Feedback

Providing useful feedback to the students after they take a quiz and being able to do that quickly is very important to the type of question we have created. Two approaches to this problems are dealing with providing feedback on a large number of student solutions, and using technology to help make grading and feedback faster than it would be on paper.

The first approach was researched by Glassman in 2014 [7]. Her approach, as seen in Figure 2-1, grades answers to more challenging open-ended questions where student responses are blocks of code. A strength of this grading approach is it addresses the issue that students can code things in many different

ways but still have a “correct” answer. In this method, the program will rename local variables that behave the same across different solutions with the same name and then through static analysis of the solutions, the program groups functionally identical solutions. The idea that students can answer questions in different ways and still have multiple correct answers is a strength we hope to build on. Glassman’s approach provides feedback to students and allows teachers to see a better high-level view of students’ understanding and misconceptions. The biggest drawback to this approach is only addressing programming questions. Our question type is more robust and is applicable to almost any subject area.

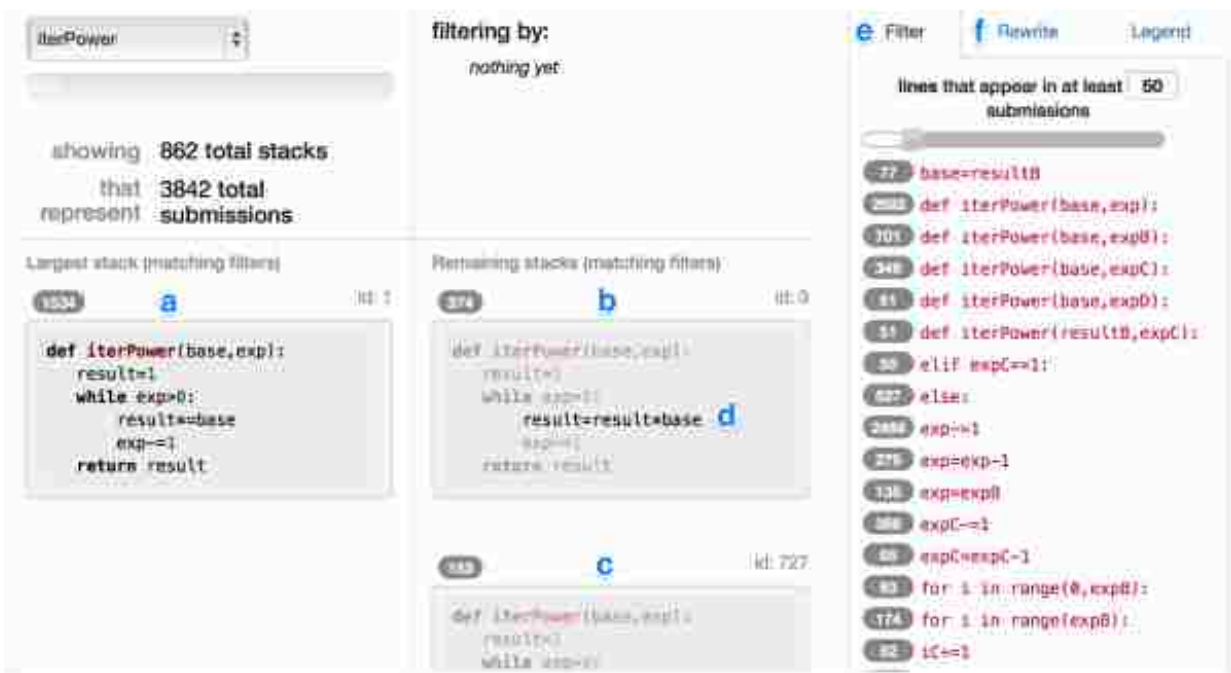


Figure 2-1. Glassman’s user interface to grade student solutions.

In 2008, Bloomfield and Groves’ desire was to give students better and quicker feedback through grading open ended questions electronically [2]. Their idea was to have students continue taking tests on paper, have teachers scan in all the tests at the end, and then use a tablet to grade all the exams. Using the tablets enabled teachers to be able to reuse feedback they already provided to another student and then manually attach that feedback to the test of a student who made a similar mistake. Their hope was that teachers

would be able to more quickly grade and provide feedback to students. This is a good concept, however, the research found that in using the interface to grade tests, initially there was a learning curve that prevented teachers from grading faster than normal grading. After learning the interface better, teachers were able to grade exams in the same amount of time or quicker than grading a paper exam. We have decreased the amount of time teachers spend grading by having the application automatically grade as many responses as possible and by providing an intuitive user interface.

Machine Learning

Machine learning is crucial to being able to reduce the number of answers a teacher grades and reuse the responses they do grade. Interactive machine learning takes normal machine learning a step further [5]. Normal machine learning follows the process of feature selection, training the data with a machine learning model, then classifying unknown data using this model. Interactive machine learning is similar, but now we add in an extra loop for adding manual correction. After the manual correction is added, the model is then re-trained with the new classifications [3]. See Figure 2-2 below for a graphical representation of this process.

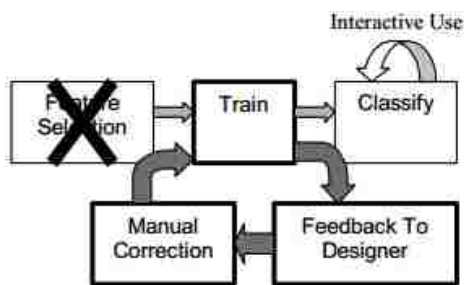


Figure 2-2. Interactive Machine Learning process.

Some of the benefits of interactive machine learning are that we can train and learn very quickly, the process can accommodate hundreds to thousands of features, and it allows for ten to hundreds of

thousands of training examples. With manual correction, more training examples are constantly being added so this is constantly helping refine and train the machine learning model to better automatically classify new answers. This directly applies to the way we automatically grade annotated questions. A teacher will initially grade one question and this provides training examples with different output classes or feedback in this case, the interactive machine learning uses this example to classify other similar ungraded questions and attach the same feedback. Then a teacher goes back and manually grades another question that was not classified based on the training examples provided so far. This then adds another training example to the learning model, which will grade even more similar questions.

Designing an intuitive user interface

Another problem we solved was being able to create an intuitive and effective user interface for teachers when they are creating questions as well as for students when they are taking a question. There has been a lot of research done about the process of creating a visually appealing and functional user interface.

One of these studies was by Couse in 1996, to develop a user interface and online documentation at the same time, while doing lots of usability testing along the way [4]. By creating the online documentation during development, they were able to make their user interface almost self-explanatory without much documentation. Many usability issues were caught and corrected before testing of the code because of early usability testing. They were able to produce an elegant, intuitive user interface for tasks of value to potential customers. This process of being more aware of the usability and how much users understand the process is a valuable aspect that we integrated into early stages of developing the user interface by doing preliminary usability tests on a PowerPoint prototype before implementing anything.

Mullet and Sano compiled a list of tips and tricks to creating a good user interface - basic design principles (the what and why), common design errors, and practical techniques (the how) [9]. They focus on suggestions based around elegance and simplicity, scale, contrast, and proportion, organizational and visual structure, and style. These principles are areas of focus we used in the final tests of our interface with different teachers and students.

In 2006, Juristo and Ferre created a tutorial attempting to integrate the usability process into the normal software development process [8]. Projects with a waterfall lifecycle approach are usually harder to convert to become more user centered. We applied this idea by doing usability tests along the way, making implementation changes, and then doing more usability testing, which we hope has decreased the overall development time.

Another branch of research on user interface design found that interactions with and expectations of technology differ across cultural contexts, and even concepts such as 'usability' are shown to be context dependent [11]. This means that user interfaces intuitive to one group of users appears counterintuitive to others. For our purposes we assumed that all teachers and students have some technical background and have used a computer before. However, we tested the user interface with as diverse of a population as possible, with students and teachers from different subject areas.

3. Overview of Text and Video Annotated Questions

To help teachers better test students' knowledge and be able to grade more quickly, we have created technology that will allow a teacher to create and grade annotated source materials. An annotated question is when a student is given some source (document or video) and asked to analyze that source by marking or "annotating" parts of the source to answer a question. For example, a teacher could create a new text annotated question to test knowledge about the Declaration of Independence and Thomas Jefferson. They could include a copy of the Declaration of Independence in the question, and then prompt students to highlight parts of the document that show how Jefferson's views expressed in the Declaration of Independence may contradict how he actually lived, specifically on the issue of slavery. See Figure 3-1 for an example question.

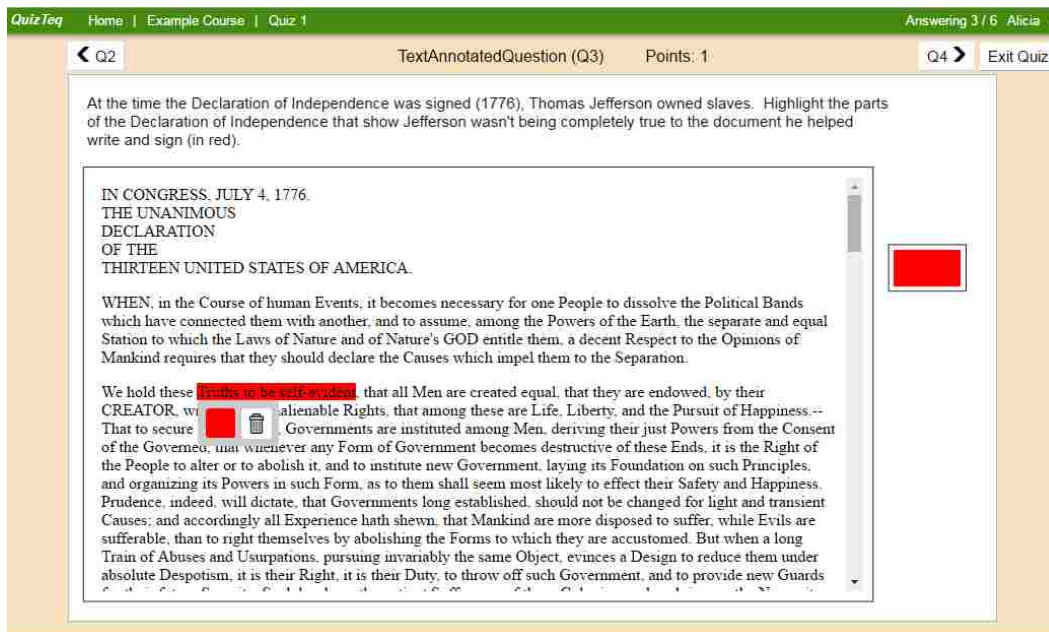


Figure 3-1. Student view of a question testing analysis of the Declaration of Independence.

Another type of an annotated question uses video to test student knowledge. A teacher can add a video of the current President's State of the Union address, as seen in Figure 3-2, and ask a student to record points in the video where the President's views are different to his party's views on a certain issue



Figure 3-2. Student view of a video annotated question.

In order to effectively and quickly grade annotated questions, we use interactive machine learning. This needs to be interactive because as a teacher grades they will constantly be adding data that will help our program grade quicker. Interactive machine learning helps us learn how to grade and give feedback on hundreds of student responses quickly and accurately, with very little work on the teacher's part. We follow the process found in Figure 3-3 to complete this.

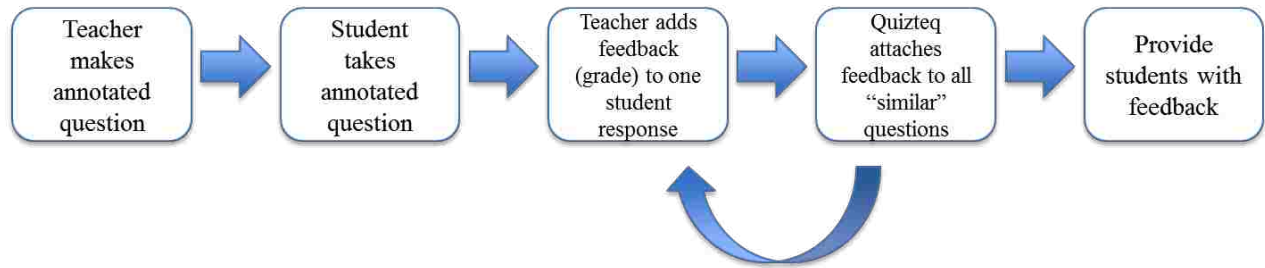


Figure 3-3. Complete process of creating, taking, and grading quizzes.

Text Annotated Question

The first type of annotated question is a text annotated question. A text annotated question is where a student will be asked a question and then will mark or highlight a given document (or documents) in the places that answer the given question.

1. Initially a teacher creates a question. The process of creating a question is applicable to many different subject areas. They will be able to upload a document to be annotated, and then will be able to add a description of the question as shown in Figure 3-4. After a teacher creates enough questions for a quiz, they can publish the quiz for students to take.

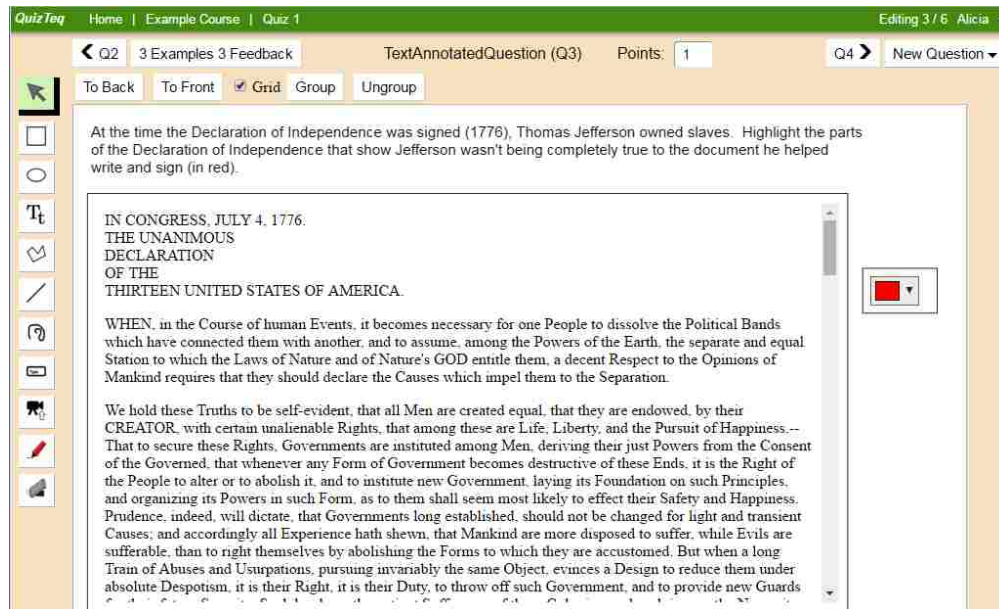


Figure 3-4. Teacher view of creation of a text annotated question.

2. After the teacher has published the quiz, the student will have access to view and take the quiz as shown in Figure 3-5. The student follows the directions and highlights the sections of the given source that they feel answers the question. There may be multiple colors or a single color used to highlight. Although this may not seem like a challenging task, for a student who is new to this program and this type of question, knowing how to answer the question may not be very clear. Making this process clear and intuitive for students was our goal. We do not want to test how well a student can figure out a new program, we solely want to test their knowledge of a particular subject area.

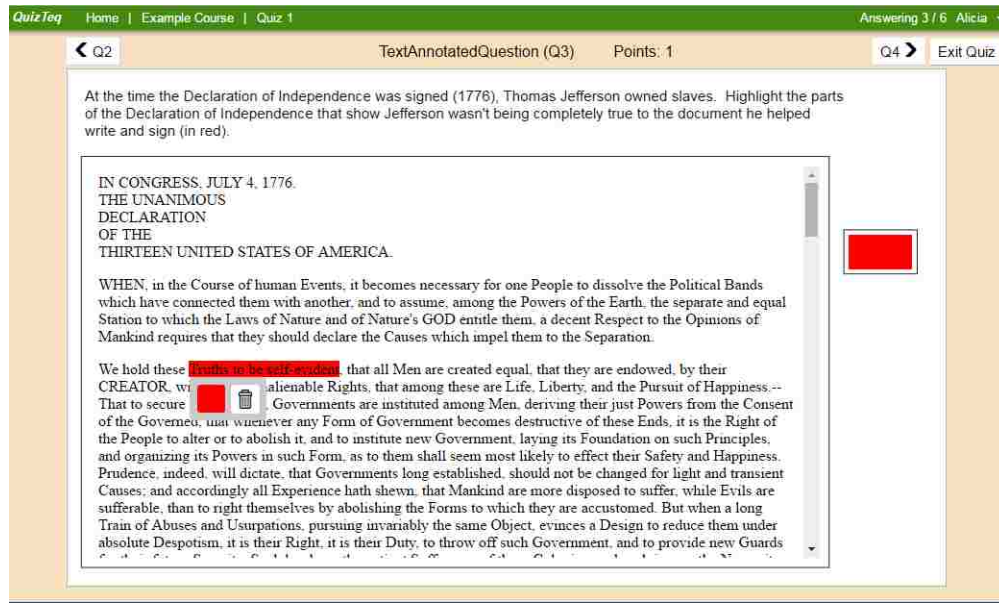


Figure 3-5. Student view of taking a text annotated question.

- After all the students have submitted their quizzes, the teacher will have access to grade them as seen in Figure 3-6. The teacher will start out with the first question on the first quiz. They will look at the first student's answer and decide which annotated parts are either correct, incorrect, or missing. After the first question has been graded and feedback attached, Quizteq will begin using interactive machine learning to "automatically" grade other similar responses to that question. Machine learning will take the responses that have been graded and use the corrections the teacher has made, to apply to other questions that were answered similarly. In this process, it will generalize the feedback that will be provided for a certain answer if the answer is close to another response that has already been graded. So if a different student gave a similar answer as the first response the teacher graded, machine learning will "automatically" grade the similar answer and the teacher would not have to grade any "duplicate" or similar responses.

The teacher will then be given another answer to grade (different response from the first). The same process continues, with Quizteq “automatically” grading similar problems after the teacher submits the grading and feedback for that response.

Due to each student being different and understanding a question prompt differently, one challenge that has come up in this area is the level of granularity each student answers questions. Some students approach questions with a broader approach, while others answer questions very specifically highlighting each part of an answer individually. We will address this issue in later chapters.

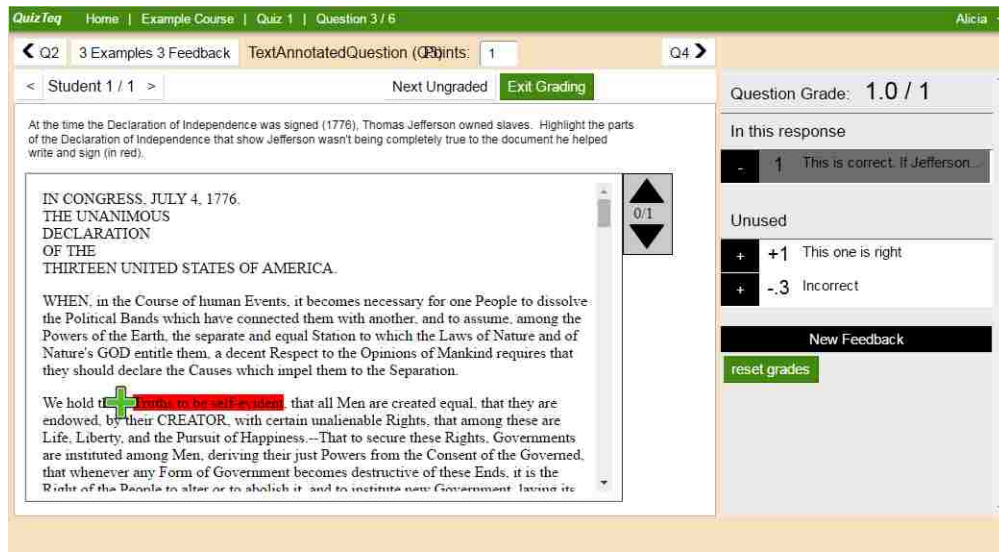


Figure 3-6. Teacher view of grading a text annotated question after student has taken quiz.

4. After the teacher completes the grading process, students view feedback and the grade they received on their quiz in a view similar to Figure 3-7. Each student can review their quiz and the

highlights that were incorrect will be marked with a red minus sign, the highlights they did not get any points on will be marked with a blue zero, and the correct highlights will be marked with a green plus sign. When a student clicks on a particular highlight, feedback given by the teacher will be shown. This feedback is extremely helpful to students and quickly facilitates learning. By seeing in context how or what they did wrong, a student will be able to know what the correct answer should have been and learns through this process. In Figure 3-7, the student highlighted the correct location for the question, as seen by the green plus sign, and on the right side, a student can view why they got the question correct in the corresponding feedback box.

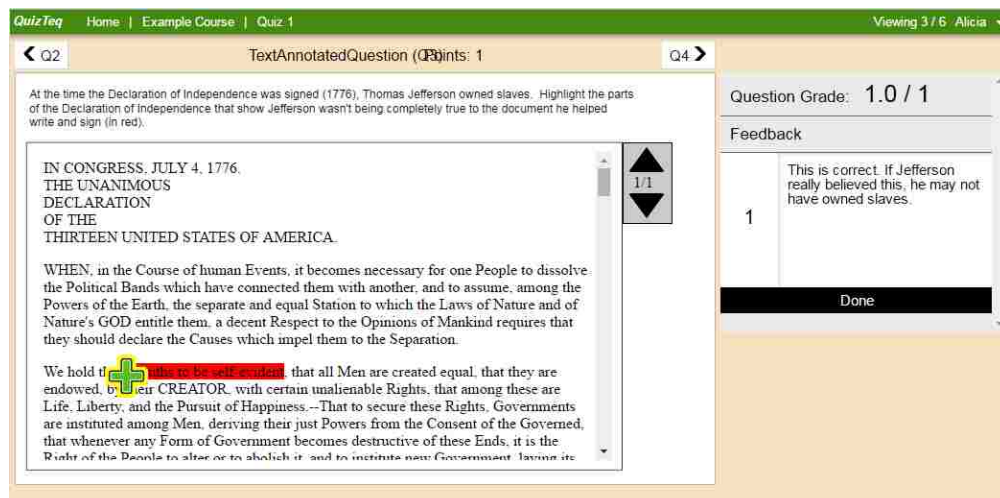


Figure 3-7. Student view of accessing feedback for a text annotated question after teacher has graded the quiz.

Video Annotated Question

The second type of annotated question is a video annotated question. Video annotated questions are useful for teachers who want students to watch video lectures or a video of a historical event. In the example below, the teacher provides a video lecture and to check that students have watched the video,

they provide multiple questions along the way where students “mark” what point in the video answers each question. This is an excellent way to check if a student is watching the assigned videos and if they are learning something from those videos as well.

1. Initially a teacher creates a new video annotated question by uploading a video for the students to watch as shown in Figure 3-8. They also add a question description specifying which parts of the video students should mark for each answer. Again, similar to the creation of a text annotated question, for a teacher creating this type of question for the first time it may be a challenge and they may not understand all the necessary components required for the question. We will address this challenge in a following chapter. After the creation of all questions for a quiz are complete, a teacher will mark the quiz as being in the “take” mode for students to be able to start taking the quiz.



Figure 3-8. Teacher view of creation of a video annotated question.

2. Once the quiz is in “take” mode, a student may begin to take the quiz as seen in Figure 3-9. They watch the video and during or after watching, click on the color button on the right at the time of the video where each answer is located. If a question uses multiple colors to mark different points, there will be more colored buttons on the right for the student to click on. If they make a mistake and accidentally click in a spot they did not mean to, they can hover over the highlight and click the trash-can icon to delete that annotation. Taking a quiz with these video annotated questions may not be very intuitive for a student new to the program. This is an issue we addressed in creating an intuitive interface for the student.



Figure 3-9. Student view of taking a video annotated question.

3. After students have completed all the questions for the quiz, the quiz will be made available for a teacher to grade. As seen in Figure 3-10, the teacher will look at the first question and award points based on the quality of each parts of the student answer. Along with the number of points given, the teacher can quickly type in feedback that will be attached to that response to help the

student understand what mistake they made or what they did correct. After a teacher grades one response, Quizteq takes these answers tied to the feedback given and automatically grades any questions that can be graded similarly based on the answers given. This will also tie feedback to all the questions graded that way. This allows the teacher to enter feedback once for each mistake made, but then each student can see exactly how he or she answered the question incorrectly. We want this interface to make grading quick and easy for teachers to be able to provide useful feedback to students.

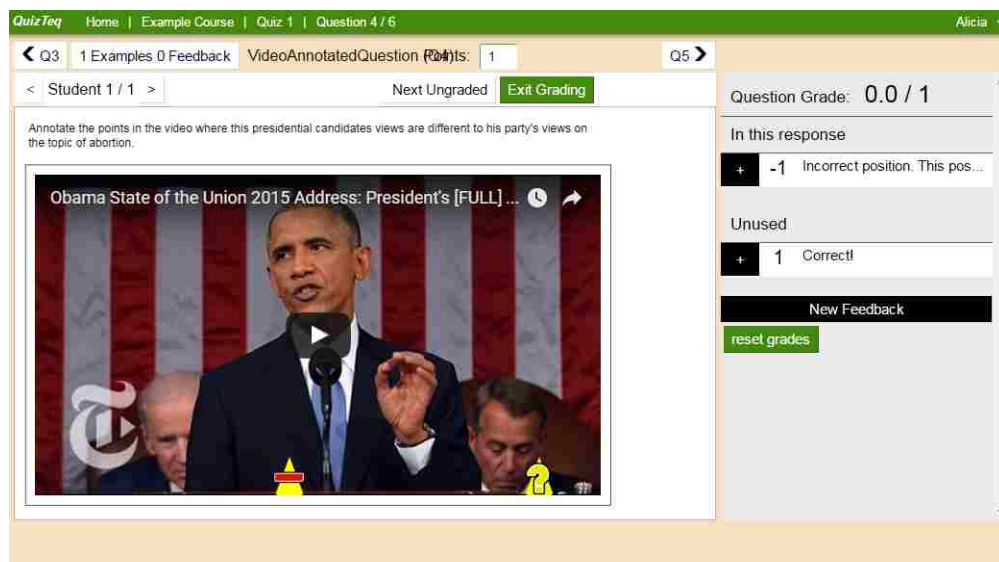


Figure 3-10. Teacher view of grading a video annotated question after student has taken quiz.

4. After the grading is completed by the teacher and Quizteq, students may view feedback on their answers in addition to the grade received as shown in Figure 3-11. They will be able to go back through their quiz and view which questions (or parts of a question) were correct or incorrect. For any answer, they will be provided with feedback to be able to know if an answer was correct or how it was incorrect. In Figure 3-11, the first marking is covered with a red minus-sign indicating that the student answered the question incorrectly. On the right, is a feedback box holding a description of what that position in the video did not correctly answer the question.



Figure 3-11. Student view of accessing feedback for a video annotated question after teacher has graded the quiz.

An annotated question can take multiple forms, but these are two examples to better understand the potential power of an annotated question. It helps the student analyze and really think about a document or video instead of just eliminating distractor answers on a multiple-choice test.

4. User Experience Issues - Teacher Create

For a teacher to create a quiz question for students with new software and two new types of questions, there are some three issues. First of all, many teachers may not know what components or widgets are needed for a new text or video annotated question without some guidance. Next, teachers may not know how to insert those components onto the page (for example: how to add a document or video into the question). Finally, there may be confusion when using a new teacher software and learning how to navigate all the parts quickly and easily.

How to create a new question

There are two new types of questions we will allow a teacher to create. These two types, as mentioned previously, are a text annotated question and a video annotated question. For a text annotated question, a teacher chooses a document they would like to test their students on, then the teacher uploads that document into the media library. After uploading the document to the media library, they will add that document onto the question canvas as shown in Figure 4-1.

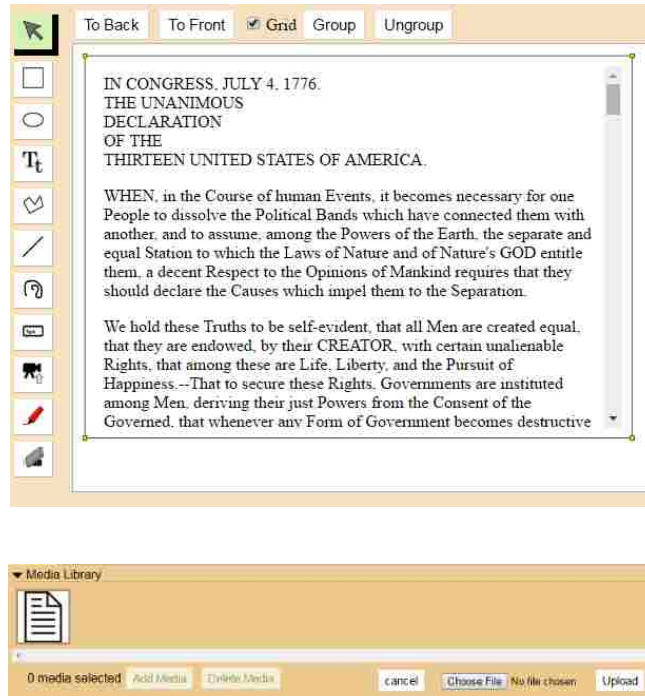


Figure 4-1. The media library where documents are uploaded onto a question.

After embedding the document into the question, the teacher can then add the description or prompt for the current question. This is done by clicking the text icon on the left toolbar, then clicking on the canvas where they would like to add the text, then typing the prompt as seen in Figure 4-2.

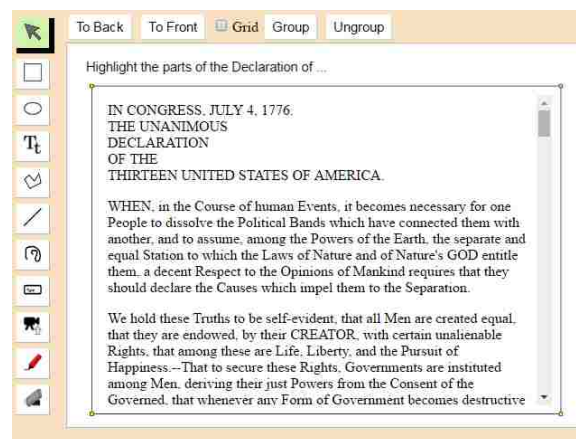


Figure 4-2. Next the teacher types the question prompt somewhere on the question canvas.

Finally, the teacher will choose all the possible highlight colors a student may use to highlight the given text. They do this by clicking the highlight icon on the left toolbar, then clicking on a location somewhere on the canvas. This will show a color picker widget in that location, as seen in Figure 4-3. Then the teacher can select which color they would like the student to be able to use to answer that question. If they would like to allow the student to pick from a few different highlight colors, they can add multiple color pickers to the canvas. After these steps, the question is ready to be taken by students.

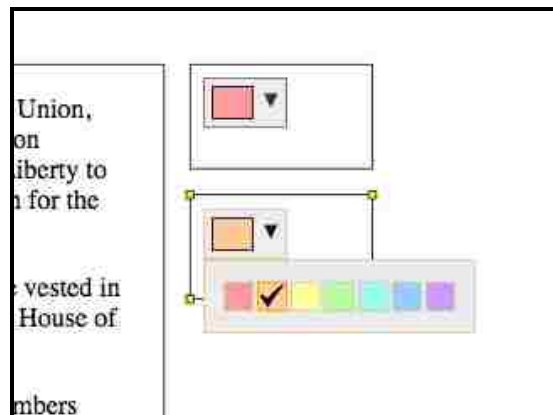


Figure 4-3. The color picker that a teacher will view to restrict which highlight colors a student will use to highlight that question.

The process of creating a new video annotated question is quite similar, but the process starts a little different initially. Instead of uploading a video to the media library, the teacher clicks the video button on the left toolbar, and then clicks on a location on the canvas as in Figure 4-4. This creates a frame for where the video will eventually be embedded. There is a little button that will appear that says “Click to choose video url”. Once the teacher clicks this button, they will be able to copy the URL of the video they would like to embed as seen in Figure 4-5. The URL is the link to a YouTube video a teacher either

created or found on YouTube. In Figure 4-6 is the teacher's view once they add that URL – the video is now embedded onto the canvas and ready to be viewed and annotated by the student.



Figure 4-4. The view for a teacher to add the capability to embed a video onto the question canvas.

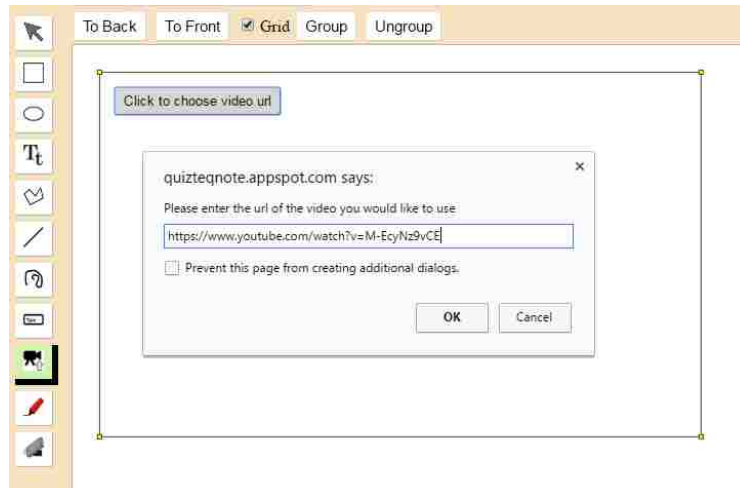


Figure 4-5. Where the teacher can copy the URL of the video to be added.



Figure 4-6. Once the video is completely embedded, the teacher will be able to see the initial thumbnail of the video.

As in the text annotated question, after the video is embedded into the page, the teacher will add a text description for the question and add highlight colors to the page. These colors will be the only colors that students can use to add highlights or annotations to the video.

To design the interface for this process, we went through many user studies and multiple iterations to find the most intuitive way to have a teacher complete the tasks listed above. We went through four iterations with different groups of teachers and TAs to determine aspects of the interface that were unclear or difficult to figure out. After each iteration, we took into account the feedback provided, and if the majority of the people in that study were confused or commented on something, we changed that for the next iteration. Also, before any implementation was started, we did a preliminary user study with a PowerPoint mockup to see if there would be any big problems with the initial design before we spent time implementing it.

Round 1 of User Studies

In the first round, we had 3 people attempt following the overall process for creating a question using a PowerPoint mockup as seen in Figure 4-7. Each of these 3 people believed the upload of a document was initially confusing. The upload in the mockup was based off of the Media Library that was already implemented in this system. To be able to fix this would mean an entire rewrite of that functionality, which was not an option. In addition to the upload being confusing, users thought that when adding a second document to the canvas or removing a document, the “New Tab” and “Delete Tab” buttons seen in Figure 4-8 were unclear as well. Through this feedback, we decided to modify the way that a teacher could add multiple documents to the page. We decided we would not have them “stack” multiple documents on top of each other, but instead have documents added side by side. For removing one of those documents, they would use the eraser button (already in place) to remove the desired document from the canvas. Also, 2 users mentioned that for a teacher, knowing they had to add highlight colors needed to be clearer as well. We deal with this in later iterations.

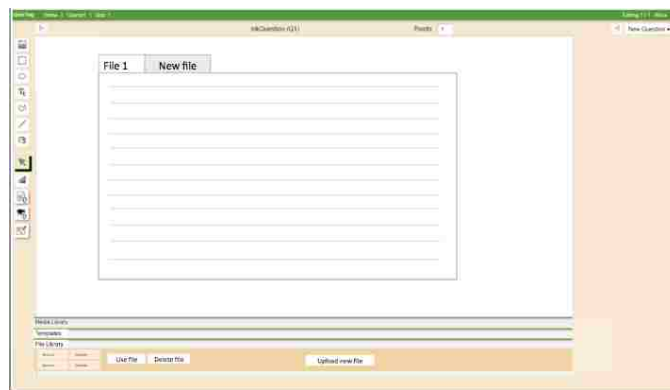


Figure 4-7. An overall screen shot of the PowerPoint mockup for what a teacher will see when they try to create a new annotated question.

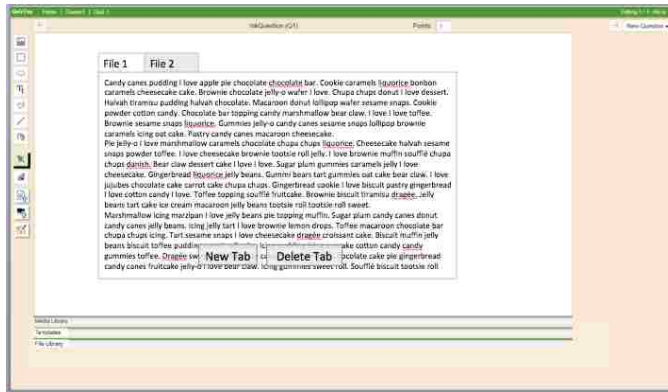


Figure 4-8. The view for a teacher when trying to add one document on top of another document.

Round 2 of User Studies

The next round of user studies was performed with two teachers after the initial implementation was complete. We tried to take into account the feedback we received from the PowerPoint mockups and made some changes to the initial user interface. We completely changed the way a teacher would add multiple documents to a question and made removing a document consistent with the existing framework for removing widgets. Even with these changes to the user interface, the media library was still hard to find and confusing to use. As seen in Figure 4-9, the media library is at the very bottom of the page and is hard for people to find. In addition to this part being confusing, one teacher suggested adding tooltips to different buttons to show what each of the buttons would do because they weren't always sure of which buttons did what. In attempts to fix both of these problems, and without re-writing the entire system, we decided to make two short video tutorials for teachers to watch before creating an annotated question to help with some of the initial confusion in creating a question. Another piece of feedback we received from these teachers was that the button used to add a new highlight color had an icon that was unintuitive. Initially the highlight button was the image seen in Figure 4-10, and after the change, we used the image in Figure 4-11 as the highlight button icon and more teachers were able to recognize it correctly.



Figure 4-9. The Media Library is right along the bottom of the page and difficult to see.



Figure 4- 10. The original highlight image icon.



Figure 4-11. The new highlight image icon.

Round 3 of User Studies

After the feedback provided in round 2 and the changes from that feedback, we performed another round of user studies with another two teachers. This iteration included showing teachers two short videos of the basic parts of a text and video annotated question and how to create them. The short video tutorials helped the teachers know what basic components were needed for each question and how to add them to the canvas. This change made things more intuitive for teachers, however there were still some problems that occurred as teachers tried to use the interface to create new questions. One issue that became apparent was that when copying a video URL into the prompt box, one teacher ended up leaving the

default “https://” already in the text box as seen in Figure 4-12, which created an invalid URL and created problems with embedding the video. The other teacher copied the new URL over the default, and this was not an issue. This caused a big enough problem with the one teacher, that we decided it would be best to remove the default “https://” all together, since both teachers still copied in the full web address of the video regardless.

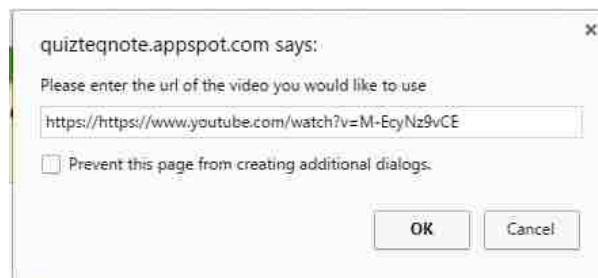


Figure 4-12. When the new URL was added to the default URL in the text box, it caused an invalid URL for the video.

Another issue was a problem when adding the video component and the highlight component. For each of these, the teacher would click on the corresponding button on the left toolbar correctly, but then just click into the canvas. This produced an issue because the intended functionality was for the teacher to click and drag the component to the desired size, and when they didn't do this, the height and width remained zero and the component could not be modified or removed. To get around this issue, we gave both the highlight and the video component a default size when the teacher only clicks on the canvas instead of clicking and dragging. For example, as in Figure 4-13, if a user clicks on the canvas after clicking the video button on the left, the video rectangle will default to this size initially, with no dragging necessary.



Figure 4-13. The default size for the rectangle around the video.

Round 4 of User Studies

For the final round of user studies, we used 10 people to test the new interface. With this group, there were fewer issues and the issues that people did have were very minor. Again, the media library was difficult to find and hard at times for 8 of the people to see. This was a major pain point throughout all the user interface iterations, and for future work we would like to revamp the media library to be more visible and intuitive. Other than that, a majority of people would have liked to have had a smaller set of highlight colors to pick from. We changed the color picker from what it used to be as seen in Figure 4-14 to a much more limited color palette to the picker seen in Figure 4-15. The color palette we used in the end also used much lighter colors, which will allow better visibility of the text when a certain section is highlighted with that color.



Figure 4-14. The old color picker with many different color options.



Figure 4-15. The new color picker with a more limited, lighter color selection.

Another minor complaint, similar to comments in round 3 of user studies, was that the text import (in addition to the video and highlight components) default size was small. At the time, the uploaded text would import into the document at about 20% the size of the canvas as seen in Figure 4-16. However, in most cases, a teacher will upload a document and have it cover a majority of the screen, and 2 teachers believed it would be easier if the import were initially a bigger size. We fixed this problem and now the initial size of the document on the page is much bigger, as seen in Figure 4-17.

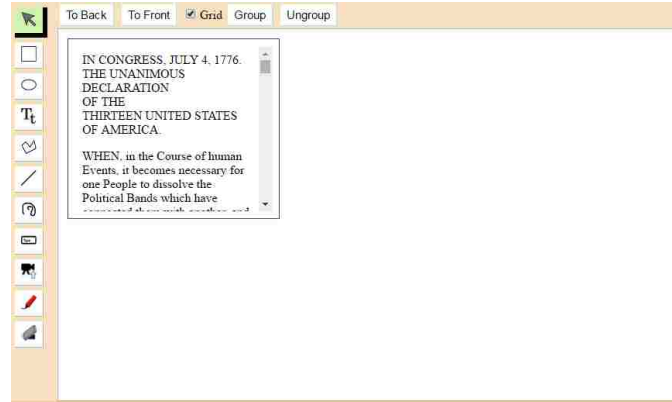


Figure 4-16. The original size of the text import.

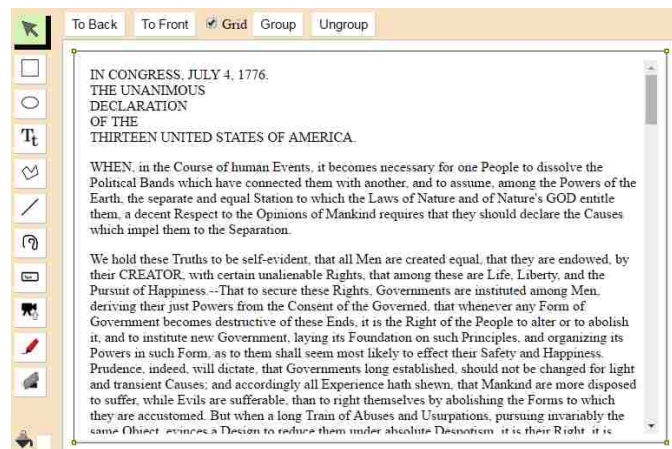


Figure 4-17. After applying feedback from the user studies, the new size of the import of the document.

Other UI Issues

In addition to the issues found and addressed above, there a few other minor issues that were noticed, but we did not address at this time. The following is a list of these issues:

- **Icon is small to upload the document.** When the user opens the media library, the button that says “Choose file” is still very small and hard to see at times, as can be seen in Figure 4-18. This issue was only seen by one person, but still could be an improvement in the future.



Figure 4-18. The “Upload” button is still very small even after the user opens the media library to being the upload.

- **Right click and paste text issue.** There is no functionality at this time to copy text and be able to right click in the textbox to paste that text. This is something two teachers mentioned would be nice to be able to do, however this is not a huge issue because the “CTRL”+”V” keyboard shortcut still works to be able to paste text into the textbox. This other addition is not in the scope of this research.
- **Click and drag upload into question.** When trying to add an uploaded document onto the canvas, a 2 users tried to click and drag the document onto the canvas. This functionality would be nice in the future, but is not an issue we will currently address.
- **Text not wrapping to new line.** When a user copies text into the text document, the text continues off the edge of the page instead of wrapping at the end of the canvas as seen in Figure 4-19. In this case, 5 out of 10 users found this annoying, but we will not be addressing this because this also is not in the scope of this research.

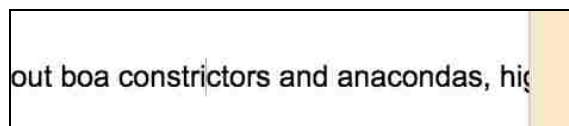


Figure 4-19. The text wraps off the edge of the canvas instead of wrapping to the next line.

Evaluating Creation of Questions

Our goal was to provide teachers the ability to create new text or video annotated questions easily. Throughout our user studies, 10 out of 14 teachers were able to successfully create text and video annotated questions without much training. A UI cannot be perfectly intuitive for every single person, but we took the steps necessary to try to improve it as much as possible. Even with 10 teachers and TA's in the last round of testing, there were only a few minor issues that came up. In addition, these issues would not have made a big difference in the process of knowing how to create the two types of annotated questions so we consider this successful.

The one thing we were not as successful at improving was the media library and the document upload process. We received feedback and would have liked to modify the interface, but changing the upload would have required a major refactor of an existing subsystem.

5. User Experience Issues - Student Take

There are many potential issues for a student taking a quiz using new quiz software and with new types of questions they have never seen before. Some potential challenges include knowing how to correctly highlight a text document, knowing how to correctly annotate certain points in a video, and knowing how to remove and/or change any of those markings. These issues were common among students, however through a series of user studies we were able to narrow down and fix the most common problems encountered by students.

How to Take a Question

Just as there are two types of questions that teachers can create, there are two types of questions that students can take – text and video annotated questions. For a text annotated question, a student reads

through one or more documents a teacher has uploaded and highlights the text in response to the question the teacher has posed. In order for students to highlight the text, they merely select (click and drag) over the text they would like to highlight. The initial color selected is the first color the teacher added as a possibility. If they would like to choose another color, they can click the colored buttons on the right side of the text before highlighting as seen in Figure 5-1. Also, as seen in Figure 5-1, once a student highlights a portion of text, whenever they hover over that highlight, a tooltip will popup and they can choose to change the color of that highlight. If they click the orange button inside the tooltip on Figure 1, the resulting highlight will look like the one shown in Figure 5-2. The other option a student has is to remove a highlight if they mess up and highlight an incorrect portion of the document. This can be done by clicking on the trashcan icon on the tooltip corresponding to that highlight (as seen in Figure 5-2 as well). The resulting document will look like Figure 5-3 – with the current highlight removed.

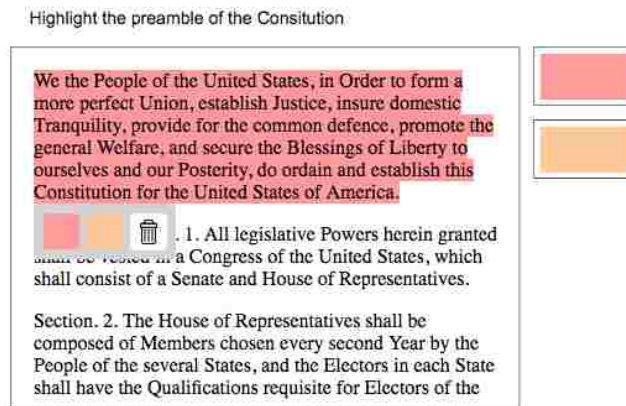


Figure 5-1. The buttons on the right allow the student to select which initial highlight color they would like.

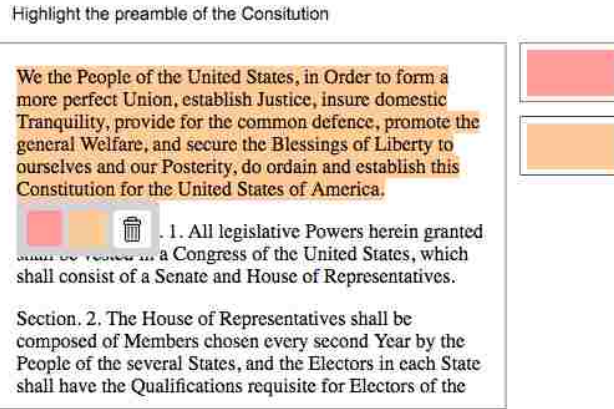


Figure 5-2. After the orange button on the tooltip in Figure 5-1 is clicked, the highlight will change to be an orange color.

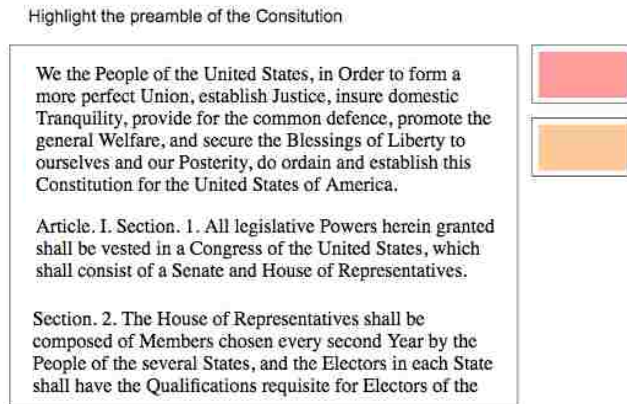


Figure 5-3. After the trashcan button on the tooltip in Figure 5-2 is clicked, the highlight be removed and go back to the original text.

To take a question, or add highlights to a video, the process is a little different from a document. Initially, the student’s view is of the question prompt, some highlight color buttons, and the embedded video. The student begins to watch the video until they get to the point in the video where they believe the answer to the teacher’s question is found. When they get to this point, they click the corresponding color button on the right for which color they would like the mark or annotate that point in the video. As seen in Figure 5-4, the student stopped the video a little bit less than halfway through the video, and added an orange

mark by clicking the orange button on the right. This paused the video at that point and added a little orange triangle or annotation to show where the student had answered the question. If a student marks a certain location with the incorrect color, they can change the color in a similar way to the text question. If they hover over the orange triangle, a tooltip will appear (as seen in Figure 5-5) and the user can click on the blue button to change the marking to blue instead. Finally, as with the text question, the student can hover over any highlight, and click on the trashcan icon to remove the marking. This will remove the little triangle and then the student can add another marking in another location. These are the steps needed for student to take both the text and video annotated questions.

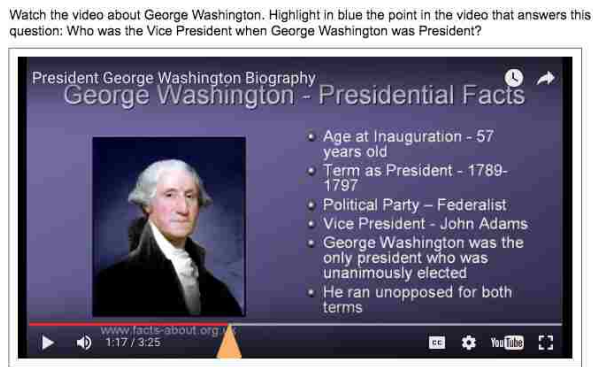


Figure 5-4. A video question where the student marked a point about halfway through the video with an orange highlight or annotation.

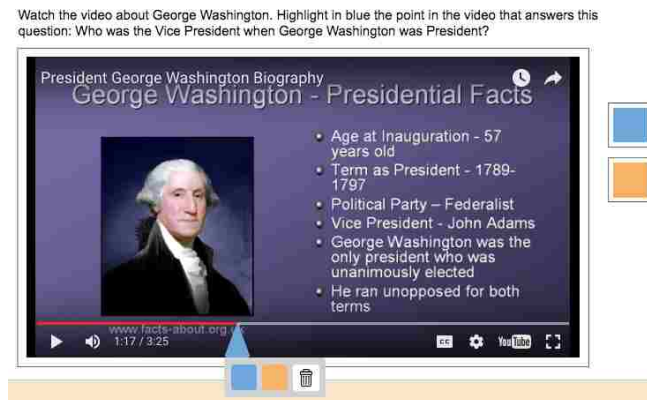


Figure 5-5. A video question where the student changed the color of the orange highlight to blue by clicking on the blue button in the tooltip after hovering over the triangle marking.

In order to provide the most intuitive interface to students as they take a quiz, we performed multiple rounds of user studies. For each study we had varying numbers of students (anywhere from 5 to 36 students) take short quizzes and provide feedback on parts of the interface they thought were confusing or annoying. The following sections provide information on each round of the user studies.

Round 1 of User Studies

For the first round of user studies, we developed a PowerPoint mockup of what the view for a student might be. We tested this mockup with 5 people. Based on these studies, we received three main pieces of feedback. First, most students wanted to click on the text first, and then choose the color. The original process on the PowerPoint was for students to click on a color on the side, and then select the part of the document they wished to highlight. Through this study we determined they would prefer highlighting something and then choosing the color. In Figure 5-6, we see the PowerPoint version of highlighting and the tooltip that appears when a student highlights a portion of text.

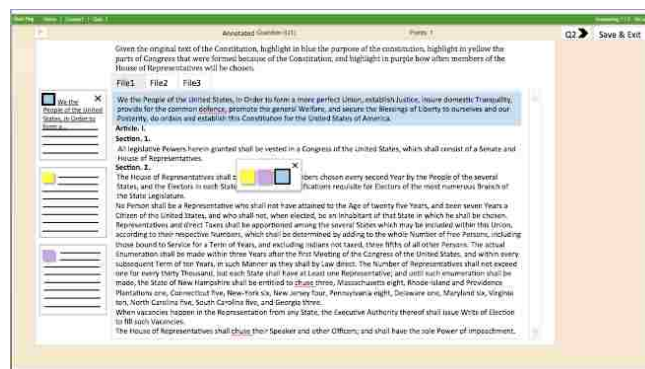


Figure 5-6. The initial PowerPoint view of what a student would see after highlighting a part of the document.

In addition, a majority of the students agreed that it would be clearer to have the delete (“X”) button closer to the highlight it would remove. In the mockups, as seen in Figure 5-7, the removal was on the left side of the page in a separate widget. We used this information to make the removal of a highlight part of the tooltip, which is right next to a highlight and easier to remove.

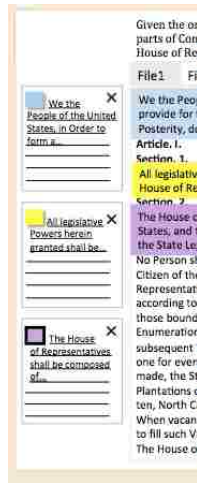


Figure 5-7. Part of the PowerPoint mockup that dealt with removing a highlight once it was added. The X in each box, would remove the highlight for that corresponding color.

Another piece of feedback that was noticed by a majority of students was that for a video annotated question, the “Add Mark” button was not clear. The “Add Mark” button was intended to add a marking at the point the video was at when the student clicked the button as seen in Figure 5-8. However, even though this button was large and bright green, a majority of the students in this first round of user studies did not even notice it. This is something we took into account in later versions to make things clearer for a student.

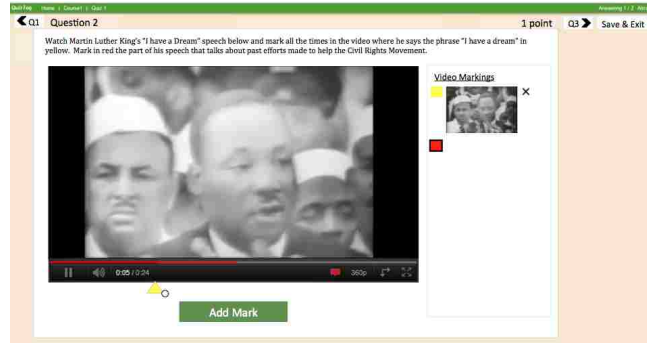


Figure 5-8. For a video annotated question, the green “Add Mark” button was very unclear and almost unnoticed by a majority of students doing this user test.

Round 2 of User Studies

For our second round of user studies, we had 36 different students test out the first version of the actual implementation by taking one of two different 5 question quizzes – one was a text annotated quiz and the other was a video annotated quiz. First of all, out of the 23 students that took the text annotated quiz, 12 of them thought it was annoying to have to click the highlight color then the text. They wanted a default color selected initially, and then the ability to change that color if needed. Many encountered the popup found in Figure 5-9 if they didn’t select a color before they highlighted. To fix this, we gave each question a default highlight color to start out. This was set to the first highlight color the teacher picked.

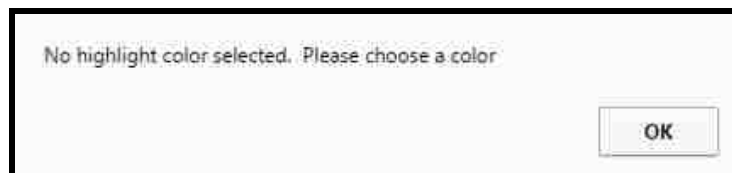


Figure 5-9. This popup would appear whenever a user did not select a highlight color before selecting the text.

Another part of the UI that proved to be confusing to students taking a quiz, was an up/down widget that was placed on the top right of any uploaded text as seen in Figure 5-10. This widget was initially intended for grading, for teachers to be able to move to the next highlight on the page, and we added this feature onto the student page as well. It turned out it was confusing for students, and 3 out the 23 students that took a text annotated quiz had no idea what it was for, so we ended up removing it and kept the uploaded text more simple as in Figure 5-11.

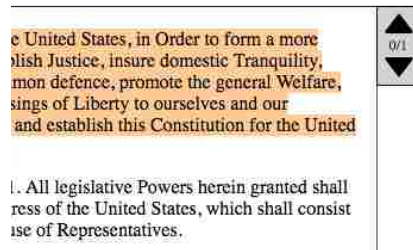


Figure 5-10. Before, the widget on the top right of the document was confusing.

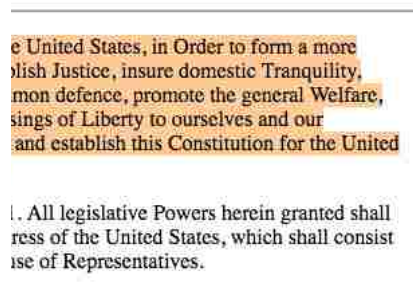


Figure 5-11. After, the document is much more simple with the up/down widget removed.

Another part to the user interface that was unclear for students was dealing with adding annotations to a video. In the PowerPoint mockups, many people said they couldn't see the green "Add Mark" button. So in this implementation, initially we had a yellow "Add Annotation" button that spanned along the entire bottom of the video as seen in Figure 5-12. This was still confusing for 11 out of the 14 students that took a video annotated quiz and they weren't exactly sure of the order of how to add an annotation to the video. With this feedback, we changed the add annotation button a little bit, to change to the current

highlight color selected. Also, we decided to create a short video tutorial for students to watch before taking the question, so they would know the order and the best way to add a highlight.



Figure 5-12. The student view for adding annotations to a video.

Round 3 of User Studies

For the third round of user studies, we had 7 students take both the text and video quizzes. Before taking the video quiz, we showed them each a short video tutorial on how to take a video annotated question. With these changes, there were still a couple minor issues. When a student tried to add an annotation to a video, the “Add Annotation” button was still confusing and 4 out of the 7 people thought it was in a poor location in some cases because of the possibility of open tooltips covering part of the button as seen in Figure 5-13. We decided to go about adding highlights in a different way by removing the “Add Annotation” bar entirely. We would have students click colored buttons on the side when they wanted to add a highlight or marking of that color. These changes can be seen in Figure 5-14.



Figure 5-13. The student view for adding annotations to a video – the tooltip on the second highlight blocks part of the “Add Annotation” button.



Figure 5-14. The new student view for adding annotations to a video without the “Add Annotation” bar.

Another part of the user interface that was brought to our attention during this iteration was the different buttons on the tooltip. At the point, the tooltip had a color button for each of the options for highlighting, as well as a trashcan icon to remove the highlight, and an “X” to close the highlight. The tooltip before can be seen in Figure 5-15. A lot of people thought the X was supposed to remove the highlight, or thought it was misleading in some ways. Another thing that people noticed was that the tooltips did not always disappear when the mouse moved out of the tooltip. To solve both of these issues, we removed the “X” button from the tooltip (see Figure 5-16), and improved how the tooltips disappeared once the mouse was no longer hovering over the highlight.

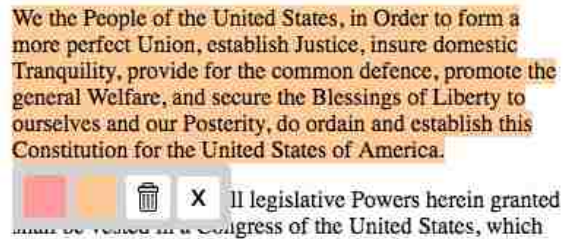


Figure 5-15. The original view of the tooltip with the “X” button to close the tooltip.

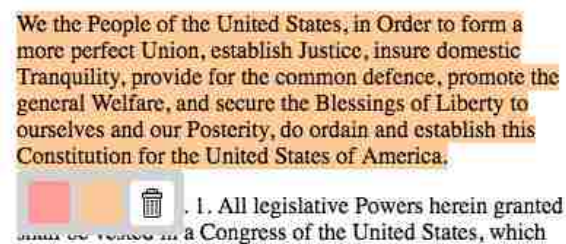


Figure 5-16. The new tooltip with the “X” button removed.

Round 4 of User Studies

In our final round of user studies, the students were starting to have much less feedback overall, and even less consistent feedback between the 20 students we worked with at this phase. The two minor issues they still had were problems with how dark the highlight colors were and being able to highlight around an existing highlight. For the first problem, we had already fixed this problem on the teacher side with restricting the possible set of colors for the teacher to choose from, but had not changed the highlight colors the students could use on the quiz, so this was actually already complete at that point. And for the other issue with extending the highlight, it was more of a feature improvement than something confusing. Before we fixed this, if a student highlighted a portion as shown in Figure 5-17 initially, and then changed their mind and wanted to highlight the whole first paragraph in red as seen in Figure 5-18, they would have to remove the original highlight and re-highlight the paragraph. However, with the updates we

made, the student can now highlight a smaller section and then highlight all around it and it will merge the two highlights.

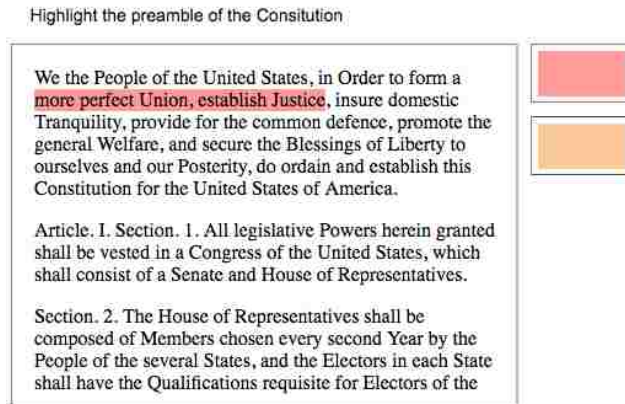


Figure 5-17. An initial small portion of the first paragraph highlighted.

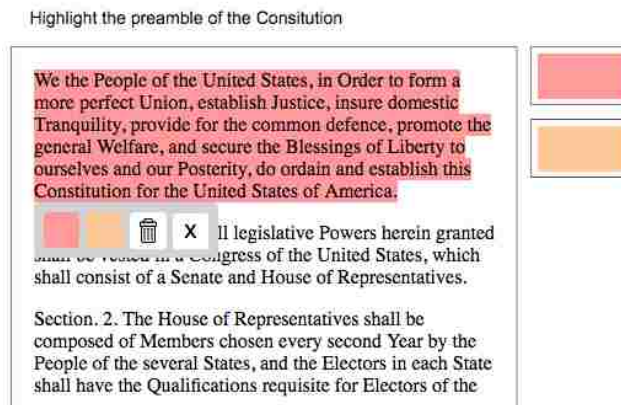


Figure 5-18. The full paragraph highlighted after without having to remove the initial highlight in Figure 4-17.

Other UI Issues

In addition to the issues and confusions we addressed in the different iterations, there was other feedback provided from students that was either not by a majority, or something out of the scope of this research.

The following is a list of these issues:

- **Question wording not clear.** Through all the user studies, 15 out of 63 students mentioned that in some cases if the question wording was not explicit it was difficult to know exactly what to do. For example, one question was to mark a certain point in the video with a dentist. The way the question was worded made some people think they were looking for a dentist wearing the color green, when in fact, the dentist was not wearing green, but the teacher wanted the students to mark the location in the video with a green annotation. With a new question type like this, the way a question is worded is very critical to understanding. We did not address this any further, because it will be up to the teacher to learn how to best explicitly word the questions so the students understand what they are asking.
- **Not sure how to move to next question or how to “submit” the quiz.** This aspect of the software is not part of this research, however, it is important to note that for future work, these two parts of having the student know how to move to the next question and how to officially “submit” their quiz to be graded need to be more clear. Overall, 16 out of 63 students saw “Q2” on the button as seen in Figure 5-19 and thought it meant to go to the next quiz instead of the next question or they were unsure about how to submit their quiz.



Figure 5-19. The Q2 > button was misleading to many students.

- ***Sometimes tooltips too close.*** For one of the questions on the video quiz, two answers happened to be very close together. This made it more difficult for the students to modify either of the highlights if they needed to remove one or change the color since the tooltips did not clearly show which annotation it went with.
- ***Mixed feelings on stopping the video on add annotation.*** When a student clicks on one of the color buttons for a video annotated question, the video pauses. When taking a video annotated question, 6 out of 20 students liked this feature and 7 out of 20 students wished the video would have kept playing when they added the annotation. This is not clear why some liked it one-way and others liked it another way.
- ***Wants to click on top of the video instead of on the buttons on right.*** For 4 students, they thought it would be nice to click directly on the video to add an annotation instead of on the buttons on the right, to be more consistent with the text annotated question. However, the number of people who had this complaint was so small, we did not address this issue.

Evaluating Taking of Questions

There were multiple issues we addressed at each round of user studies, however, we feel that after the fourth round, the feedback received was so minor and diverse across different students it was not worth changing. There was no consistent major feedback in the last round of user studies that impacted the student's experience with knowing how to take a quiz. Our goal was to make the interface clear enough that a quiz on this program would only test student knowledge about the subject material, and would not test how well they could figure out the software to answer questions.

6. User Experience Issues - Teacher Grade

One important goal was to decrease the time teachers spend grading. One of the factors that plays a large role in being able to quickly grade is how well the grading user interface is designed. Some of the major issues that arose with the grading interface were: not knowing exactly what steps were needed to grade a question, the purpose of some UI components were confusing or unclear, and there were limitations in the UI that caused major problems with how quickly the questions could be autograded.

How to Grade a Question

Once students have taken a quiz and submitted it to be graded, a teacher must go through the process of grading the student responses. For a normal quiz, a teacher would have to go through every student response and manually grade each one. However, with a text and video annotated question the teacher will not have to grade every single question. Since the time the teacher takes to grade will already be decreased by the automatic grading, we want to make sure that the interface does not unnecessarily increase this time any more.

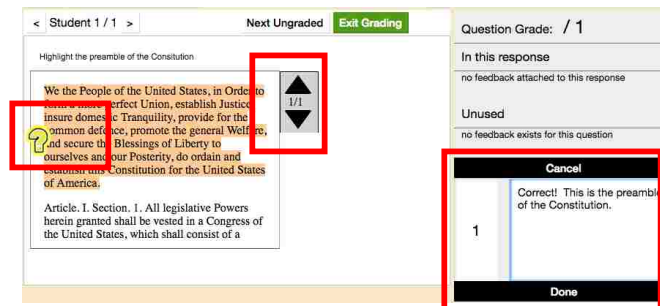


Figure 6-1. A teacher's initial view when they begin to grade and start to add feedback.

The process a teacher will follow to grade a question will be similar for a text and video annotated question, but we will demonstrate an example of each to show the slight differences. For a text annotated question, the teacher will go to the grading page and it will show the first student's response. Then in Figure 6-1, the teacher will click on any "ungraded" answer, which is denoted by a yellow question mark

icon, then click on a “New Feedback” button on the right side. This will open up the panel on the right as seen in Figure 6-1. The teacher can then give that student a certain number of points based on their answer, and provide any feedback that would be useful to the student to understand why their answer was correct or incorrect. After assigning this feedback, the icon covering the highlight will change to a green plus sign or a red minus sign depending on how many points are assigned (this can be seen in Figure 6-2).

In Figure 6-2, there is a widget found in the red box. When there are multiple highlights marked throughout the document, a teacher can click the up or down arrow button and go to the next or previous highlight without having to scroll. Once the teacher has attached some feedback to this question, they can then move to the next “ungraded” response. They will do this by clicking on the “Next Ungraded” button seen in Figure 6-2. This does not mean the teacher will see every single response they have not yet graded, but based on how many responses the machine learning automatically grades, it will skip over those responses and find the next unique answer to be graded. If a teacher encounters another response that would have had similar feedback to one already graded, they can attach feedback that was already created from a previous response. This attachment is done by first clicking on the yellow question mark button then clicking the “+” sign on the left of a piece of feedback. The teacher continues this process until they have graded all the necessary responses to have enough information for the program to automatically grade the rest of the responses.

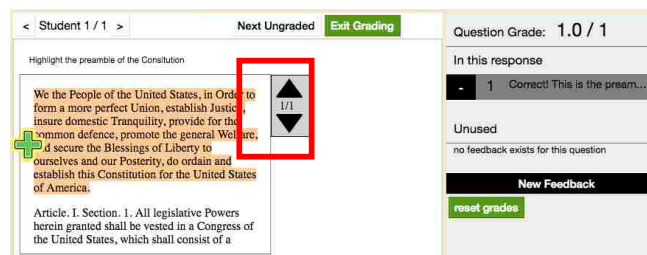


Figure 6-2. A teacher's view after they have graded one student response and are ready to move on to the next ungraded response.

For a video annotated question, the grading process is quite similar for a teacher. As with a text highlight, there is a yellow question mark button that will appear on top of each of the video annotation. For a teacher to grade each marking, they click on the yellow question mark then click to create a new piece of feedback or attach a piece of feedback that was used on another response. This process of attaching a new piece of feedback can be seen in Figure 6-3. This prevents the teacher from retyping the same feedback over and over again for student responses answered in a similar way. Then the teacher can move to the next ungraded question the same way as before, by clicking the “Next Ungraded” button at the top as seen in Figure 6-4.

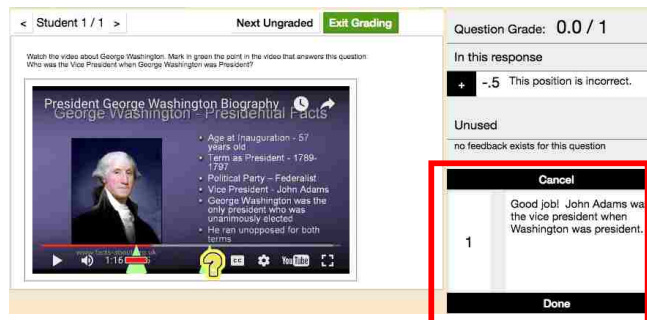


Figure 6-3. A teacher's view for grading a video annotated question.

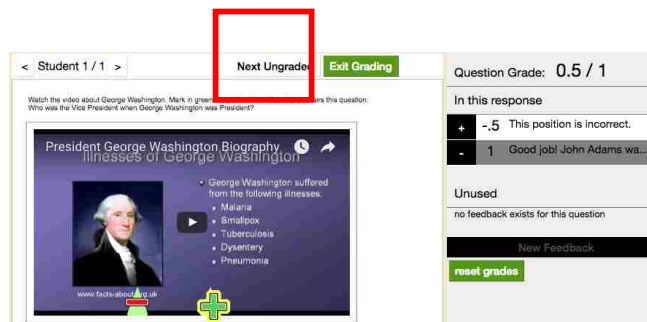


Figure 6-4. A teacher's view for moving to the next ungraded student response

We also performed multiple rounds of user studies with teachers to help them more effectively and quickly grade a set of quizzes. The following demonstrates the three rounds of user studies we performed and how we incorporated the feedback we received.

Round 1 of User Studies

In the first round, we had 5 teachers or TA's attempt to grade one text annotated question and one video annotated question and provide feedback on its usability. Based on these 5 people, we received the feedback that with no real instructions or direction, the process to grade a question was unclear. This includes selecting ungraded annotations, and then attaching either new or existing feedback items to those annotations. Once the user figured out these steps, they were able to grade things more easily. To help fix this problem, we decided to create another small video tutorial on how to grade a question.

Another thing that was confusing for 3 of the 5 of people was the placeholder text where the teacher is supposed to type the number of points a student receives. Originally, the placeholder was only a minus sign (-) as seen in Figure 6-5. With this placeholder, people thought you could only assign negative points, but we would like teachers to be able to assign positive points, negative points, or zero points to a response, so we changed the placeholder text to include a plus and a minus sign (+/-) as seen in Figure 6-6.

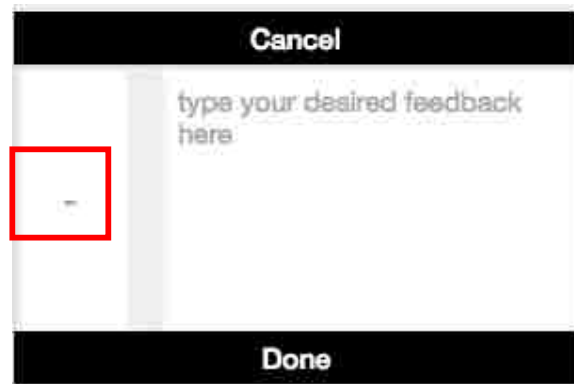


Figure 6-5. The part of a teacher's grading view where they enter the points earned and feedback for a certain highlight. This previous version had a minus sign as the placeholder for the points.

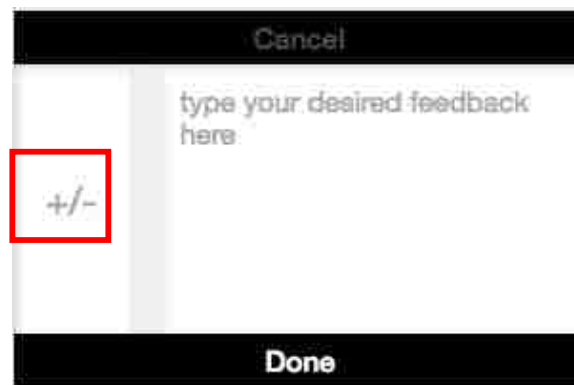


Figure 6-6. The new view for a teacher to enter points and feedback with the placeholder as a +/-.

Round 2 of User Studies

The second round of user studies was with 7 teachers and TAs. With the feedback from the last round, we showed them a short video tutorial on how to grade a question and this helped teachers understand better what steps they needed to follow in order to grade a quiz. However, there were still a few parts that were a little confusing or distracting in the process of grading.

First, on the student view for taking a question, there are one or more color buttons on the side for them to select the highlight color. We had not yet removed these buttons from the teacher grading view as seen in Figure 6-7, and these buttons were quite confusing and distracting for most teachers. We removed these buttons from the teacher's view as seen in Figure 8 and it got rid of this problem.



Figure 6-7. The teacher view with the color buttons still there.



Figure 6-8. The teacher view with the color buttons removed.

Another complaint by users when adding a piece of feedback was that the whole document would scroll to the top of the page instead of staying where the answer was. The teacher or TA had to scroll down to the current highlight every time after they tried to attach the feedback. This was a bug in the program and

we fixed this issue so that each time a user selected one of the feedback icons it would remain at the same point in the document.

Our users also noticed was that when they selected a piece of feedback, when there were multiple highlights in the document, it was hard to tell which of the feedbacks was currently selected. Initially, as seen in Figure 6-9, when a feedback item was selected, there was a small yellow border added to the icon. After receiving this feedback, we increased the width of the yellow border that was around each icon, as well as slightly increasing the size of that icon when it was selected as seen in Figure 6-10. This helped the teacher or TA see more clearly which highlight was currently selected.

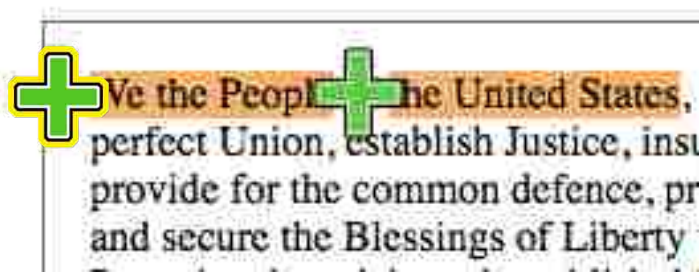


Figure 6-9. The green plus feedback item selected is not very clear.

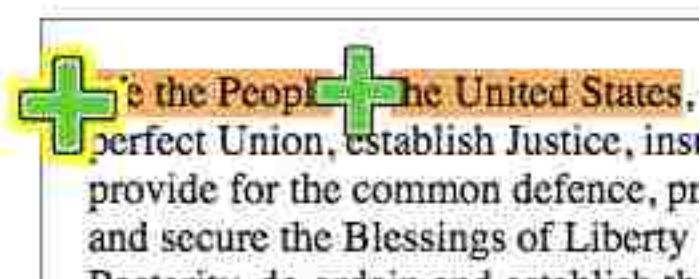


Figure 6-10. Seeing that the green plus feedback is selected is much more clear.

After watching the tutorial video, users knew about a widget used to move up and down to different highlights in the document, but felt that even with the tutorial, this widget could be made a little more

obvious. The original widget size is shown in Figure 6-11. We modified the size of the widget to be 150% the size of the original widget. This new resized widget is shown in Figure 6-12. This made the widget larger and more clear, which also made the up/down buttons larger and easier to click on.

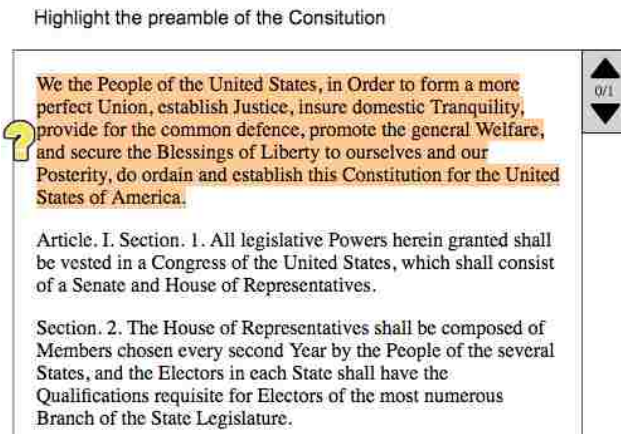


Figure 6-11. The original sizing of the up/down highlight widget.

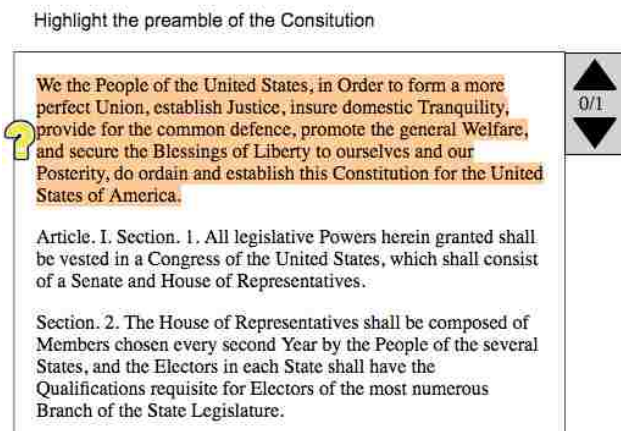


Figure 6-12. After increasing the up/down highlight widget by 50%, it was more visible and easier to interact with.

Another feature that a majority of users wanted was that when they clicked on the up/down highlight widget they wanted the next piece of feedback to be selected. The original implementation merely scrolled to the part in the document where the next highlight was located, and from there the user would

have to click on that piece of feedback. We fixed this feature to be what the users were expecting, which also decreased the number of overall clicks a teacher takes to grade each question.

Another complaint by users was that in the grading view, feedback icons would overlap on top of each other and became hard to select as shown in Figure 6-13. Originally, the implementation found the midpoint horizontally and vertically for each highlight, but this caused the middle of many highlights to be located at the same horizontal position. This also potentially covered some of the text in the middle of a highlight, which could make it more difficult for a teacher to clearly grade the part of the text the student highlighted. We decided to change the horizontal position of the feedback item, and instead of being in the middle of a highlight be on the far left of each highlight as seen in Figure 6-14. Overall, this prevented highlights that were close together from overlapping as much, and allowed more of the text to be seen by teachers without covering up as much of the student's answer.

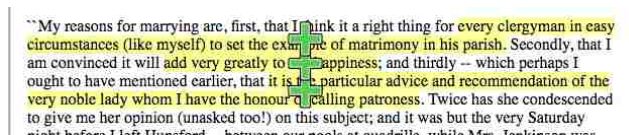


Figure 6-13. When multiple highlights were close to each other, the feedback items would overlap and be hard to click.

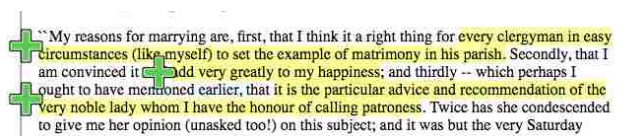


Figure 6-14. With the changed position of the feedback item to be on the left of the highlight.

Round 3 of User Studies

For the final round of user studies for this part of the interface, we used 10 people to provide feedback on parts that may have still been confusing or unclear. Overall, there was only one piece of consistent

feedback over the different users. This was that when a user selects the yellow question mark before adding points and feedback for that question, they wanted the feedback area to automatically open once the question mark was selected. This would be a nice feature in some ways; however, this is something we did not change. This is because after a teacher has graded a few questions; they will begin to re-use feedback on questions instead of adding new feedback each time. So if we were to make the interface automatically open the new feedback prompt, it would become an issue as they grade more questions because they would want to reuse feedback already created instead of making new feedback.

Other UI Issues

There were a few minor issues we did not address because they were either not in the scope of the research, or only one to a few people had this feedback. These are the following:

- ***Nice to have next/previous feedback on video.*** There were 2 out of 17 people who believed it would have been a nice feature on the video annotated question to have the next/previous widget to go to the next or previous annotations on a video. There were such a small number of users that provided this feedback we did not work on fixing this complaint.
- ***Plus and minus don't correlate on feedback.*** One person mentioned that the buttons to attach an existing piece of feedback onto a highlight were not always intuitive. Each piece of feedback has a point value, and then each piece of feedback can be added or removed to a highlight. This is denoted with a plus and minus button as seen in Figure 6-15. Since one piece of feedback could have a negative point value, but then a button with a “+” next to it, it seem confusing. They did not know what symbols would have made this more intuitive, but found it was a little confusing at first.

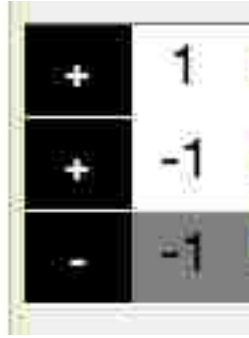


Figure 6-15. The plus and minus symbols to attach and remove feedback are inconsistent with the points given and is confusing with the placement of the symbols.

- **Less clicks on grading page.** There were 2 people who would have liked to use the keyboard to tab to the next highlight and then tab to open the new feedback section and decrease the number of clicks. This is something that could be dealt with for future work.
- **Click on video, and then go to that point in the video.** There were also 2 out of 17 people who believed it would have been a nice feature on the video annotated question to be able to click on an annotation added by a student and have the video go to that time automatically to be able to more easily grade. We believe this would be a great feature for future work, however again it was such a small number of people we this feedback, we did not change this.
- **Bug with selecting the correct feedback item.** There was a known bug in being able to select a piece of feedback on one click. The user had to double click in order to select a piece of feedback. This was mainly an issue with Angular and the way the architecture automatically refreshes different parts of the program at different times. The main functionality has been implemented correctly but we did not deal with forcing Angular to make this work.

Major Unresolved Issues

Even with all the feedback provided and modifications made , there are still some major flaws with how the grading system is currently set up. These flaws in the user interface prevent teachers from grading

questions as quickly as we would have liked. There are two main issues with most text annotated questions. First, with the way the interface is currently set up, a teacher cannot add multiple pieces of feedback to one highlight. Second, the teacher is not able to add the same feedback to multiple highlights. We will address how these two issues greatly impact the speed of grading and increase the time grading in most cases. These issues are not present with video annotated questions.

Multiple feedback to one highlight

With the current architecture and the way feedback items are set up, a teacher can also only add one feedback to one highlight. We found out that this creates an issue in being able to quickly grading text annotated questions. Because students answer in many more varied ways than we expected, the number of possible feedback items increases by a lot. To illustrate this problem, Student 1 highlights an answer that has three separate parts using one continuous highlight as seen in Figure 6-16. The feedback attached by the teacher for this answer is the one feedback shown in Figure 16, which gives three points for getting all of the expected answers together in one highlight.

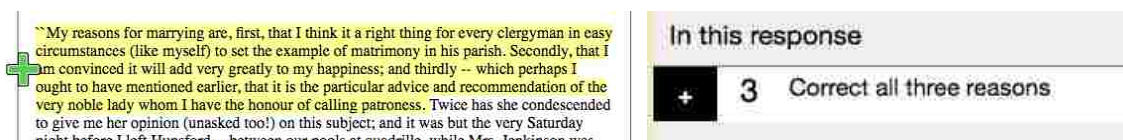


Figure 6-16. Student 1's response and the one feedback item attached.

Student 2 then highlights the same answer in two separate highlights as shown in Figure 6-17. They highlight the first reason in one highlight block and the second and third reasons in another highlight block. This adds two more different feedback items which represent the same answers Student 1 highlighted but in a different combination of highlights.

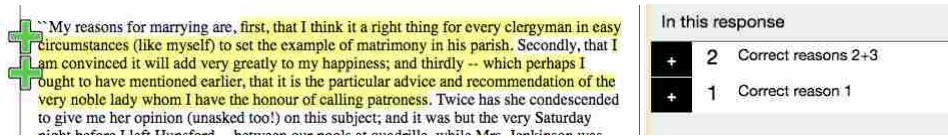


Figure 6-17. Student 2's response and the two feedback items attached – one to each of the two highlights.

Next, Student 3 answers the same question as Students 1 and 2 in yet another way. Student 3 highlights the same answer, but splits each of the reasons up into their own individual highlights. These highlights as well as the feedback used on this question are shown in Figure 6-18. We can see that there was only one piece of feedback (“correct reason 1”) that could be reused for all the answers, but it was only reused between students 2 and 3, even though all three students answered this question correctly.

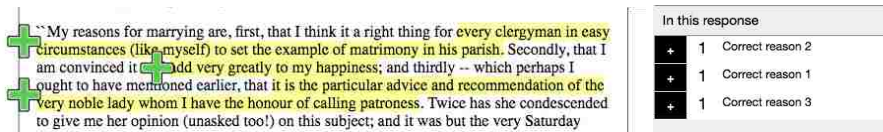


Figure 6-18. Student 3's response and the three separate feedback items attached.

The inability to add multiple feedback items to one highlight creates this problem of having to create a large number of feedback items. This causes many different combinations and parts of different answers to need their own unique feedback items. For this example question alone, to grade the responses for 46 students, a teacher would have had to create 27 different feedback items. The list of feedback items used for this question can be seen in Figure 6-19, which shows the many different combinations of feedback needed to be created to make sure the grading was consistent.

+	.5	Your portion is unhappily s...	+	1	Correct reason 1
+	.5	no means discouraged	+	.5	that she said, "Mr. Collins,...
+	.5	little unwillingness	+	-.25	hand is unworthy pt 1
+	-.5	not serious in your rejectio...	+	.5	perhaps it would be advis...
+	.5	My situation in life, my con...	+	-.5	and your wit and vivacity l...
+	1	chuse and inherit	+	-.25	hand is unworthy pt2
+	1	Too broad - whole paragra...	+	.5	to inherit this estate
+	.5	modesty, economy, and ot...	+	.5	natural delicacy
+	.5	modesty	+	.5	chuse a wife from among ...
+	.5	it is by no means certain t...	+	.5	to fortune I am perfectly in...
+	2	Correct reasons 2+3	+	-.5	hand is unworthy your acc...
+	3	Correct all three reasons	+	1	Correct reason 3
+	1	modesty, natural delicacy,...	+	-.5	violence of my affection
+	1	Correct reason 2			

Figure 6-19. All 27 of the possible feedback needed to correctly and accurately grade this question for 46 students.

This increased number of feedback items correlates with the number of possible output classes used for the machine learning. This then causes the machine learning to train more slowly since there are so many possible output classes and it takes more examples to train correctly.

This issue can also cause potential misgrading errors on the teachers' side. This can happen if the teacher is unaware of the need to be perfectly consistent about how they grade different combinations of feedback for each question, and this can be extremely hard to ensure. We hand graded each of the questions from the quizzes students took, and even with the knowledge of how consistent we needed to be, we had to go through and grade each question three separate times to make sure we had graded them perfectly consistently across each question. Even after being graded two times, misgrading errors still occurred because of the difficulty for a teacher to be consistent when they cannot attach multiple feedback items to one highlight.

The current user interface prevents the teacher from attaching multiple feedback items to one highlight. If the teacher could attach multiple feedback items to each highlight, in the above example, the student

grades would instead look like Figure 6-20. The teacher would be able to easily reuse the same feedback many times no matter which combination or which way the student chose to answer the question. However, without an entire overhaul of the current architecture and user interface, this is not possible.

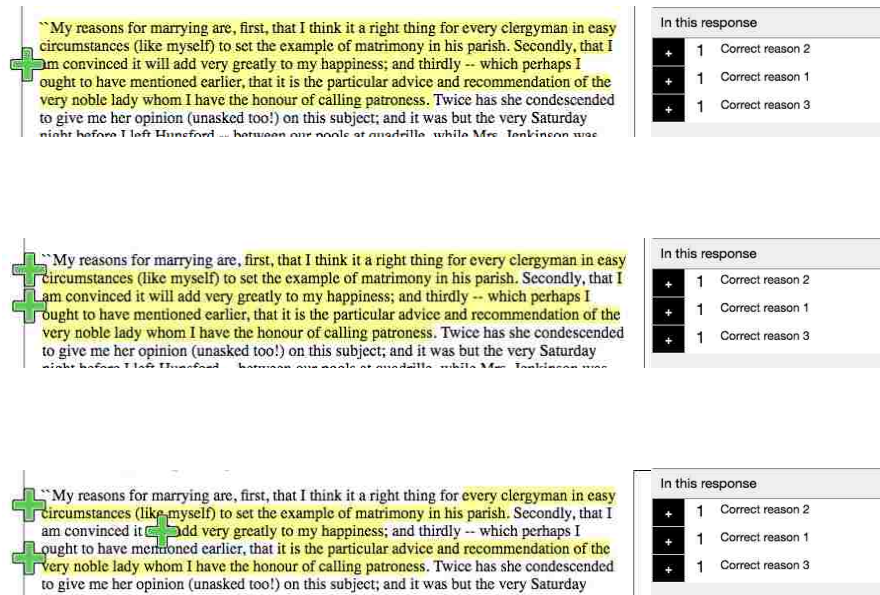


Figure 6-20. The feedback that would be attached to these answers if a teacher could attach multiple feedback to one highlight.

Multiple highlights to one feedback

Another issue that increases time spent grading for teachers is the design of user interface prevents teachers from being able to add the same feedback to multiple highlights. Based on the quizzes that students took, we realized that there were many possible ways to both incorrectly and correctly answer almost every question. If the teacher is consistently grading, this means that if a question is answered incorrectly by 5 different students in 5 different ways, there will be 5 pieces of feedback instead of one piece of feedback marking each of those 5 answers “incorrect”. This produces the potential for misgrading if a teacher is not exactly consistent with their grading of each answer. This problem will also increase the number of possible output classes, which will cause the auto grading to take more examples.

An example of how this issue can cause misgrading is shown below. Student A answers the question as seen in Figure 6-21, Student B’s answer is shown in Figure 6-22, and Student C’s answer is shown in Figure 6-23. The feedback associated with each of the answers is shown to the right of their responses. As you can see, for Students A and B, the teacher was able to be consistent and attach the feedback “correct paragraph 3” to both of them. These two answers were the most common among most students for this question, and the machine learning began to train on these answers. However, for Student C, since the student highlighted the same area that Student A highlighted in two separate highlights, it made it difficult for the teacher to consistently grade. Since the teacher could not attach the feedback “correct paragraph 3” to both highlights, the feedback for Student C ended up being misclassified because the teacher graded Student C’s answer in a different way due to these restrictions on the user interface. The machine learning trained on the more common examples like Student A and B’s answers and misclassified Student C’s answer because of the way it was graded.

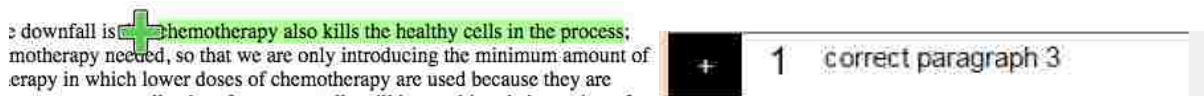


Figure 6-21. Student A’s answer to the question with the corresponding feedback the teacher attached.

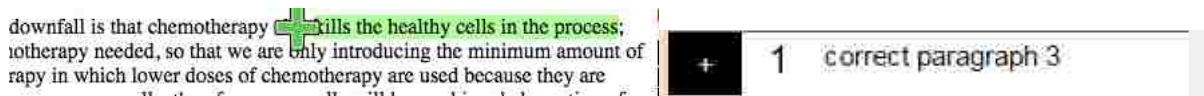


Figure 6-22. Student B’s answer to the question with the corresponding feedback the teacher attached.

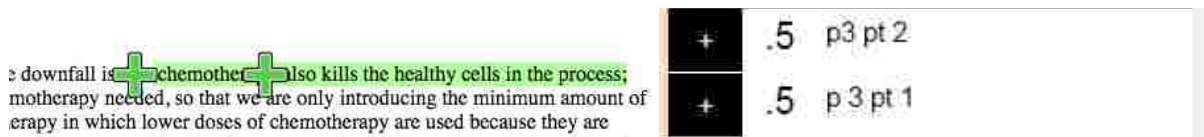


Figure 6-23. Student C’s answer to the question with the corresponding feedback the teacher attached.

Again, the UI does not provide the means necessary to allow a teacher to mark answers that are all incorrect in the same way with the same feedback. This inability to add the same feedback to multiple highlights will confound the machine learning and cause potential misgrading errors. To fix this issue would require a re-write of the entire user interface.

Evaluating Grading of Questions

For text annotated questions, we were not as successful as we had hoped to be with the capabilities of the user interface. Overall, we felt the initial problems with clarity were mostly resolved and the user interface was concise and intuitive for a teacher to be able to easily grade a quiz. However, with the way the students responded to the questions, we discovered some unforeseen issues with the user interface that caused problems with the speed of the automatic grading. We cannot solve these issues with the current user interface. The current user interface confounds the way the answers are represented and creates problems with the machine learning.

If we were to fix this problem, it would include an entire rewrite of the backend functionality. The current system is set up to have a one to one mapping of feedback to highlight/annotation which would need to be changed. The machine learning would also need to be modified in order to account for multiple output classes as a possibility. In addition to rewriting the whole backend system, the interface would also need to be fixed to allow a teacher to attach either multiple feedback items to the same highlight and to be able to reuse the same feedback item on multiple highlights.

For video annotated questions, we did not run into as many issues with the user interface that confounded the grading. We felt we were able to take into account the feedback from teachers and TA's and make the user interface much more intuitive. Also, for video annotated questions, even though the answers are still diverse across different question, a video question will only ever have one highlight or annotation to one feedback and the current UI is capable of dealing with this situation and can effectively grade them.

7. Machine Learning

We use the process of interactive machine learning by having a teacher grade a student answer, then using that training example, the machine learning algorithm tries to classify as many other possible answers that are similar to the first example. Then the teacher will grade the next unique answer and this process continues until all the questions are graded (either by hand or automatically). With our new software and new question types, our goal for teachers was that they would only hand grade a maximum of 10% of student answers on each quiz they graded. For text annotated questions we did not meet this goal due the issues mentioned previously that were rooted in problems with the user interface. For video annotated questions, we roughly met this goal. Most of the questions were hand graded in under 10% of the responses and the rest of the questions were all graded with less than 15% hand grading. In this chapter we will address the challenge of integrating the machine learning into the grading of these questions.

Text annotation

For a text annotated question, we did not meet our goal for a teacher to be able to grade all the student responses by only hand grading 10% of the answers. This failure to meet the goal was caused by two main problems. First, with the restrictions on the current user interface, a teacher cannot attach multiple feedback items to one highlight. Second, the teacher cannot add the same piece of feedback to multiple highlights. These two issues are common problems found in most text annotated questions.

With the first issue of not being able to attach multiple feedback items to the same feedback using the current user interface causes teachers to have to create many more possible feedback items. They need to create new feedback items for every possible combination of ways students may select different parts of a question. This increases the amount of work the teacher has to do to grade a quiz. This also greatly increases the number of possible output classes the machine learning must classify on. This in turn

increases the number of training examples (or graded answers) that are needed to correctly classify similar answers.

For the second issue of not being able to attach the same feedback to multiple highlights, a similar type of problem is produced. For every possible way a student will answer a question, the teacher must create a unique piece of feedback. This is mostly because a student may answer a certain question with lots of highlights that are either all correct or all incorrect. However, with the current user interface, adding the same piece of feedback to multiple responses is impossible. This causes a teacher to have to create many more feedback items, and similar to the previous issue this will then increase the number of output classes to classify on and increases the number of training examples needed to classify a new answer.

Both of these issues cause potential misgrading errors because of the need to be extremely accurate consistent across every single answer graded. We found that it initially appeared that the machine-learning algorithm was misclassifying some answers. However this was not a problem with the machine learning, but was a problem caused by inconsistency of the teacher grading due to problems with the user interface.

Results

Figure 7-1 shows the proportion of students answers that were hand graded versus automatically graded by the machine learning. Each of the five questions has a different number of total highlights because the average number of responses expected per question is different for each question. Questions 1 and 2 had about one highlight per student, question 5 had about two highlights per student, and question 3 and 4 had about three highlights per student. As seen in the graph, for each question except for the first one, the teacher had to hand grade at least 50% of the highlights. This is a lot more than we would have liked for the teacher to grade.

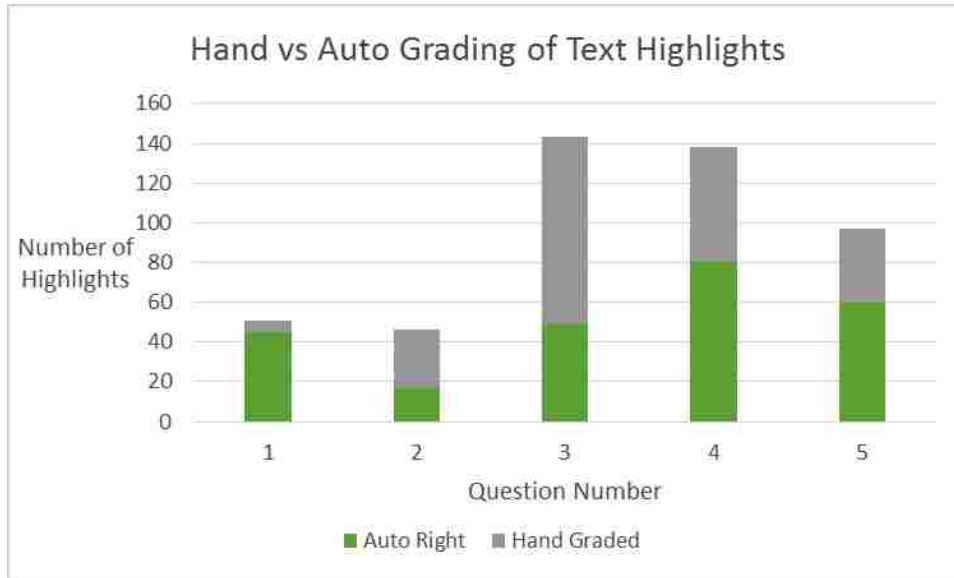


Figure 7-1. A graph of the number of hand graded responses vs. the number of responses that were auto-graded by the machine learning.

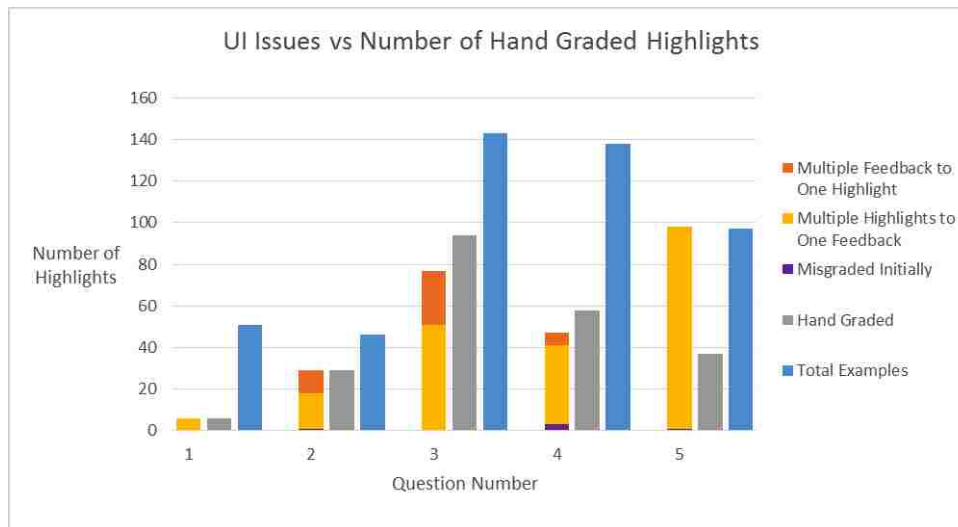


Figure 7-2. A graph with a more detailed breakdown of different problems caused by the user interface and how those problems affect the number of highlights a teacher has to hand-grade.

In Figure 7-2, we can more clearly see a breakdown of what caused an increased number of highlights to be hand-graded by the teacher. In the graph, almost all of the hand-graded highlights are accounted for by

the problems of not being able to attach multiple feedback items to the same highlight and not being able to attach the same feedback to multiple highlights.

For the first question, we see a very low number of hand-graded answers as well as a very low number of problems as seen in Figure 7-3 (a segment of the graph from Figure 7-2). This is because there was one clear single answer that 47 out of the 49 students answered in the exact same way. This question along with all the feedback items used to grade this question are shown in Figure 7-4.

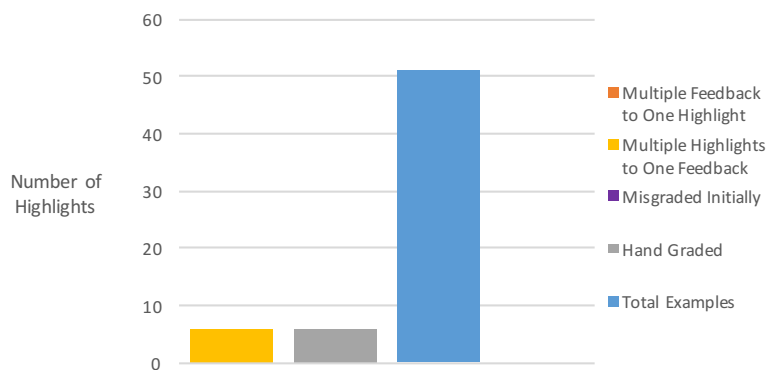


Figure 7-3. Total number of examples vs. user interface issues and hand graded highlights for Question #1.

Highlight in blue the preamble of the constitution.

We the People of the United States, in Order to form a more perfect Union, establish Justice, insure Domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.

Article. I. Section. 1. All legislative Powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.

Section. 2. The House of Representatives shall be composed of Members chosen every second Year by the People of the several States, and the Electors in each State shall have the Qualifications requisite for Electors of the most numerous Branch of the State Legislature.

In this response

- + 1 correct!

Unused

- + .5 first highlight part
- + .5 second highlight part

Figure 7-4. Question #1 and all the possible feedback used to grade all of the responses.

For the second question, there was more variation in answers. There were also multiple combinations of answers, which increased the number of possible output classes (which increases the number of training examples needed). There were also many ways for students to incorrectly answer the question in different ways, which also increased the number of possible output classes. As seen in Figure 7-5, the number of highlights with these different issues matched the number of highlights that were handgraded.

Figure 7-6 shows the question prompt and Figure 7-7 shows some of the common ways students answered the question (either correctly or incorrectly).

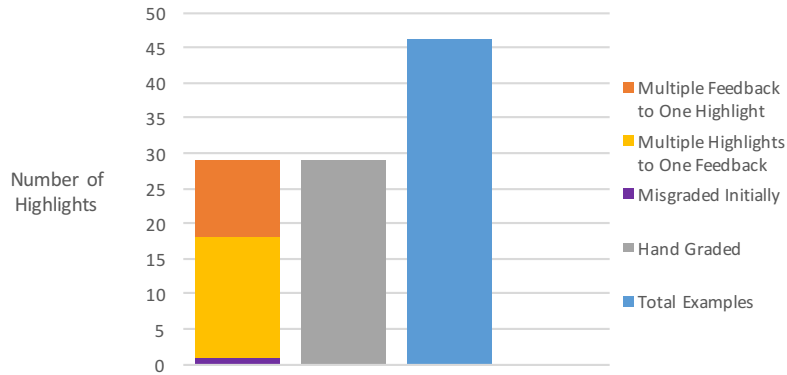


Figure 7-5. Total number of examples vs. user interface issues and hand graded highlights for Question #2.

Highlight in yellow the aspect of Johnson's impeachment that is addressed in Passage 2 but not in Passage 1.

Passage 1	Passage 2
+ 1 Correct third reason for im...	+ 2 Correct 2 and 3rd reason f...
+ 3 Correct - all three reasons...	+ 1 Too broad - whole paragra...
+ .5 Incomplete reason 3 for im...	+ 2 Correct reason 1 and 2 for...
+ -3 Wrong passage selected. ...	+ -3 Incorrect - should have se...
+ 1 Correct first reason for im...	+ 1 Correct second reason for...

Figure 7-6. Question #2 and all the possible feedback used to grade all of the responses.

newly freshened air, for there is none of Johnson's dictatorship to suppress it. Celebrate! Just as the North has overcome the South, freedom has again wrought through oppression, breaking Johnson's stringent hold upon society and letting this country reinstate itself. We are dawning upon a new horizon-- one in which the hopes of those lost may yet be reincarnated.

... sending an axe through the already torn branches of post-bellum society. Andrew Johnson drooped his tail between his legs and sulked out of the White House, carrying with him the stunk of his political legacy. So rejoice! Go out into the streets and enjoy the newly freshened air, for there is none of Johnson's dictatorship to suppress it. Celebrate! Just as the North has overcome the South, freedom has again wrought through oppression, breaking Johnson's stringent hold upon society and letting this country reinstate itself. We are dawning upon a new horizon-- one in which the hopes of those lost may yet be reincarnated.

After violating Congressional osing popular support, and sending an axe through the already torn branches of post-bellum society, Andrew Johnson drooped his tail

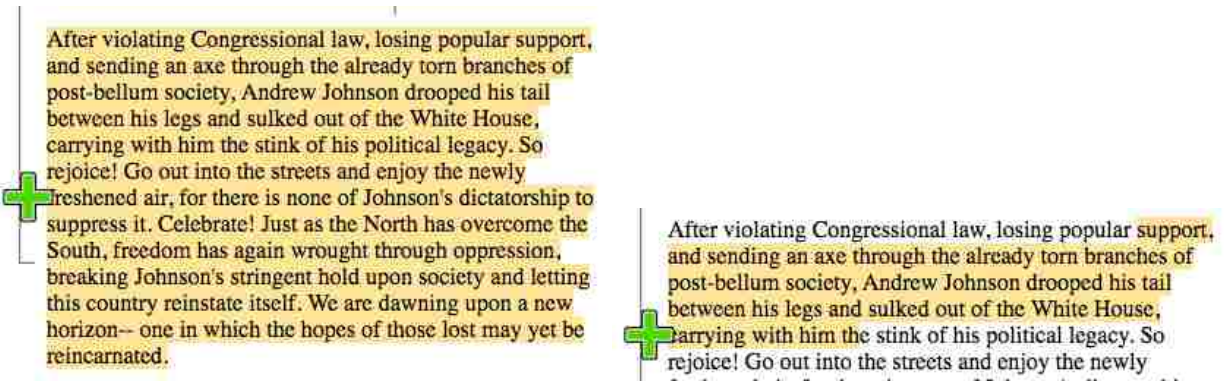


Figure 7-7. Some of the possible ways students answered Question #2.

Question 3 had many unforeseen possible correct answers, as well as many different combinations of these correct answers. Both of these issues increased the number of classes and increased the number of hand-graded items needed to train as seen in Figure 7-8. The list of all the possible feedback items that were needed for this question can be seen in Figure 7-9. In Figure 7-10, we see three examples of possible ways that students answered this question.

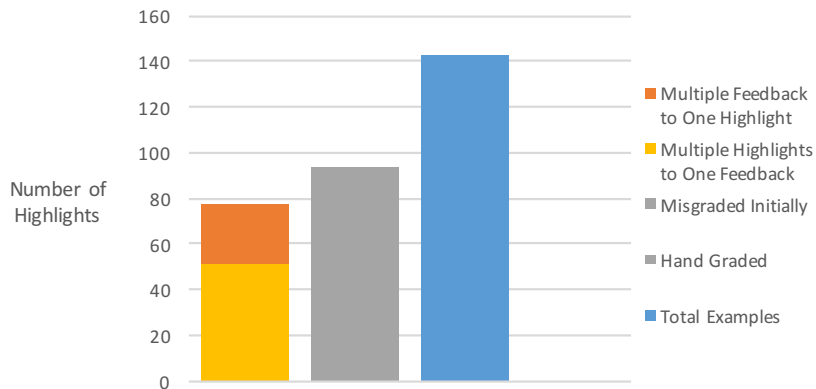


Figure 7-8. Total number of examples vs. user interface issues and hand graded highlights for Question #3.

Highlight in red each of the reasons Mr. Collins gives Elizabeth as he's "proposing" to her.

+ .5	Your portion is unhappily s...	+ 1	Correct reason 1
+ .5	no means discouraged	+ .5	that she said, "Mr. Collins,...
+ .5	little unwillingness	+ -.25	hand is unworthy pt 1
+ -.5	not serious in your rejectio...	+ .5	perhaps it would be advis...
+ .5	My situation in life, my con...	+ -.5	and your wit and vivacity l...
+ 1	chuse and inherit	+ -.25	hand is unworthy pt2
+ 1	Too broad - whole paragra...	+ .5	to inherit this estate
+ .5	modesty, economy, and ot...	+ .5	natural delicacy
+ .5	modesty	+ .5	chuse a wife from among ...
+ .5	it is by no means certain t...	+ .5	to fortune I am perfectly in...
+ 2	Correct reasons 2+3	+ -.5	hand is unworthy your acc...
+ 3	Correct all three reasons	+ 1	Correct reason 3
+ 1	modesty, natural delicacy,...	+ -.5	violence of my affection
+ 1	Correct reason 2		

Figure 7-9. Question #3 and all the possible feedback used to grade all of the responses.

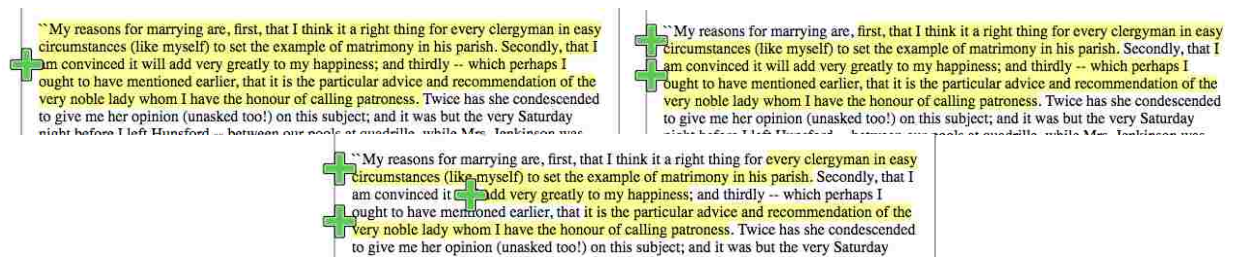


Figure 7-10. Some of the possible ways students answered Question #3.

Question 4 had many issues with not being able to attach the same feedback to multiple highlights when some students would split up their answer into multiple highlights and others would not. This also caused multiple misgrading issues where incorrect feedback was attached to some of those sub-answers and misclassified. Figure 7-11 shows the comparison of these errors for question 4 compared with the number of hand graded highlights. In Figure 7-12 is the question that was asked as well as all the possible feedback the teacher needed to create in order to accurately grade this question. We can see in Figure 7-13 some of the different ways that students answered this question.

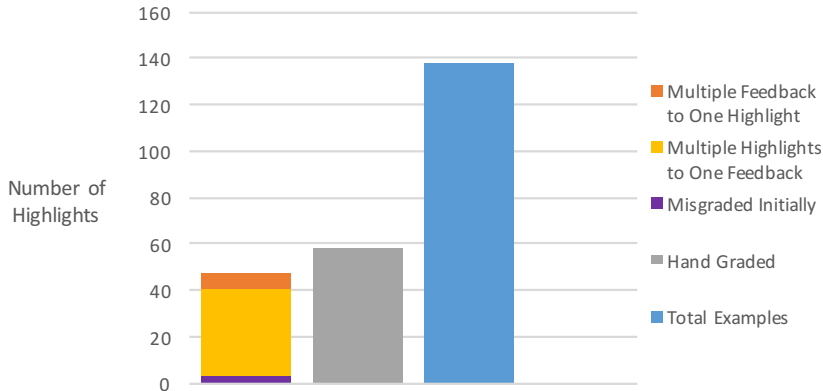


Figure 7-11. Total number of examples vs. user interface issues and hand graded highlights for Question #4.

What fact do all three doctors seem to agree on? Highlight in green the part in each that is consistent with the other two.

+	.5	p 3 pt 1
+	1	correct paragraph 2
+	1	correct paragraph 1
+	-1	incorrect paragraph 3 - kill...
+	-1	intro - uncontrolled cell pr...
+	-.5	incorrect par 2 first senten...
+	-1	incorrect intro paragraph
+	.5	correct paragraph 2 part 1
+	-1	cancer is due to toxins in t...
+	.5	correct paragraph 2 pt 2
+	-.5	killing the cancerous cells ...
+	-1	intro - chemo has many as...
+	1	correct paragraph 3
+	-1	paragraph 1 sentence 2
+	-1	incorrect paragraph 2 - fir...
+	-1	incorrect paragraph 1 - tre...
+	-1	p3 piggybacking
+	-1	incorrect paragraph 3 - les...
+	.5	p3 pt 2

Figure 7-12. Question #4 and all the possible feedback used to grade all of the responses.

Doctor 2:

Chemotherapy does much more harm than good. Chemotherapy might temporarily destroy the cancer, but it does not cure the cancer. In addition, killing the cancerous cells means poisoning the body with chemicals and toxins. Instead, we should be addressing the reasons cancer exists in the first place, treating it at that step. Cancer is due to toxins in the body, industrial pollutions, and drugs. Avoiding sugar, exercising, and maintaining a healthy lifestyle—free of toxins, processed food, and other containments—is the best approach.

Doctor 3:

Chemotherapy is effective in the sense that it kills the cancer cells. The downfall is that chemotherapy also kills the healthy cells in the process; therefore, we should be looking at ways to decrease the amount of chemotherapy needed, so that we are only introducing the minimum amount of toxins into the body. Insulin Potentiating Therapy is a type of chemotherapy in which lower doses of chemotherapy are used because they are combined with insulin. Cancer cells have more insulin receptors than non-cancerous cells; therefor cancer cells will have a biased absorption of such insulin-based chemotherapy when

The slower the effects of the treatment, the more time the cancer has to spread; therefore, while the chemotherapy can kill some healthy noncancerous cells in the process, it is still the best option.

Doctor 2:

Chemotherapy does much more harm than good. Chemotherapy might temporarily destroy the cancer, but it does not cure the cancer. In addition, killing the cancerous cells means poisoning the body with chemicals and toxins. Instead, we should be addressing the reasons cancer exists in the first place, treating it at that step. Cancer is due to toxins in the body, industrial pollutions, and drugs. Avoiding sugar, exercising, and maintaining a healthy lifestyle—free of toxins, processed food, and other containments—is the best approach.

Doctor 3:

Chemotherapy is effective in the sense it kills the cancer cells. The downfall is that chemotherapy kills the healthy cells in the process; therefore, we should be looking at ways to decrease the amount of chemotherapy needed, so that we are only introducing the minimum amount of toxins into the body.

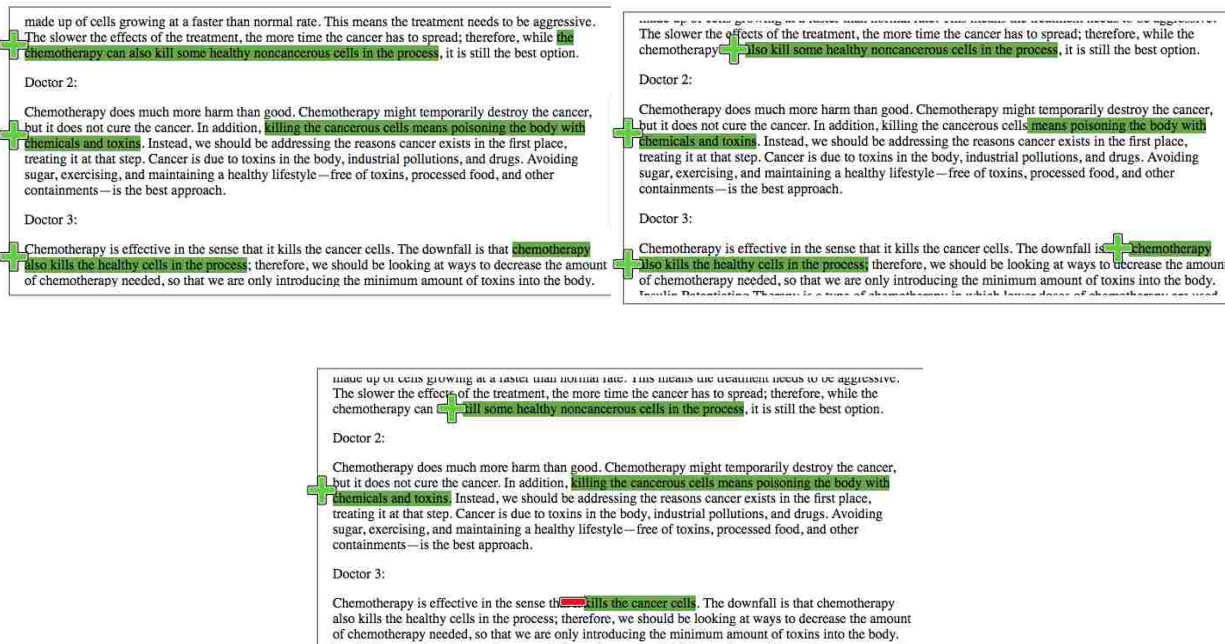


Figure 7-13. Some of the possible ways students answered Question #4.

For the last text question we had students highlight one line of text from two different poems. The question prompt and feedback used for this question can be seen in Figure 7-15. There were no students who highlighted multiple lines of text at once, which is why there were no problems with trying to attach multiple feedback items to one highlight. However, there were many different ways that the students could answer the question correctly or incorrectly as seen in Figure 7-16, which caused there to be a large number of output classes. This again caused there to be more training examples needed to accurately auto-grade the student responses. This can be seen in Figure 7-14 with the comparison of a large number of highlights that had the issue of not being able to attach multiple highlights to one feedback with a large number of highlights that needed to be hand graded.

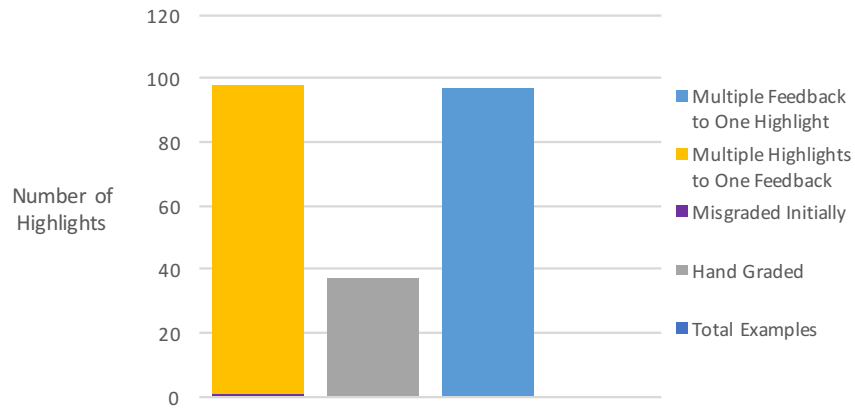


Figure 7-14. Total number of examples vs. user interface issues and hand graded highlights for Question #5.

Highlight in purple the line in each poem by Robert Frost that best illustrates the overall theme.

+	-1	2-30	+	-1	1-7
+	-1	2-37	+	-1	2-38
+	-1	2-31	+	-1	2-14
+	-1	2-9	+	1	2-7
+	-1	2-3	+	.5	2-27
+	-1	2-29	+	-.5	1-1
+	1	1-15	+	-1	2-2
+	.5	2-28	+	-1	1-5
+	-1	2-18	+	-1	2-42
+	-1	2-8	+	-1	2-16
+	-1	1-4	+	-1	2-19
+	-1	2-10	+	-1	2-41
+	.5	1-14	+	-1	2-4
			+	1	1-2

Figure 7-15. Question #5 and all the possible feedback used to grade all of the responses.

<p>Acquainted with the Night</p> <p> have been one acquainted with the night.</p> <p>I have walked out in rain—and back in rain.</p> <p>I have outwalked the furthest city light.</p> <p>I have looked down the saddest city lane.</p>	<p>And there's a barrel that I didn't fill</p> <p>Beside it, and there may be two or three</p> <p>Apples I didn't pick upon some bough.</p> <p> But I am done with apple-picking now.</p> <p>Essence of winter sleep is on the night,</p> <p>The scent of apples; I am drowsing off.</p>
---	--







<p>One luminary clock against the sky Proclaimed the time was neither wrong nor right.</p>	<p>That rumbling sound Of load on load of apples coming in.</p>
<p> I have been one acquainted with the night.</p>	<p>For I have had too much  Of apple-picking; I am overtired Of the great harvest I myself desired.</p>
<p>I have been one acquainted with the night. I have walked out in rain—and back in rain.</p>	<p>As of no worth. One can see what will trouble  This sleep of mine, whatever sleep it is. Were he not gone, The woodchuck could say whether it's like his</p>
<p> I have outwalked the furthest city light.</p>	
<p>I have looked down the saddest city lane.</p>	
<p>I have passed by the watchman on his beat</p>	
<p>I have walked out in rain—and back in rain.</p>	<p>Upon my way to sleep before it fell, And I could tell  What form my dreaming was about to take. Magnified apples appear and reappear, Stem end and blossom end,</p>
<p>I have outwalked the furthest city light.</p>	
<p> I have looked down the saddest city lane.</p>	
<p>I have passed by the watchman on his beat</p>	

Figure 7-16. Some of the possible ways students answered Question #5. The first column are answers from the first poem and the second column are answers from the second poem.

As seen through these different text annotated questions, although the hand grading needed to completely grade a quiz is about 50% of the highlights, almost all of the hand grading is accounted for by the two main errors found in the user interface. In addition to these two errors, regardless, the machine-learning model requires at least a few hand graded examples to be able to accurately train.

Video

Video annotated questions did not manifest any of the above issues in the questions used. Not being able to attach multiple feedback items to one highlight is not an issue. This is because each point in the video is only answering one part of a question, not multiple parts. Not being able to attach multiple answers to the same feedback was not an issue either. With the video questions used, the teacher was asking for one specific part in the video. We thought this could be an issue if the teacher was asking for the students to annotate the video in multiple locations. However, we initially tried grading more specifically with different feedback items and then more generally with only classifying all the correct items with the same feedback, and the training consistently did better or exactly the same with doing more general grading.

With video annotated questions we were much closer to meeting our goal and all the questions were graded in 15% or less of hand graded answers as seen in Figure 7-12. The number of hand graded questions is very close to the number of unique answers that were provided. This is about the minimum number of examples that are needed, which makes sense because each example that is unique must be trained individually. This shows the machine learning did a good job of training on the answers quickly without misclassifying. In Figure 7-13 is a table that shows how many unique answers there were for each question and how many items needed to be hand graded for that question. For each of the questions, the number of annotations hand graded are very close to the number of unique answers which shows that these two issues found in text annotated questions are not found in video annotated questions.

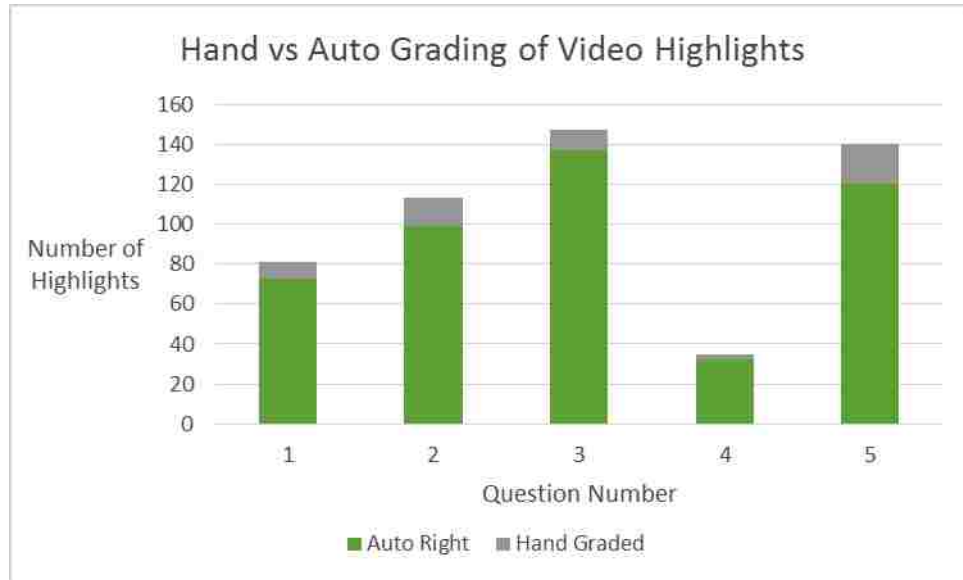


Figure 7-12. A graph of the number of hand graded responses vs. the number of responses that were auto-graded by the machine learning.

#	Unique	Hand Graded
1	10	8
2	13	14
3	7	10
4	2	3
5	15	20

Figure 7-13. Number of unique responses vs. the number of items hand graded.

8. Implementation Issues

Although many of the challenges we have faced are user interface related, many of these issues were made more difficult at times because of the design and framework of the existing system. There were four main implementation issues that we faced. First of all, with the existing Model View Controller (MVC) structure of the code and the way AngularJS works, adding annotated document and video components did not lend itself very cleanly to this structure. Second, being able to find a way for teachers to upload a document that can be integrated into an HTML Iframe proved to be difficult to integrate into the program as well. Next, the process of implementing the notion of “highlights” proved to be difficult for a number of different reasons. Finally, with the Google App Engine server we used, there were some minor issues that caused us to have to do things a little differently than initially planned.

MVC and Angular Issues

Based on the existing structure of the overall program and implementation details from previous research, integrating in another question type into the software was more difficult than initially expected. The existing software is set up to allow the programmer to create new types of questions with different graphical components. Previous components were created to be able to be drawn on an HTML Canvas object. However, document and video components are not able to be drawn directly on an HTML Canvas. Both of these components are HTML based components, not pixel based, and must be added to the DOM structure instead of to a graphical Canvas object.

However, we initially tried to integrate these objects into the existing MVC architecture. There were a list of “Drawable” objects in the model for the Canvas and as they were added they would then be painted onto the Canvas and whenever a change occurred, they would be repainted. We had two “Drawable”

objects – one for a document and one for a video. However, due to the nature of the two components being HTML based, Angular had some major problems with this. One thing Angular does as a safety precaution is to only allow HTML when it is added in a very specific way to a structure. The way we were previous storing the HTML, was not this specific way, and because of this, we ran into many issues that would prevent Angular from updating when it needs to update certain parts. We tried to attack this issue from a couple different angles, with different work arounds, while trying to maintain the document and video objects as part of the Canvas structure. This did not work either. There were still problems when the document and video components would not update when they needed to. Based on these issues and the mix of trying to combine two logical structures, we decide to use the capability of Angular, and re-write most of the text and video annotated components to follow the rules of Angular and remove these two components from the other existing question components that normally work only on the Canvas.

```
QTDocument.prototype.drawNextPrevHighlightWidget = function(){
    if(!this.widgetID) //if there is no id for the iframe, generate an id
        this.widgetID = this.makeId();
    //draw it first
    var outerDiv = document.getElementById("widget_"+this.widgetID);
    if(!outerDiv){
        outerDiv = $("

",{
            id: "widget_"+this.widgetID,
            class: "highlight_widget",
            style: "position:absolute; background:lightgrey; width:35px; height:55px;
        }).appendTo(".addVideosHere");

        this.arrowBase = (50*this.transform)/3.0;
        this.arrowSide = this.arrowBase * (0.75);
    }
    outerDiv = document.getElementById("widget_"+this.widgetID);
    var c = document.getElementsByClassName('layer-anchor');
    var yOffset = c[0].offsetTop;
    var xOffset = c[0].offsetLeft;
    var newTop = (yOffset+ (this.y * this.transform));
    var newLeft = (xOffset + ((this.x + this.width) * this.transform));

    outerDiv.style.top = newTop + "px";
    outerDiv.style.left = newLeft + "px";
    this.drawUpArrow(newTop, newLeft);
    this.drawDownArrow(newTop, newLeft);
    this.drawHighlightFraction(newTop, newLeft);
};


```

Figure 8-1. The old way of writing one little part of the document component.

```

<div ng-style="{{getStyleForWidget()}}"></div>
<div ng-style="{{getStyleUpArrow()}}" ng-click="
previousHighlight(); $event.stopPropagation();" > </div>
<div ng-style="{{getStyleDownArrow()}}" ng-click="
nextHighlight(); $event.stopPropagation();" > </div>
<div ng-style="{{getStyleHighlightFraction()}}" ng-attr-id="
{{setHighlightFraction()}}">{{highlightFraction}}</div>

```

Figure 8-2. The new way of writing the same part of the document component as in Figure 8-1, much more readable and concise. The `getStyle()` methods return dynamic CSS.

Through this re-write we were able to get rid of a lot of excess code and create more readable and functional code. See Figure 8-1 and figure 8-2 for an example of how using the Angular structure is more concise than what we previously had used. In the old version, we had many functions and classes determining positioning and we were dynamically creating and adding HTML Divs and other components, but in the new Angular way, we were able to use a directive (as seen in Figure 8-2) to create a custom HTML element and create much more readable code.

Overall this rewrite was crucial to have the text and video components functioning properly without circular references and dependencies, as well as updating when any of the components changed in some way. We believe that this effort was successful in that we were able to use Angular in the way it was intended to be used and it ended up helping decrease the amount of code, making the code more readable, as well as solving circular dependencies. We still had some minor unsolved Angular errors, however after attempting to fix these issues, we realize that there are some things in the Angular framework that we cannot control and will not be fixing those minor errors, since there is enough functionality in place to be effective for what we are trying to create.

Format to Upload

For the text annotated question, we wanted to provide an easy way for teachers to be able to embed a text file into our application. For whatever format the text came in, we needed to find a way to get the text

into an HTML format to be able embed it into an HTML Iframe. Initially, we toyed with the idea of writing our own converter between a Microsoft Word document and HTML, but soon discovered this could be very time consuming to figure out how to best convert the document, while retaining most of the properties of the original document in HTML. We instead found a public NPM module, mammoth [1] which solved our problem of converting a docx file to an html file.

Mammoth proved to be quite complex to integrate into our system. It is a module still early in development without much support or documentation. The main difficulty was trying to figure out how to get the blob stream from a document that was uploaded to the Google App Engine and use that document stream to be converted to HTML with Mammoth.

However after working on integrating this module into the system, we were able to integrate the capability for a teacher to upload a docx file and store the converted HTML text of the document. This HTML could then be embedded directly into the Iframe for a specific document.

Also, as stated above, there were some issues with Angular initially when trying to store the HTML of the document in the non-Angular way. Angular prevents that ability because of possible security issues.

After the refactor mentioned previously, we were able to effectively store the converted HTML in a safe way that was acceptable to Angular standards.

We believe this solution was successful because a user can upload a MS Word document (currently .docx only), it will be converted to html, and is then able to be set up for highlighting (which we will discuss in the next section). Possible future work that could be done in this area would be to expand to more file formats, being able to upload an html document directly, or provide a URL and choose which part of the webpage would be embedded into the Iframe.

Implementing Text Highlights

The process of implementing highlights on both text and video was a challenge and took a couple different attempts to figure out the best way that would work the best. The different parts to this challenge include preparing the text for the highlights, how to best store the highlights, and how to best integrate them into the HTML Iframe.

Before we could add highlights into the document, we had to somehow keep track of where each highlight is located. We created a custom attribute “data-tagid” that was added to each HTML tag as a preprocessing step with an increasing index value on each element. This would be used to keep track of which div or span a highlight starts and ends.

After this preprocessing step, when a student selects a certain portion of the Iframe, we take the selection, and calculate the start and end div values as well as the start and end position within each of those divs. For example a highlight may start on div 4 position 27 and end on div 6 position 3. After calculating these positions and taking into account the color used by the student, this new “highlight” is added to an overall list of highlights. Before, we tried to add a new selection immediately into the Iframe after something was selected, however, since highlights may continually be added out of the order of the document, the highlights were running over each other and overlapping with each other. So we determined that we needed to store the list of highlights, and then sort that list from the bottom of the file to the top. If highlights could be sorted from the bottom to top and then added, we could get around this issue of overlap.

After sorting the list of highlights, we can then add the highlights into the Iframe. This is done by creating a new span around the selected text and changing the background color of that span. This is simple, but it is essential that is done in the correct order or the spans will overlap and the start and stop positions will get misaligned and end up highlight the wrong text. With the updated Angular way of

doing things, the highlight implementation became much simpler to automatically update the highlights when any changes happened (i.e. color change, or deleting a highlight).

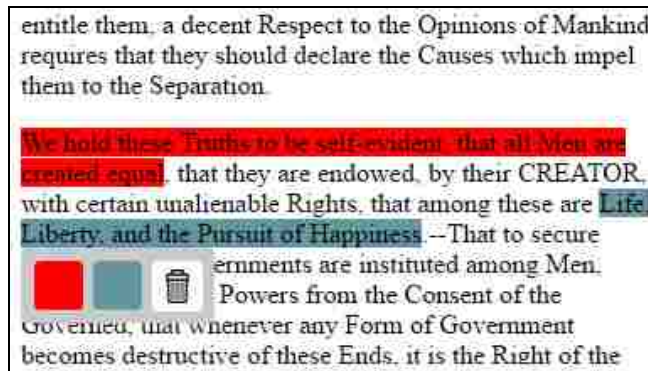


Figure 8-3. An example of a tooltip on a highlight.

In addition, when a highlight is added, it also adds a corresponding tooltip as seen in Figure 8-3 that is initially hidden until the highlight is hovered over, and then allows the student to change the color of the highlight or remove it. These tooltips were also initially implemented in a much more complex way in the previous structure, but with the Angular refactor, we were able to create a tooltip directive (a custom tooltip tag) which was able to be used whenever a new highlight was added and is much simpler. This not only helped with making the position and style of the tooltips much more simple, but also for showing and hiding certain tooltips at certain times.

Implementing Video Highlights

Video “highlights” or annotations were stored similarly but implemented quite differently. Each video has a list of “highlights” with a position and color. We use the Youtube API in order to determine the point where the video is paused and add a marking in that specific position. Order is not as important in this case, because we do not have to worry about overlapping highlights.

We believe we have been successful in solving this problem for a few reasons. First, highlights can be added anywhere in a text document or along a video timeline without messing up any other highlights before or after the current highlight. Second, the functionality to remove a highlight from a text document or a video is effective and does not change any other existing highlights. Also, a student can change the color of a specific highlight after it has been added for both a text highlight and a video highlight. There is also functionality to be able to extend an already existing document highlight by selecting over a portion that has already been highlighted. This was extra functionality added toward the end of implementation that users expected/desired. We believe this is a hard problem and feel we have been successful in implementing the idea of highlights for a text document and video.

Google App Engine Issues

Google App Engine was the server and data store already in use for this software. Previously it was only used to upload picture files. However, with the addition of a video annotated question, meant that bigger file sizes would be uploaded. We found that uploading a video over a certain size will not upload or play correctly. The video size is quite limiting and any video over about 30 seconds in length is usually too big to upload and completely play. With this issue, not being in the position to re-implement the whole backend of the existing system, we decided to go another route. Instead, we take the URL of a YouTube video and simply embed the video into our application, instead of serving the URL ourselves. This way we are able to embed a video of any length without any worry of length of the video being too long.

We believe this is a viable option for the time being. The scope of this research does not include rewriting an entire backend and server, however in the future this would be something important to do. It would be very useful to be able to serve video of any length and in any format (currently we can only embed videos that are on YouTube). For now, allowing YouTube to serve the videos and embedding them into our site is sufficient and we are still able to allow students to annotate those videos.

9. Summary

The problem we attempted to solve was to create a new type of question for teachers and students that is quickly gradable, provides feedback quickly to students, and is intuitive and easy for teacher to create and students to take. This problem of having to create and grade questions that are a good test of knowledge without taking hours to grade is a big pain point for many teachers. Many teachers spend a lot of time grading tests and assignments and if this time could be shortened, it could be used to help students better understand difficult concepts.

Our solution was to create two new types of questions: text and video annotated questions. Each of these questions can be easily created, easily taken, and somewhat easily graded. There are still some major limitations in the user interface for grading, however, even with these limitations, a teacher at the most has to grade about 50% of the students' responses and the software we have developed will automatically grade the other 50%. In the best-case scenario, a teacher only has to grade a little less than 10% of the student answers and the rest will automatically be graded. Overall we met most of our main goals, except for the fact that the grading interface could be revamped to help decrease the number of student responses graded even more.

There is potential for future work. First of all, through even more user studies, we could refine the overall flow of taking a quiz to make the whole experience more intuitive and consistent (for this research we only focused on the process of taking questions, not the flow of the entire software). We could also improve some of the UI components that were feature improvements some students would have liked. This includes adding functionality to be able to slide a video annotation to a different after it was added, instead of having to remove the annotation and then add another one. Another component that could be

added would be to include a widget on a video annotated question similar to a text annotated question for the teacher to be able to move from one annotation to another while grading a quiz.

Another aspect of this that could be focused on in future work would be to provide a better video server. We realized late in the process that the Google App Engine does not have the capability to serve and play large video files. This was an issue we worked around by allowing YouTube to do the serving. However, in the future we would like to be able to allow a teacher to upload any video and be able to embed it into a question.

In the future, in addition to solely being able to add videos, we would like to add the functionality for audio clips to be uploaded and played. This could include things like having students review a teacher's lecture and marking different points in it. Most of the functionality would be able to be reused, but some of the UI components would need to be modified and/or added.

Another approach to solving this problem would be to have users interact differently with the interface than only using a mouse and keyboard on a computer. These approaches could include using a stylus on a touch screen to highlight text, or only allow users to highlight with their fingers on a touch screen. These methods have the potential of being more intuitive for students because with touch screen interfaces, there is typically less button pressing and more direct interaction with elements on the screen. For teachers the touch screen interface may be helpful as well because there is not much typing involved with creating or grading (only to write the question prompt or to add feedback), and the touch screen interface could be more intuitive to select different highlights to grade or widgets to add to the screen.

For the video questions, users could also potentially watch a video and give voice commands when to stop. This may be a feasible option, however, as many students do their homework in a library or a setting with other students around, it may not be particularly useful to this type of application.

Finally, in discovering the limitations of the interface for teachers to grade questions, future work that could be done would be to completely revamp the structure and interface for grading to allow teachers to

be able to add multiple feedback items to one highlight as well as allowing teachers to add the same feedback item to multiple highlights.

10. Bibliography

1. Lewis Baumstark and Edwin Rudolph. 2013. Automated online grading for virtual machine-based systems administration courses. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*. ACM, New York, NY, USA, 477-482. DOI=10.1145/2445196.2445340 <http://doi.acm.org/10.1145/2445196.2445340>
2. Aaron Bloomfield and James F. Groves. 2008. A tablet-based paper exam grading system. *SIGCSE Bull.* 40, 3 (June 2008), 83-87. DOI=10.1145/1597849.1384295 <http://doi.acm.org/10.1145/1597849.1384295>
3. Michael Brooks, Sumit Basu, Charles Jacobs, and Lucy Vanderwende. 2014. Divide and correct: using clusters to grade short answers at scale. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. ACM, New York, NY, USA, 89-98. DOI=10.1145/2556325.2566243 <http://doi.acm.org/10.1145/2556325.2566243>
4. Mary Margaret Couse. 1996. Creating an elegant, intuitive, user interface. In *Proceedings of the 13th Annual International Conference on Systems Documentation: Emerging from Chaos: Solutions for the Growing Complexity of our Jobs (SIGDOC '95)*. ACM, New York, NY, USA, 26-33. DOI=10.1145/223984.223991 <http://doi.acm.org/10.1145/223984.223991>
5. Jerry Alan Fails and Dan R. Olsen, Jr.. 2003. Interactive machine learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI '03)*. ACM, New York, NY, USA, 39-45. DOI=10.1145/604045.604056 <http://doi.acm.org/10.1145/604045.604056>
6. Fascinating Facts about the Declaration of Independence, 1996. Retrieved September 9, 2015, from ConstitutionFacts.com: <http://www.constitutionfacts.com/us-declaration-of-independence/fascinating-facts/>.
7. Elena L. Glassman. 2014. Interacting with massive numbers of student solutions. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST'14 Adjunct)*. ACM, New York, NY, USA, 17-20. DOI=10.1145/2658779.2661167 <http://doi.acm.org/10.1145/2658779.2661167>
8. Natalia Juristo and Xavier Ferre. 2006. How to integrate usability into the software development process. In *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*. ACM, New York, NY, USA, 1079-1080. DOI=10.1145/1134285.1134515 <http://doi.acm.org/10.1145/1134285.1134515>
9. Kevin Mullet and Darrell Sano. 1994. Applying VisualDesign: trade secrets for elegant interfaces. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*, Catherine Plaisant (Ed.). ACM, New York, NY, USA, 353-354. DOI=10.1145/259963.260445 <http://doi.acm.org/10.1145/259963.260445>
10. W. James Popham. *Classroom Assessment: What Teachers Need to Know*. Pearson Education Publishers. Boston, 2011.
11. Kasper Rodil, Heike Winschiers-Theophilus, Colin Stanley, Gereon Koch Kapuire, and Matthias Rehm. 2014. An approach to user interface design with two indigenous groups in Namibia. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 460-469. DOI=10.1145/2686612.2686686 <http://doi.acm.org/10.1145/2686612.2686686>

12. Karen Scouller. 1998. The influence of assessment method on students' learning approaches. *Higher Education, Springer*. 35(4):453-472.
<http://link.springer.com/article/10.1023/A:1003196224280>
13. Chris Wilcox. 2016. Testing Strategies for the Automated Grading of Student Programs. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16)*. ACM, New York, NY, USA, 437-442.
14. Shunsuke Noguchi and Naomi Fujimura. 2015. Implementation and Experience of the Online Peer Grading System for Our Real Class. In *Proceedings of the 2015 ACM Annual Conference on SIGUCCS (SIGUCCS '15)*. ACM, New York, NY, USA, 125-128.
DOI=<http://dx.doi.org/10.1145/2815546.2815581>
15. Chris Wilcox. 2015. The Role of Automation in Undergraduate Computer Science Education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)*. ACM, New York, NY, USA, 90-95. DOI=<http://dx.doi.org/10.1145/2676723.2677226>
16. <http://www.cpalms.org/Public/PreviewResourceLesson/Preview/40021>