Flipping the Classroom in CS1101
An Interactive Qualifying Project Report

in partial fulfillment of the requirements for the
Degree of Bachelor of Science at
Worcester Polytechnic Institute

Submitted to:
Professor Joseph Beck, Advisor

By:

______________________________
Michael Andrews
Computer Science Class of 2016
Abstract

The objective of the “Flipping the Classroom in CS1101” Interactive Qualifying Project is to study and improve the use of the ASSISTments online teaching system in CS1101, an entry level course in Computer Science at Worcester Polytechnic Institute. This report details the project, analyzes data collected during a number of studies, and further explores options for improving the course.
Contents

1 Introduction 2

2 Background 3

3 Methodology 6
   3.1 Problem set improvements 6
   3.2 Studies and research 7
       3.2.1 Study One 8
       3.2.2 Study Two 8
       3.2.3 Study Three 9
       3.2.4 Results 9

4 Results and Analysis 10
   4.1 Changes in CS1101 Online Content 10
   4.2 Study Results and Analysis 11
       4.2.1 How do comprehension questions enhance student learning? 11
       4.2.2 How do comprehension questions affect student confidence? 11
   4.3 Correlation Details 12
       4.3.1 Quiz / Self Evaluation Correlations 12
       4.3.2 Quiz 3 Correlation Tables 14
   4.4 Parameter Estimates for Post-Quiz Self Evaluation 14

5 Conclusion 15
Chapter 1

Introduction

*Flipping the Classroom in CS1101* is an Interactive Qualifying Project studying and improving the use of supplementary online teaching systems at *Worcester Polytechnic Institute*. In particular, we audited the quality of an existing assistive solution, sought comprehensive feedback on the system, and studied the effects of its use on various measures of confidence and comprehension. Through multiple studies, we collected a volume of data relating problem understanding, confidence in knowledge, and the use of interactive comprehension questions during and after video lectures. Beyond studying the effects of video lectures and comprehension questions, we improved the quality and organization of existing questions and located other online resources for use in the course. In this paper, we detail the execution of the project, and we outline a number of significant findings in our collected data.

The focus of this project, the ASSISTments online teaching system, is an interactive online tutoring system with a variety of features and uses. ASSISTments is a flexible system, whereby teachers may divide their classes, organize homework assignments, integrate arbitrary web resources into homeworks, and receive detailed reports on students' answers and performance. ASSISTments is used by more than 600 teachers globally, and may represent a significant step towards intelligent interactive online tutoring systems.

CS1101 is the introductory course to Computer Science at *Worcester Polytechnic Institute*. CS1101 uses the Racket programming language, a scheme variant, to teach the basics of data structures, program design, and algorithms. CS1101 students used ASSISTments to study and complete assignments online.

This project was twofold. In the first part, we significantly improve the quality of CS1101’s online homework system through thorough auditing, student polls, student comments, and professor feedback. Investigating and improving ASSISTments was the primary goal of this project. In the second part, we perform three sets of trials over the course of the term to analyze the effects of online homeworks and lectures on confidence and performance. We hope to use this information to determine how effective online lectures and homeworks are and how to improve them.
Chapter 2

Background

For the duration of this project, ASSISTments was used to reflect the course work of CS1101. Each week, the students attend lectures, attend a lab, submit homework, and watch video lectures while answering comprehension questions online. This project sought to determine the efficacy of integrated online lectures, and in particular, the efficacy of adjoined comprehension questions. As the online tutoring system was both integrated with and refined to the specifics of the course, CS1101 made an ideal candidate for studying the effects of online tutoring systems.

ASSISTments had been used with CS1101 previously, and the C15 offering made use of the previous session’s assignments. These assignments typically consisted of an embedded video lecture and a number of multiple choice or short answer questions related to the video. These questions were typically short, requiring between five and ten minutes and dealt with understanding, evaluating, and completing segments of code and interpreting relevant vocabulary. As online assignments may be used multiple times without significant alteration, the previous assignments acted as a pool of relevant preexisting assignment material for each new session. Different offerings of CS1101 may take slightly varying paths throughout the term, the course staff is able to select relevant assignments without recreating problem sets. While assignments enjoyed some improvements and modifications, the existing base of problem sets proved to be a valuable asset to the staff of CS1101.

Beyond providing an efficient means of administrating and reusing daily course assignments, ASSISTments is capable of intelligent problem handling. ASSISTments provides a general framework for defining problems, problem sets, course-wide and fine grained problem set assignment, multiple response evaluation methods, and many more relevant features. These capabilities not only allowed the efficient execution of this project, but also semi-intelligent tutoring for every assignment for each student. This course organized its assignments into multiple problem sets, generally arranged around a relevant video lecture. Each night, there may be multiple assigned problem sets, depending on the breadth of the day’s lecture. These problem sets contained a list of problems.
An image with a recursive structure is called a \[\text{__________}.\]

Figure 2.1: Example text response question in a typical ASSISTments assignment.

In generative recursion, what is done first?
• Base Case Test
• Self Reference Test
• Recursive Test
• Function Definition
• Stub

Figure 2.2: Example multiple choice question in a typical ASSISTments assignment.

Each problem was of a type, multiple choice, multiple answer, text response, or open response. ASSISTments automatically shuffles answer indices to dissuade dishonesty. Using ASSISTments’ problem mark up editor, HTML embedding, and the available question type, creating, assigning, and evaluating assignments was efficient, regardless of assignment type or content.

CS1101 ASSISTments problems can be typified in Figure 2.1 and Figure 2.2. Such questions would be asked in sequence, and would either include or follow an embedded video lecture. These two questions were asked in the C15 video lecture assignment following a video explaining recursion and fractals. ASSISTments provides graders with a report on class performance and individual answers. It automatically grades questions with defined answers, and allows for multiple correct answers in the case of text response (to allow for minor formatting differences.) There may be a future project dealing with racket code equivalence and other useful question types. The history and content of the supplemental ASSISTments material is thoroughly reviewed in [1].

As mentioned previously, the video lectures used CS1101 ASSISTments problem sets were provided by Systematic Program Design, a series created by Gregor Kiczales for use in another online course. CS1101 staff were able to embed these videos into the assignments directly without additional support from the ASSISTments maintainers, as ASSISTments allows arbitrary HTML tags in problem bodies. YouTube, a video streaming service, hosts the video lectures and provides a convenient embedding API. Using embedded frames with arbitrary size, and the youtube embedded video API, problems could include
properly sized and properly timed relevant videos without implementing a larger hosting infrastructure or modifying ASSISTments. A number of questions in the assignments included images and other elements from the web. These were all available as online resources to the ASSISTments system.
Chapter 3

Methodology

This project encompassed multiple goals and results. The later work was heavily focused on study and analysis, while the early work focused on assessment and improvement. This chapter will discuss both the improvements and modifications to the ASSISTment assignments over the project as well as the experiments and their methodology.

3.1 Problem set improvements

One of the primary goals of this project was to directly improve the quality of the existing online assistive homeworks in use in CS1101. To achieve this, we talked with course staff, interviewed the students, analyzed student comments, and reviewed the problem sets manually. After learning of problems and potential improvements, we tested corrections and solutions, and corrected all available problem sets prior to assignment to the class. This gave the students a less confusing and more coherent experience with the online learning materials.

Beyond corrections, there were a number of basic format and usability improvements we were able to perform. Some basic corrections were applied to every problem throughout the term. These included usage directions, improvements on the video and image embedding, and some other formatting notes. There are a number of specific examples. A few to note are the following:

- Videos were inconsistently sized. Larger videos were problematic for small screens. All videos were corrected to be 640x480.

- Due to ASSISTments’ testing structure, the user was often unclear of whether or not to play a video. All ambiguous situations were given clear direction above the video and problem.

- Video intervals were sometimes included. This allows students to view a particular part of a lecture dealing with the specific question. Multiple inconsistencies were corrected.
• We originally discussed autoplay for simpler direction. After polling students, it was determined that this was unpopular, and was removed.

These modifications made the questions much more understandable.

Beyond formatting improvements, we also audited all assigned materials for corrections and potential improvements. While searching comments, talking with students, and reviewing material, we found and corrected a number of mistakes including the following:

• Spelling mistakes, grammar mistakes, or otherwise unclear questions.
• Incorrect answers, multiple correct answers, or questions irrelevant to the problem set.
• Problem sets embedding the incorrect video lecture.
• Timing intervals and video breakup being unrelated to the problems being asked.

CS1101 had 82 available video lectures to be used throughout the term. Correcting these issues and reviewing all problem sets for the course was a nontrivial task, and multiple other varieties of errors were discovered. The majority of mistakes were minor, but there were cases where assignment problems caused the class and course staff a significant deal of grief. The primary example for this was the incorrect embedded video lecture. This lead to an entire class of confused students.

This process of gathering feedback, reviewing, improving, and clarifying was of great value to the course. There are not exact numbers for the volume of modifications. We are able to say that multiple modifications were made to each problem set, typically including many formatting improvements and some minor corrections.

3.2 Studies and research

The second aspect of this project was directly studying the effects of online learning with measures of confidence and performance. To assess the effects of combining subject comprehension questions and video lectures, we ran a number of trials using the ASSISTments system.

The design of the trials is as follows:

Each participant was assigned to one of four groups, denoted groups A, B, C, and D. For each trial, two groups were selected as a control, and two were selected for the experimental treatment. For each trial, the control group was given the video lectures to watch, while the experimental group was given the videos paired with relevant comprehension questions. At the end of each video, they were asked to rate their understanding of the material. At the end of each round of trials, both groups were given a short quiz. This quiz
included one or more open-ended comprehension questions. After
the quiz, students were again asked to rate their understanding of
the material.

This design allows us to control for longterm performance gains and effects
between trials. The primary condition of these trials was the effect of added
comprehension questions on the performance and confidence of students.

The students were assigned to one of four groups, A, B, C, and D. There
were approximately thirty students in each group. For the video lectures and
the comprehension questions, we took existing assignments and created a con-
trol problem set, without the comprehension questions. For the control group,
this was simply the video lecture, along with a confirmation of completion. The
experimental group watched the lecture in appropriate sequential segments, an-
swered comprehension questions, and completed the problem set as per usual.
The quiz was written from scratch for each trial. It included comprehension
questions related to the material, typically involving a lengthy or creative re-
sponse. These questions were manually scored and entered into a spreadsheet,
where they were made available for statistical analysys.

Over the term, we conducted three separate trial sets. Outside of the trials,
the students continued their regular course of video lectures and comprehension
questions. The intention and format of each trial was guided by the previous.
The first trial set was midway through the term, the second took place two
weeks later, and the final study was held one week after the second. The second
and third sets had self evaluations on each trial and the final quiz, whereas the
first only had self evaluations after each trial.

3.2.1 Study One

The first study used post test questions not much more complicated than regular
comprehension questions. This particular topic was on defining local variables,
and was mostly concerned with details of the language. While watching the
previous video lectures with comprehension questions, students were asked to
rate their understanding. After the question, students were given the full answer,
and continued without again rating their understanding. The students were
arranged into AB for treatment and CD for control.

3.2.2 Study Two

The second study improved on the methods of the first. This study concerned
itself with fractals and generative recursion. While watching the previous video
lectures the students had again rated their understanding. After the first study,
we decided to add an additional post-quiz self rating. This was useful in deter-
mining changes in confidence when faced with involved, open ended questions
on the new subject. The students were arranged into AD for treatment and BC
for control.
3.2.3 Study Three
This study was more involved than study two, and included two related post test comprehension questions. The first was a termination argument for a sudoku solving algorithm. The second was completing a partially implemented recursive algorithm. This study was chronologically sequential to study two, and also used AD for treatment and BC for control.

3.2.4 Results
The self evaluations from each video, the performance on each quiz, and the self evaluation after each post test quiz were compiled into a database. As student proficiency is almost always the dominant predictor in educational research, the CS1101 professor made the initial course performance available, and it was added to the database. The initial course performance was an average of quiz performance for the first quizzes of the course. We used SPSS statistics software to search these datapoints for meaningful correlations and potential causality. We detail the results from this analysis in the Results and Analysis chapter.
Chapter 4

Results and Analysis

The results of this project are twofold. First, there are the tangible changes to CS1101 and its online component. Over the project, a number of improvements were made to the online homeworks. Secondly, there are the results from the three studies conducted using the ASSISTments system. Using the results of the study, we are able to search for interesting correlations and patterns. In this section, we present both the results of tweaking CS1101’s ASSISTments assignments, and an analysis of the three studies.

4.1 Changes in CS1101 Online Content

This project included a number of tasks to gauge the efficiency of online homeworks and to use this information to improve the quality of the course. Among these tasks were subtle changes, such as catching grammatical errors and correcting confusing questions, and more noteworthy changes, such as correcting mistaken answers. Along with various tweaks in formatting and correctness, we added some new essay questions, gathered all related online materials into an easy to access spreadsheet, and fixed a number of problems with the old online lectures. We also accepted student comments through ASSISTments and through other channels, which helped us notice other problems with the homeworks. By auditing these lectures, we were able to improve the value of the online assignments, correct issues with incorrect videos or questions, and make the lectures more coherent.

Along with manual auditing, we polled the students, and found the number of problems and other suggestions for the system. Beyond formatting, correctness, and other glitches in the homeworks and ASSISTments’ interface, the students were also interested in integrating Racket’s runtime environment, possibly as an interpreter, with the ASSISTments assignments. This would allow them to test their code, without switching between DrRacket, Racket’s IDE, and their browser. By providing a complete sandbox for students to test their code in, students may spend more time playing with Racket in the context of questions,
rather than trying to copy over given functions to complete templates. Along with an auto-grading system, there is now an independent project investigating these possibilities.

### 4.2 Study Results and Analysis

Once the study results were compiled into a manageable database, we were able to search for correlations and ask questions about the data. Before analyzing the data, we outlined a number of points of interest. These points were used to guide our search and explain the effects of comprehension questions when used with online video lectures. The answers to these questions explain the role of comprehension questions, their efficacy in this situation, and outline our topic of interest. For each question, we present figures and tables from the collected data to help communicate the answer.

#### 4.2.1 How do comprehension questions enhance student learning?

To determine the immediate educational effects of comprehension questions, we analyzed the performance of each group, and determined if there was a statistically significant gain for the experimental group. As a somewhat surprising result, over the three studies, the experimental group’s quiz performance did not differ from the performance of the control group in a statistically significant manner. Looking at the data, it became clear that the coupled comprehension questions had little significant or immediate impact on the quiz results over unaccompanied video lectures.

While there was no measured significant impact on performance, we go on to analyze the effects of comprehension questions on subject confidence, and its relation to actual student performance.

#### 4.2.2 How do comprehension questions affect student confidence?

Throughout the study, we asked the students to rate their understanding on a scale from 1 - 10. Using this data, we can explore the effects of comprehension questions on self rated confidence in the material. As in the above section, each question is answered using the data collected during the studies.

**How do comprehension questions affect accuracy of confidence?**

The final two trial sets included self evaluations directly after their respective quizzes. [Table 4.1](#) and [Table 4.2](#) give a basic idea about the correlation between confidence and performance for these questions for both groups. When split into two groups, as shown in [Table 4.3](#) and [Table 4.4](#) there is a slight difference in in the correlation. The difference between the two correlations is unreliable,
but for both trials, it is lower for the control group, possibly prompting further investigation. Table 4.5 and Table 4.6 show the parameter estimate for quiz self evaluations, based on initial quiz scores and the control condition. In both cases, the control condition was negative, but statistically unreliable. This result may prompt further investigation.

How do self evaluations change over time?

In each set of trials, there were multiple successive video lectures and self evaluations. When studying the data, we found no significant trend in self evaluations in successive questions.

4.3 Correlation Details

4.3.1 Quiz / Self Evaluation Correlations

In the following sections, we include a number of tables exploring the correlation strength between measures of confidence and ability during the study, controlling for various factors and separating between experimental and control groups.

Overall Score and Self Evaluation

Table 4.1 shows the correlation between quiz two scores and self evaluation. This table includes the results from both groups, and controls for the initial quiz scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>Correlation</td>
<td>0.397</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 4.1: Experiment Two Overall Score/Self-Eval Correlation

<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>Correlation</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 4.2: Experiment Three Overall Score/Self-Eval Correlation

<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Correlation</td>
<td>0.392</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>0.005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental</th>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.406</td>
<td></td>
</tr>
<tr>
<td>Significance</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Experiment Two Per-Group Score/Self-Eval Correlation
<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Correlation</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>0.037</td>
</tr>
<tr>
<td>Experimental</td>
<td>Correlation</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 4.4: Experiment Three Per-Group Score/Self-Eval Correlation

Table 4.5: Parameter Estimates for Experiment 2, Self Evaluation, Initial Quiz Scores

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>t</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Eta Sqr.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Intercept</td>
<td>5.842</td>
<td>0.958</td>
<td>6.100</td>
<td>0.000</td>
<td>3.941</td>
<td>7.742</td>
</tr>
<tr>
<td>InitialQuizScores</td>
<td>0.198</td>
<td>0.121</td>
<td>1.637</td>
<td>0.105</td>
<td>-0.042</td>
<td>0.439</td>
</tr>
<tr>
<td>Control Cond.</td>
<td>-0.659</td>
<td>0.388</td>
<td>-1.699</td>
<td>0.092</td>
<td>-1.429</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Table 4.6: Parameter Estimates for Experiment 3, Self Evaluation, Initial Quiz Scores

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>t</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Eta Sqr.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.442</td>
<td>0.967</td>
<td>4.592</td>
<td>0.000</td>
<td>2.524</td>
<td>6.360</td>
</tr>
<tr>
<td>InitialQuizScores</td>
<td>0.358</td>
<td>0.121</td>
<td>2.969</td>
<td>0.004</td>
<td>0.119</td>
<td>0.597</td>
</tr>
<tr>
<td>Control Cond.</td>
<td>-0.410</td>
<td>0.376</td>
<td>-1.090</td>
<td>0.278</td>
<td>-1.156</td>
<td>0.336</td>
</tr>
</tbody>
</table>
scores. Along with the correlation factor, we include the two tailed significance test. This included 99 data points, evenly distributed between the groups.

**Per-Group Score and Self Evaluation**

Table 4.3 demonstrates the quiz two score and self evaluation correlations and their significance for each group in the trial. There were 47 members of the control group and 49 members of the experimental group. Table 4.3 depicts the quiz three score and self evaluation correlations in both the experimental and the control group, controlling for the initial test scores.

### 4.3.2 Quiz 3 Correlation Tables

**Overall Score and Self Evaluation**

Table 4.2 details the overall correlation between quiz three scores and self evaluation. It follows the same format as Table 4.1 and includes 105 subjects. Along with the quiz two overall correlation analysis, this controls for the initial test scores of each subject.

**Per-Group Score and Self Evaluation**

Table 4.4 represents the quiz three score and self-evaluation correlation for both the control and experimental group. There were 51 subjects in each group. As with all of the previous correlation analysis, this table controls for the initial test scores for each student. The quiz three per-group analysis follows the same format as Table 4.3.

### 4.4 Parameter Estimates for Post-Quiz Self Evaluation

In this section, we show some linear parameter estimates and their significance for both quiz two and three self-evaluations. Table 4.5 estimates quiz two self-evaluation as a linear function between the initial quiz scores and the control condition. Effectively, it estimates the self-evaluation, on a scale from 1 to 10, as $((0.198) \ast \text{initial} + (-0.659) \ast \text{control} + (5.842))$. Table 4.6 does the same for quiz three, and it estimates quiz three self-evaluation as $((0.358) \ast \text{initial} + (-0.410) \ast \text{control} + (5.842))$. In both cases, the control condition had a negative factor. While this is an interesting result, it is not statistically reliable. These results, if demonstrated statistically reliable under replication, would suggest that comprehension questions help students calibrate confidence in their knowledge of the material. In Table 4.5, the 0.198 parameter for initial quiz performance indicates that students with stronger initial quiz performance had higher self evaluations.
Chapter 5

Conclusion

This project succeeded in reaching its original goals. While first focusing CS1101’s supplementary online homeworks, this project grew to include a brief study on the effects of video lectures and homework questions on student performance and confidence. As its primary goal, this project significantly improved the quality and relevance of CS1101’s ASSISTments assignments. Over the course of the term, this project used student feedback and careful auditing to improve the online homework and lecture system, correcting a variety of problems, adding multiple new questions, and organizing all video lectures. As a secondary goal, this project unveils surprising results relating to the efficacy and miscellaneous effects of mixing comprehension questions and online lectures. While multiple interesting trends were not statistically reliable, this study did illuminate a number of potentially useful relationships, as outlined in the results section.

There were a variety of Improvements to CS1101’s ASSISTments based online homework and lecture system. We started by fixing minor bugs in the problem sets, accumulating student feedback, and curating comments submitted through the online system. We enumerated and analyzing student problems, we were able to identify problem questions. From there, we improved the formatting of the homeworks, making the assignments more clear. To help organize the course, we collected all related online lectures from the lecture source, organized them by subject and order, and linked them to existing problem sets in the ASSISTments online system. Along with the studies, we produced a number of comprehension and essay questions for use by the course. While these changes may seem minor, each either addressed concerns expressed by students, or helped the organizers in tangible ways. By making online resources available in the form of easily accessible spreadsheets, linking their URLs, ASSISTments problem sets, and other useful course information, course organizers were able to more easily decide what lectures to work with, and how to add them to the course. These tools, corrections, and content significantly improved the quality of CS1101’s online component. In the future, when a well established and audited database of lectures and questions is established, course staff should take great care to ensure the quality of video lectures and comprehension questions.
For the moment, a relatively well audited set of questions is sufficient for our purposes.

The results of the study largely showed remarkably insignificant or statistically unreliable, but interesting, patterns. The most surprising result is the lack of effect from comprehension questions on immediate quiz performance. While not statistically strong, there was evidence that comprehension questions have a deflating impact on self evaluations, possibly making students more realistic. While these studies were not initially within the scope of the project, these studies present multiple directions for further exploration. Given further time and resources, investigations into other factors, such as previous programming experience, or more features from the online homework data, may provide further interesting insights.
Bibliography