April 2012

Randomized Controlled Experiments Conducted With The ASSISTments Infrastructure

Jeffrey Alexander Namias
Worcester Polytechnic Institute

Randomized Controlled Experiments Conducted With The ASSISTments Infrastructure

A Major Qualifying Project Report
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science

by

_____
Jeffrey A. Namias

Date: April 26, 2012

Approved:

_____
Professor Neil T. Heffernan, Advisor

Abstract

ASSISTments, a web based educational tutoring system, has been used with thousands of students to provide real time feedback and targeted assistance in learning. In addition to helping students, ASSISTments serves as a research catalyst for academics by providing a highly integrated, minimally disruptive avenue for conducting educational studies. The work described herein leveraged ASSISTments to design, implement, and analyze studies which explored the impact of several treatment interventions on student learning. From November 2011 to April 2012, research questions were investigated pertaining to the following: the effect on student learning after re-mastery of prior knowledge skills and the effect of motivational video on student mindset and learning. Statistically significant learning gains were discovered in the former study, while the latter study suggested an effect and requires additional investigation.

Table of Contents

Acknowledgements

I would like to say thank you to everyone that contributed to these studies and made the research possible. This research was primarily supported by Neil Heffernan, a WPI professor and founder of ASSISTments. Additional collaborations with Kim Theinpont, Sidney D'Mello, Amber Chauncey Strain, and Cristina Heffernan were also integral to this research. I would also like to extend my thanks to all of the school teachers we worked with for accommodating our studies with their students and all the other individuals who collaborated on these projects. Thank you!

Introduction

ASSISTments is a non-profit research project that aims to benefit all parties that are involved in its use. First and foremost, students benefit from the immediate correctness feedback and from custom hints tailored to their current problem. Teachers enjoy the ability to leverage existing content or create their own in the robust content creation framework that ASSISTments provides. In addition to serving as a learning tool, teachers can access up-to-date reports on their student's performance to help assess their teaching. The flexibility of the ASSISTments team and versatility of the software itself allows for operational alignment with the goals of its funders and partners. Academics are provided a powerful instrument from which to conduct real world research, operating with minimal disruption to the thousands of students that naturally use ASSISTments for their classes. Everyone that benefits from the system also contributes positively simply by participating. This is the goal and vision of the ASSISTments, to foster a greater learning environment which benefits everyone involved.

ASSISTments, which is a web based educational tutoring platform, is freely accessible at http://www.assistments.org/ for public use. Use of the site requires a quick and easy registration which grants users access to the functional features of the system as well as an entire library of content, primarily created by WPI students and actively engaged teachers. With the support of federal grants, much of the existing math content has been certified, categorized, and catalogued by members of WPI into the WPI Fine Grained Skills, accessible via the following web address http://teacherwiki.assistment.org/wiki/WPI_Fine_Grained_Skills for public use. Those who wish to development new or additional content can make use of the

comprehensive build tools to modify an existing problem set or even begin from scratch. Once new content is created, all users are able to access and benefit from the additions.

This project primarily involved work on two studies completed in fulfillment of WPI's Major Qualifying Project (MQP) requirement for a Bachelor of Science degree. The work for the first study overlapped with parallel work completed for the Initial Qualifying Project (IQP), while the second study was completed separately; the two studies will be abbreviated Prior Knowledge and Video Motivation respectively hereafter. Work for the Prior Knowledge study, conducted as part of a team effort, focused on exploring the impact of re-mastery of prior knowledge skills on the retention of new mathematical content. This study posed the research question asking if students exposed to ASSISTments skill builders (homework assignments typically requiring three consecutively correct answers to complete) on topics relevant to the upcoming class curriculum would perform better than those not exposed to such skill builders. The Video Motivation study, which was completed as an individual effort, investigated the effect of motivational video targeting a student's appraisal with respect to two elements: their control over a situation, and the value they saw in the content of their coursework. Based on current psychological literature, particularly by Carol Dweck, this study posed the question asking if motivational video could target either of these two appraisals (control and value) to produce a measurable gain in a student's learning or mindset. Both studies were conducted within the ASSISTments infrastructure, working with 7^{th} and 8^{th} grade students to explore real world outcomes.

Study Design

Video Motivation Study – Trial 1

The Video Motivation study aimed to examine the effect on students exposed to motivational video targeting control appraisal and value appraisal. To explore this, the need for a carefully tailored script became evident. There were also design questions about who should recite the scripts in the video recordings. Logistically, it was necessary to provide assessment opportunities where impact could be observed. Finally, appropriate distribution of students into two groups was necessary.

The first design questions raised in the Video Motivation study pertained to the nature and circumstances of the videos themselves. There were questions about who might be most effective in the video – would a familiar teacher be a positive influence for students or perhaps might some students not relate to their teacher? Perhaps both would be true depending on the student and teacher. Would a "cool" celebrity or similar figure which the students admired be more effective? Ultimately, it was decided that a familiar teacher provided the best connection to students and was also the most accessible prospect.

Script content was based off of popular psychology literature, mainly inspired by psychologist Carol Dweck. With the guidance of collaborating psychology professors, it was determined that a video targeting control appraisal (an assessment of their own ability to affect what occurs in a situation) and value appraisal (an assessment of the value of, in this case, their class's math content) would be best suited for the stated research question.

In order to measure an effect, it was necessary that students both have a chance to grow after exposure to the video and that they have a chance to demonstrate the growth. In order to show how much growth had occurred, if any, the first item given would be an in class pretest establishing a student's prior knowledge. Following the pretest, a single growth opportunity was provided via skill builder homework to allow students a chance to increase their knowledge. It was in this homework that students, depending on their group placement discussed below, might be exposed to a video treatment. A posttest the next day as well as a second posttest a week afterwards would be used to assess if knowledge had been gained since the pretest and to assess if gains might be the result of a student's exposure to the video.

Given the psychological nature and focus of the video treatments, five survey questions were also administered with the pretest, homework, first posttest, and second posttest. These questions allowed the students to gauge from "I strongly disagree", "I disagree", "I neither agree nor disagree", "I agree", and "I strongly agree" how they felt about several value and control based statements. The questions were given as follows:

Table 1 – Trial 1 Survey Questions

Question#	Survey Question
1	With hard work most anyone can be good at math!
2	If I apply myself I can improve my grade in math.
3	I like learning about math.
4	Math skills taught in school are valuable.
5	I think that knowing the math skills taught in school will be good for me when I am older.

Finally, it was necessary that students be grouped appropriately for the study to be effective. Three conditions were identified for the study, two experimental conditions which would each receive a video in their homework and one control condition. These conditions were the value-appraisal condition which would receive the video targeting value appraisal, the control-appraisal condition which would receive the video targeting control appraisal, and the traditional control condition which would receive no video and would be used as a baseline group to measure the others against.

Students were sorted into a list based on summaries of their performance in previously completed classwork. The students then were divided into the three conditions, one by one in a chiastic "ABCCBA" pattern beginning at the top and moving down the sorted list. It should be noted that at the beginning of the homework, a student would place out of the study and not receive their video (if they belonged to either of the two conditions which would get a video in the first place) if they answered the first three homework questions correctly. By design, the fourth question was the video question and so students who already were masters of the content – demonstrated by answering first three homework questions correctly – would avoid their video entirely and would be dropped from analysis.

Video Motivation Study – Trial 2

Trial 2 of the Video Motivation study aimed to make slight modifications to the Trial 1 design in response to both student feedback and observations about shortcomings in the initial design. It was determined that based on a variety of circumstances, Trial 2 should fundamentally be similar to Trial 1 and serve as a retrial, as opposed to correcting all of the

observed flaws from the first trial in the second. This was done as there was interest in demonstrating a positive result before investing in a larger reengineering of the study. The following section speaks to what changes were and were not made and the reasons for these decisions.

Primary concerns voiced by the students revolved around the repeated use of the same survey questions so many times. In Trial 1, students were exposed to the same five survey questions four times in just over a week, three of those times in a period of two days and each time with the questions consecutively loading separate webpages. To reduce frustration, Trial 2 ran a longer survey with eight questions, adding three to the original five, but opted to only administer the survey twice, once at the start of the study and once at the end. Additionally, these survey questions all appeared in a single browser window loaded once to allow students to see the end of the survey and reduce time wasted on waiting for questions to load. The following questions in Table 2 were added to those appearing above in Table 1. These questions made use of reverse scoring so that detection of lying in answers would be possible.

Table 2 – Trial 2 Additional Survey Questions (See Table 1 for questions 1-5)

Question#	Survey Question
6	If I don't understand the course material, it is because I didn't try hard enough.
7	Some people just cannot be good at math no matter what they do.
8	I do not enjoy learning about math.

Several concerns were raised outside of student feedback. One such concern regarded the amount of class time required for the study. While Trial 1 required three tests (one pretest

and two posttests) all with knowledge and survey questions to be administered during class. Trial 2 was refined so that instead of Trial 1's three surveys and tests given in class, along with a fourth survey given at home, only two tests and one survey occurred in class. Instead of using an in class pretest, Trial 2 used a six item at home pretest before the homework. For a student to place out of the study, they must correctly answer all six of the pretest items (items in test mode provide the student no feedback); this requires double the number of consecutive correct answers to place out of the study compared to Trial 1, and contrary to Trial 1 students have no feedback on the place-out component in Trial 2. Another concern that affected the population of students chosen for Trial 2 was the low sampling size of Trial 1. Trial 2 intended to use over 100 students as opposed to under twenty-five usable from Trial 1. One concern which was not addressed in Trial 2 was the element of time. The two videos used each lasted just under a minute. In both Trial 1 and Trial 2, students placed into one of the two video groups were exposed to the video and immediately expected to have incurred a detectable effect, with assessment of this effect occurring within twenty four hours and again one week later. It was proposed that such a short time period and the lack of a continuous, repetitive treatment may have negatively contributed to Trial 1, but project collaborators wanted to see some effect with a single treatment before redesigning the study for expanded treatment time.

Prior Knowledge Study

(Note: The following section focuses on work completed as part of the IQP and not MQP. As such, this following section is provided for contextual completeness but not a representation of individual work completed towards the MQP.)

The Prior Knowledge study aimed to detect if exposure to relevant prior knowledge skills would improve performance on assessment of upcoming curriculum content. Scientific control highlights the importance of having two substantially equivalent groups, with one exposed to treatment and the other not, in order for sound comparisons and ultimately valid conclusions to be drawn. Since the research question involved a strict ordering of the homework, instruction, and assessment activity, the need to define a schedule was also evident. As such, design primarily involved student grouping and the timing, size, and scope of skill builders and assessment.

Students were divided into one of two groups. The first group, the experimental group, was the group which would receive treatment (i.e. exposure to skill builders on relevant prior knowledge skills). The second group, the control group, would similarly be exposed to skill builders, but not ones which were relevant to the upcoming curriculum content. This kept constant that both groups would be doing skill builder homework, but varied whether the content was relevant (experimental group) or irrelevant (control group). A list of students was sorted based on summary performance metrics from past homework assignments. Then, students on the list were divided in alternating fashion into the two groups, beginning with the top student and working down the list. Counterbalance was introduced to the design to counteract any imbalances in academic strength that one group might possess over the other. The resulting design was that each teacher would run the study twice, swapping each student's condition between the two runs. It was also deemed necessary to swap the book used between each run so that the knowledge gained in the first run would not affect the second.

Implementation

The first step for exploring the research questions raised in each study was to create appropriate content, built using the ASSISTments suite of content creation tools. My primary content creation contribution used in both the Prior Knowledge study as well as the Video Motivation study focused on fractions, specifically on the addition and subtraction of both proper fractions and mixed numbers, requiring the creation of sixteen variabalized templates. Variabalized templates are a feature of ASSISTments whereby a template with variables can be used to generate individual, static homework problems containing fixed values in place of those variables. With ASSISTments' ability to parse and compute logical statements entered into the content creation form, what was first forecasted to require over 200 templates was able to be accomplished in just sixteen. The example snippet of code shown below in Figure 1 was written for the subtraction of mixed numbers template, one of the sixteen templates mentioned above.

While Figure 1 shows code from the variabalized template, Figure 2 depicts hints that are part of a static ASSISTment question generated from the template featured in Figure 1.

Figure 1 – Sample Code From Variabalized Template

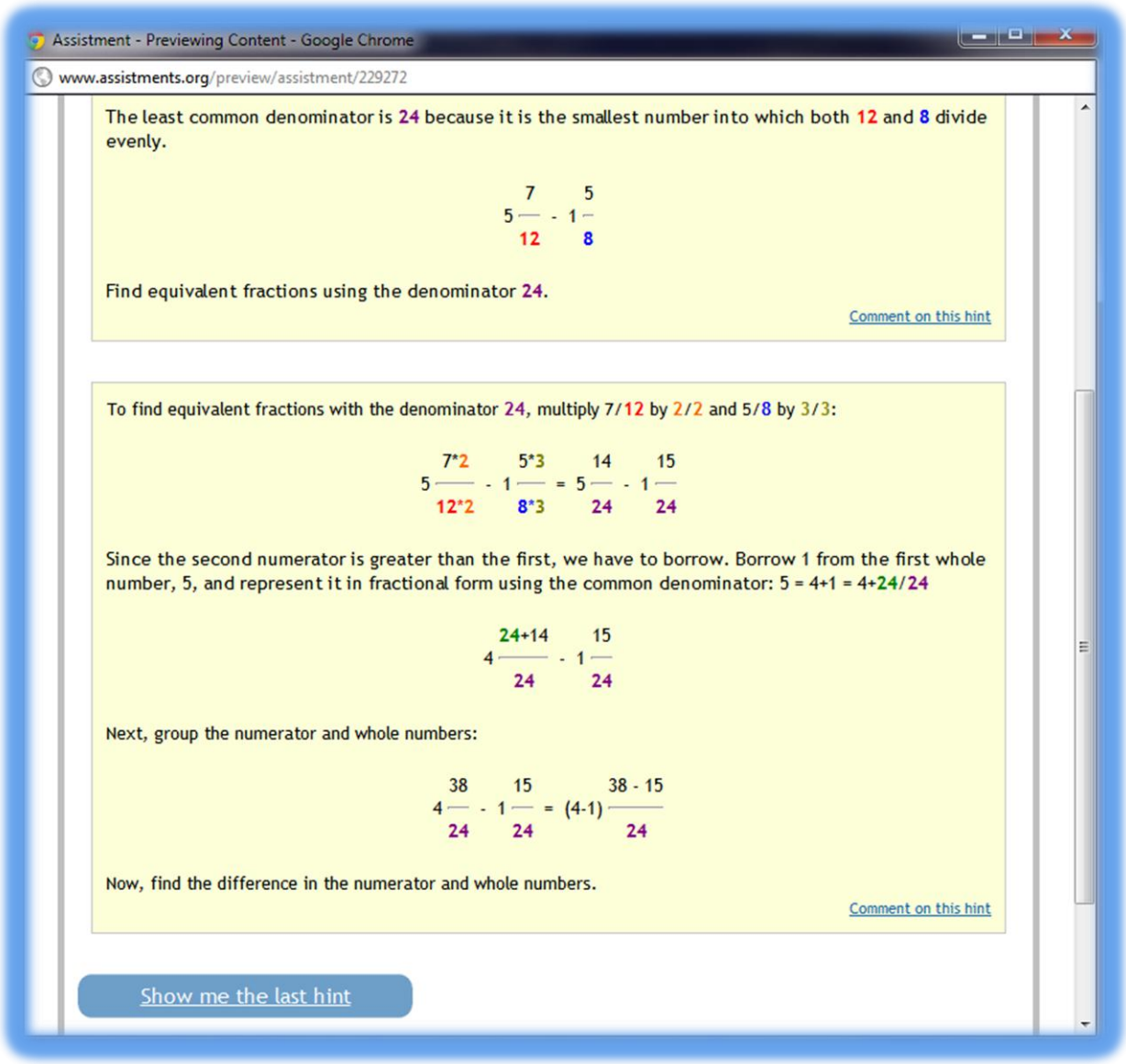

Figure 2 – Example of an ASSISTment hint, generated by code in Figure 1

The Prior Knowledge study involved a single design which was implemented multiple times with students from four different teachers. Four different mathematical text books were

used which outlined the curriculums used in these teachers' classes. Teachers were consulted to confirm the content of their planned curriculum, and then the correct skill builders, pretest, midtest, and posttests were prepared as per the design specifications.

Trial 1 of the Video Motivation study was run a single time with a single class of twenty-six students. Upon feedback, the design changes outlined for Trial 2 were implemented, and Trial 2 of the study was run with a different teacher. While over 100 students across this teacher's sections were assigned the Trial 2 homework and two posttests, fewer than thirty-five students were able to accommodate completing all three of the assignments.

Results

Prior Knowledge Study

A statistically significant learning gain was discovered in students working with Accentuate the Negative, one of the four books used in this study. With this book, students in the experimental group (who were assigned skill builders based on relevant prior knowledge skills) that completed tests and performed above average on the pretest gained 28% measuring from the pretest on curriculum content to the posttest on that content. In comparison, the control group (who were assigned irrelevant skill builders) gained only 18% between pretest and posttest on curriculum content. The p-value computed via two-tailed homoscedastic ttest showed statistical significance at 0.024.

Later, a statistically significant gain was shown in analysis by Kevin Dietz to apply more generally to the students participating in the Accentuate the Negative study than described

above. Kevin found that students in the experimental group had gain scores above those of the control group for students which completed the pretest, midtest, and posttest. Kevin's result was statistically significant and did not require filtering the data based on initial pretest score, which as described above was necessary to show an effect in our initial analysis. Kevin was able to broaden our initial findings, which first applied only to students completing the tests that also performed above average on the pretest, to then simply apply to all students that completed the tests. Kevin described "if we run the analysis as an ANCOVA, with posttest score as our dependent variable, condition as our independent variable, and pretest scores as a covariate, there is a significant difference between the two conditions at post-test, $F(1, 116) = 5.38$, $p < .05$" and these results are visible in Figure 3 shown below.

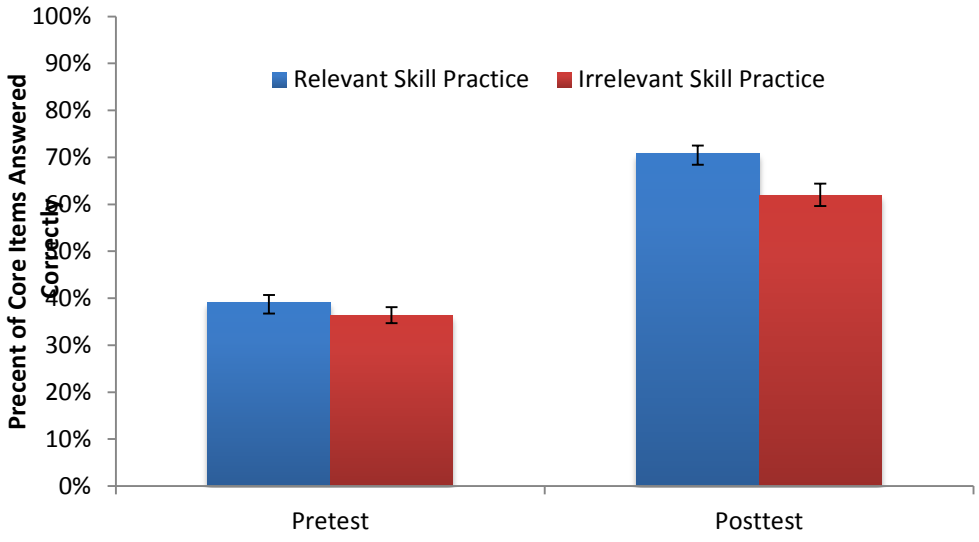

Figure 3 – Graphical Representation of ANCOVA Analysis for Accentuate the Negative Students Who Completed All Tests

Video Motivation Study – Trial 1

Trial 1 of the Video Motivation study was less effective than we had hoped. No knowledge gain was apparent in either of the video groups over the control group. These results were understandable considering the limited scope of time students were exposed to treatment, the small sample size, and the fact that of the twenty-three students who had completed enough work to be analyzed, an additional five of those students placed out of the video groups as shown in Figure 4. By chance, no students in the no-video group completed the first three homework problems correctly and thus no students in this group placed out of the study.

Condition	Students	Placed Out?
Value-Appraisal	7	2
Control-Appraisal	8	3
No Video	8	0
All	23	5

Figure 4 – Video Motivation Study Trial 1 Place Out Data

An effect was shown in the survey questions for the value-appraisal group as compared to the control-appraisal and no video groups for questions relating to value. This was an understandable result as the brief treatment period was seen as an appropriate timeframe for beginning to convince a student that external subject matter is valuable and worthwhile. Conversely, it was interpreted that to shift a student's perspective of their own capabilities and control was not a task that could be expected to successfully occur with a single treatment in under a minute. The following analysis in Figure 5 shows the value result mentioned above and

was produced by University of Memphis professor Sidney D'Mello using the SPSS software package.

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
Control Appraisal	No Video	.004	.242	.986	-.508	.517
	Value Appraisal	-.585(*)	.235	.024	-1.083	-.087
No Video	Control Appraisal	-.004	.242	.986	-.517	.508
	Value Appraisal	-.589(*)	.250	.032	-1.119	-.059
Value Appraisal	**Control Appraisal**	**.585(*)**	**.235**	**.024**	.087	1.083
	No Video	**.589(*)**	**.250**	**.032**	.059	1.119

Figure 5 – Pairwise Comparison of Control-Appraisal, No Video, and Value-Appraisal Conditions

Video Motivation Study – Trial 2

Trial 2 produced a desirable result, but based on initial analysis does not appear to be statistically significant. As was carried out in the other studies, the initial analysis here consisted of reducing the data down to students who had completed all of the testing metrics. This produced the gain score results depicted in Figure 6 below.

	Posttest1- Pretest	Posttest2- Pretest	Posttest2- Posttest1
No Video	12.89%	3.67%	-9.22%
Value Appraisal	15.09%	13.64%	-1.45%
Control Appraisal	18.20%	18.20%	0.00%

The No Video group has the lowest gain score for all categories, particularly on the posttest2 to pretest comparison which shows a sub 4% gain compared to over 3 and 4 times that respectively. It must be noted however that these results, based on the current analysis,

do not seem to be statistically significant. Continuing collaboration is underway to assist in analyzing this data further.

Reflection and Future

Several pertinent questions have been raised with respect to the previously discussed studies and could help to guide future progress if pursued. One of these questions involved the trigger for introducing treatment conditions to students. Some of the literature suggests that allowing a student to struggle, and then rewarding their successful persistence could be more effective than targeting failure. By the current design used in Trial 1 and Trial 2 of the Video Motivation study, students are targeted upon failure. In both trials, a student must have perfect performance on the beginning homework items to place out of the study. Otherwise, their failure is addressed and they are encouraged with value-appraisal video, control-appraisal video, or receive no video encouragement depending on the student's assigned group. If we targeted the same students, but waited until they successfully completed the assignment to intervene with a treatment, treatments may be more effective.

Two other questions revolved around timeframe and student grouping. The general consensus among collaborators on this project was that the treatment schedule (1 intervention, under 1 minute long) was absolutely too short for control-appraisal and that value-appraisal could also benefit from a longer schedule of treatment application. Another way to improve the effect, which may be the most promising, would be to make use of an initial survey to assess if each student even struggles with any control obstacles or value issues in the first place. If so,

could a stronger impact be identified by pre-matching students into relevant appraisal conditions? Doing so may yield strong results and should be investigated moving forward.

Conclusion

Work for this project focused primarily on two studies, the first of which examined the impact of re-mastery of prior knowledge skills on performance in upcoming curriculum work, and the second which involved designing, implementing, and analyzing data on the effect of motivational video in student learning. The results of the first study indicated a statistically significant result for knowledge gain as predicted. The second study showed a desirable result in Trial 2, but failed to be statistically significant. It is suggested that with further refinement as suggested in the Relfections and Future section, that further experimentation could be productive. The work discussed in this project was completed towards the fulfillment of the Major Qualifying Project requirement for Bachelor of Science degree at WPI by leveraging the ASSISTments platform. Work on both these studies is an ongoing effort involving the contributions and cooperation of a large network of individuals.

References

1. http://www.assistments.org/
2. http://teacherwiki.assistment.org/wiki/WPI_Fine_Grained_Skills
3. http://news.stanford.edu/news/2007/february7/dweck-020707.html