

# Does Instant Feedback on Online Homework Assignments Improve Student Learning in Introductory Economics Classes?

Veronika Dolar, SUNY Old Westbury

## ABSTRACT

*The purpose of this paper is to study the effect of receiving instant feedback on online homework assignments. Using data from a natural experiment that included over 500 students taking Principles of Micro- and Macroeconomics at a midsize public university in Ohio, I show that “Grade It Now” (GIN) option in Aplia -an online learning management system positively impacts grades on assignments. This impact is especially strong for academically weaker students and has the same impact on students’ grade as does increasing GPA by almost half a point. However, in sections with GIN, students’ performance on midterm exams and final exam was either not statistically different from sections with Grade At Deadline (GAD) option or was actually worse. Using OLS regression and controlling for various student and class characteristics, I show that Aplia’s GIN impact on students’ performance on exams is negative and does not improve student learning. One possible explanation for this might be due to students’ trying to “game” the system by increasing their grades on assignments and lowering their efforts on exams. This behavior seems to be supported with the data since there is no difference in the final grade between sections using GIN vs. GAD.*

**JEL Classification:** A20, A22, I21

**Keywords:** Economic education, Learning technology, Online assessment, Aplia, Multiple attempts, Grade at deadline, Grade it now.

## INTRODUCTION

In the past few years, the use of online assessment tools, such as Aplia and MyEconLab, has been rapidly increasing. Alongside this increase has been the publication of articles examining the effectiveness of these tools. The goal of this paper is to add some new insights to this burgeoning literature.

In this paper I study the effect of Grade It Now (GIN) option in Aplia on student learning. I use a data set from a natural experiment that includes over 500 students taking Principles of Micro and Macroeconomics at a midsize public university in Ohio. About half of the students used the older version of Aplia where they had only one set of questions to complete online and had to wait until the deadline to receive feedback on their work. In this paper I refer to this option as Grade At Deadline (GAD). The other half of students used the newer version of Aplia with Grade It Now (GIN) option, where they were able to obtain immediate feedback on their work for each question on the assignment. In addition, under GIN students were allowed two additional attempts for each question that were not identical, but very similar to the original question. The main intention of GIN is to allow students to learn from their mistakes right away,

instead of having to wait for help from the instructor or wait to see the correct answers later online.

My results show that Aplia's GIN positively impacts grades on Aplia assignments. This impact is especially strong for academically weaker students and has the same impact on students' grade as does increasing GPA by almost half a point. However, in sections with GIN option, students' performance on midterm exams and final exam was either not statistically different from sections with GAD option or was actually worse. Using OLS regression and controlling for various student and class characteristics, I show that Aplia's GIN impact on students' performance on exams is negative and does not improve student learning. One possible explanation for this might be due to students' trying to "game" the system by increasing their grades on assignments and lowering their efforts on exams. Since there is no difference in the final grade between sections using GIN vs. GAD, this behavior seems to be supported with the data.

The remainder of the paper is organized as follows. In Section 2, I begin with a description of Aplia and briefly review the literature on the impact of online learning tools on the overall student success. In Section 3, I describe the data used for this study and provide descriptive statistics of some of the key variables. In Section 4, I report my results, first by analyzing the impact of GIN on Aplia assignments grade and second by analyzing the impact of GIN on other grades. Finally, I offer concluding remarks in Section 5.

## BACKGROUND

Aplia is one of many online learning management systems available on the market today. It was developed by an economist Paul Romer in 2000 and is now owned by Cengage Learning.<sup>1</sup> Even though Aplia started as a tool for economics courses, today Aplia is available for use with more than 200 textbooks across 21 disciplines including business communication, economics, finance, and statistics (Cengage Learning, 2013). The program comes with tutorials, homework assignments, interactive activities, experiments, news analysis, and reports of the students' progress as well as online versions of the textbook that is being used in the class.

One of the most important benefits from using Aplia in economics is its ability to ask not only complex numerical questions but also questions that require the use of graphs. In Aplia, students are asked to derive curves, highlight areas on the graph, and manipulate graphs by shifting curves; all of which is also automatically graded. Since all provided questions are electronically graded it can save a great amount of grading time. In addition, Aplia helps instructors by giving them options on how to set up the assignments, how they should be graded, and when the students can expect to get feedback to their problems.

In the original version of Aplia students would get a detailed feedback that included a step-by-step explanation of the problems after the deadline of their assignments; a method called Grade at Deadline (GAD). In the fall of 2008 Aplia introduced a new tool called Grade it Now (GIN). This new option allows students to get immediate feedback on their work for each specific problem. GIN also allows students to try to answer additional attempts up to three times. These additional attempts are almost identical to the original question but use alternative numbers and examples. In other words, after answering a question and obtaining the feedback, students may decide to either move on to the next question, or try another version of the question they have just attempted. Finally, in an attempt to discourage cheating, Aplia randomizes the order of questions in each attempt for every student.

For grading purpose the instructor can choose from three different settings on how to score additional attempts. The first setting is called “Average”, where the average score based on all attempts is taken. Another setting, “Do No Harm” includes only the score in the averaging process if it does not lower the current average. Finally, the last setting is called “Keep the Highest” and it takes the highest score out of all attempts. The default option in Aplia is the “Average” option, recommended by Aplia since “Do No Harm” and “Keep the Highest” might allow students to use their first attempt to look at the explanations.

In the past few years, a number of papers have examined the effectiveness of Aplia and other online assignment systems both in economics and other fields and report mixed results (Bianco, Lindblom, & Nixon (2014), Richards-Babb Drelick, Henry, & Robertson-Honecker (2011), Bonham, Beichner & Deardorff (2006), Ball, Eckel & Rojas (2006), Butler and Zerr (2005)). For example, Bonham, Beichner & Deardorff (2006) studied computer graded homework versus human-graded homework in a large introductory physics course and found no significant difference in the performance of students doing either type of assignments. On the other hand, in an introduction in chemistry course, Richards-Babb et al. (2011) found that when in class laboratory quizzes are replaced with graded online homework there is a significant positive relationship with student performance. In addition, a study using more than 750 students and 35 instructors from 30 two- and four-year institutions of higher education in developmental English class found that students’ learning increased dramatically when using Aplia as a learning tool. Both reading and writing skills were improved and students reported that Aplia helped them prepare better for tests (85%), that the use of Aplia allowed them to keep track off their progress in the course (85%) and that it was a valuable tool in helping learn new concepts (85%) (Cengage Learning (2013)).

In economics, Lee, Courtney & Balassi (2010), using unpaired t-tests, find that there is no statistically significant difference in improvement in the Test of Understanding in College Economics (TUCE) between students using traditional instructor-assigned and graded homework and online Aplia assignments (either GAD or GIN versions). However, using OLS regression they find that students who received A and B grades and were using Aplia’s GIN option improved their TUCE scores by nearly two points over those students who used instructor-assigned and -graded homework assignments.

Similarly, Kennelly, Considine & Flannery (2011) compare the effectiveness of online (Aplia) and paper based assignments using students in one large managerial economics course in Ireland. Their results show that the format of an assignment makes no difference in how a student performs on an exam. In a follow up study, Flannery, Kennelly & Considine (2013), using panel data, find that paper assignments were generally more effective than online assignments in preparing students to answer exam questions.

Using one undergraduate level of principle of macroeconomic, with a sample size of 129, Self (2013) finds that doing well on online homework assignments does not impact test grades. However, students that voluntarily access the website to practice on additional problems are found to do better on tests.

Finally, Rhodes & Sarbaum (2013) study the impact of online homework management system, when multiple attempts on assignments are allowed. By using data from two introductory macroeconomics classes in two successive summer sessions, students are given two attempts in the first session and one attempt in the second session. Most of the questions are multiple choice questions (MCQ) with 4 or 5 options and the only feedback students receive are their total scores and the indication of which questions they missed. Given these settings, and

without controlling for any additional student characteristics, they find that multiple attempts lead to “gaming” behavior that results in grade inflation without improvement in learning outcomes.

A unique feature of this paper is that the sample size is significantly larger compared to most studies mentioned above. In addition, I am able to control for numerous individual and class characteristics and the class sections used in this study are more diverse in size. In addition, I also study the effect of allowing multiple attempts on assignments, however the “gaming” behavior (simply adjusting your guesses on each question) is not as easy since most of the questions in Aplia are not MCQ and require a numerical answer (with fill in the blanks option) or direct work with the graphs.

## DATA AND DESCRIPTIVES

The experiment for this project was conducted over six semesters and data derived from twelve sections of principles of micro- and macroeconomics classes (seven and five sections, respectively) during the Spring, Summer, and Fall semester of 2008 and 2009. All the courses were taught by the same professor in the Economics Department at Cleveland State University in Ohio. The professor taught each of the twelve sections as similarly as possible using the same textbook, covering the same material, and giving similar exams. The only planned difference in the courses was the type of the homework assigned (GAD vs. GIN).

The sample includes 504 students.<sup>2</sup> There were 286 students using the GIN version of Aplia and 218 students using the GAD version. As shown in Table 1 the size of the class sections varied from 12 to 80 students, with the average class size of 56.8 students (std. dev. 18.5). In addition, most sections were taught in mid- to late morning and met three times per week (Mondays, Wednesdays, and Fridays). Three sections however, were offered as a late afternoon/evening class, which during the summer session met twice a week and once a week during the spring 2009 semester.

Semester	Year	Morning	Grade it Now	No. Students	Class
Spring	2008	Yes	No	59	Micro
Spring	2008	Yes	No	72	Macro
Summer	2008	No	No	12	Micro
Summer	2008	No	No	21	Macro
Fall	2008	Yes	Yes	59	Micro
Fall	2008	Yes	No	54	Macro
Spring	2009	Yes	Yes	55	Micro
Spring	2009	No	Yes	30	Micro
Summer	2009	Yes	Yes	14	Micro
Summer	2009	Yes	Yes	17	Macro
Fall	2009	Yes	Yes	47	Micro
Fall	2009	Yes	Yes	64	Macro
		TOTAL		504	

As shown in Table 2, almost sixty percent of students were male, a majority of students were white (68.2%) and the biggest proportion of students were in their second year in college. The average age of a student was 22.7 years with the average GPA of 2.8 (see Table 3). The final grade was based on student's performance on assignments and exams. More precisely, the final grade was a weighted average with Aplia homework assignments worth 30%, midterm exam grade worth 30%, and the final exam worth 40%. In the fall and spring semesters the final exam was cumulative (with 70 multiple choice questions), while in the two summer sessions (50 multiple choice questions) it was not. During the summer session, students only took one midterm exam (50 multiple choice questions), while during the fall and spring semester they were given two midterm exams (30 multiple choice questions each), but only the highest of the two midterm exams counted towards their final grade. In addition, students were able to earn up to 3 additional percent added to their final grade (extra credit) based on their performance on the math assignment (math review) that was offered on Aplia in the first three weeks of the semester. Using this grading rule, and expressing all the grades in percentage terms (normalizing to 100) the average on homework assignments for all 504 students was 80.6% - this average was based on all assignments assigned in each class after the two lowest scores were dropped. In addition, the professor used the "Average" setting in Aplia, so that student's score on any question on the homework assignment was based on the average of all the attempts taken (see Section 2 for a more detail description of the "Average" setting in Aplia). The midterm exam average was 79.4%. Finally, the average final exam grade was 69.5% and the final grade was 79.0%, which is equivalent to a letter grade C+.

Variable		Number	Percentage
Sex	Male	301	59.7
	Female	203	40.3
Race	White	347	68.8
	Black	102	20.2
	Asian	48	9.5
	Hispanic	7	1.4
Year in College	Freshman	80	15.9
	Sophomore	211	41.9
	Junior	127	25.2
	Senior	86	17.1

Variable	Average	Median
Age	22.7 (5.56)	20.8
GPA	2.8 (0.72)	2.8
Homework	80.6 (14.41)	84.3
Midterm Exam	79.4 (14.17)	80.0
Final Exam	69.5 (15.64)	70.5
Final Grade	79.0 (13.26)	80.7

## RESULTS

To recap, my main interest in the empirical analysis is to discover whether Grade It Now (GIN) in Aplia affects student's performance on assignments and exams. In the first subsection below I analyze the impact of GIN on Aplia assignments and in the second subsection I examine the impact of GIN on midterm and final exams, as well as the final grade.

### GIN and Aplia Assignments:

I begin by performing a series of two-sample mean comparison t-tests by individual assignments.<sup>3</sup> Aplia assignments offered in GAD version were extremely similar (if not identical) to those in GIN version. However, the assignments in micro- and macro-economics classes were not the same. Moreover, in the introduction to microeconomics 13 assignments were assigned while in the introduction to macroeconomics there were 11 assignments. As a result, I separate the data into principles of macro- and micro-sections. You can review the list of all the assignments by general topic in the Appendix. Since at this university, the Introduction to Macroeconomics (ECO 201) is typically taught before Introduction to Microeconomics (ECO 202) I start our analysis with Introduction to Macroeconomics.<sup>4</sup> In Table 4 we can see that in Principles of Macroeconomics the average scores on Aplia assignments with GIN were higher for all but one homework assignment. For homework assignment 2 (HW2) the score with GIN is lower than with GAD, however, this difference is not statistically different. In addition, the scores for assignments 1 and 7 are higher under GIN but the differences between the two means are not statically different. It should be pointed out that assignment 1 was a very basic assignment and did not include any knowledge of economics as it was an introduction on how to use Aplia and how to complete assignments online. This could potentially explain why there is no difference between the two types of assignments. Finally, the average assignment scores (which does not included two lowest grades on assignments) is almost 10% higher with GIN compared to GAD and this difference is statistically significant at  $p < 0.01$ .

Assignment	Grade it Now (GIN)	Grade at Deadline (GAD)
HW1	95.6 (11.48)	91.7 (22.89)
HW2	74.9 (24.54)	80.0 (24.68)
HW3*	72.7 (33.32)	65.2 (25.90)
HW4***	83.6 (14.92)	67.5 (26.81)
HW5***	77.0 (25.64)	63.1 (28.17)
HW6***	79.6 (26.34)	69.6 (28.21)
HW7	80.3 (22.72)	75.1 (27.17)
HW8**	77.1 (25.05)	69.3 (26.28)
HW9**	69.2 (28.81)	61.2 (21.64)
HW10***	69.9 (27.30)	52.6 (30.25)
HW11***	70.8 (36.62)	58.0 (34.64)
HW average***	84.2 (13.28)	75.6 (16.11)

Statistical difference of the means \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

I perform the same type of t-tests for average scores on Aplia assignments in microeconomics sections and obtain similar results. The scores on assignments with GIN are higher than under GAD for 9 out of 13 assignments (all statistically significant). For 3 other assignments, scores under GAD are higher compared to GIN, however this difference is not statistically significant. The only assignment, where the scores under GAD are higher and statistically different compared to GIN is assignment 2. One possible explanation for this might be due to learning-by-doing; students are still experimenting and trying out how GIN works. For example, one property of GIN (as selected by this instructor) is that it takes the average score based on all attempts. This means, that if a student selects and clicks for a second or third attempt, but does not actually solve any problems and does not supply any answers, Aplia considers those missing answers as incorrect at deadline. As a result the average grade is lowered. Another explanation might be that students were still not taking full advantage of GIN by taking second and third attempts. Finally, the average score based on all assignments (minus two lowest scores) for GIN is higher compared to GAD (82.7% vs. 80.7%), however, this difference is not statistically significant.

<b>Assignment</b>	<b>Grade it Now (GIN)</b>	<b>Grade at Deadline (GAD)</b>
HW1	91.9 (20.09)	94.8 (17.06)
HW2***	73.2 (23.03)	84.5 (19.86)
HW3	73.2 (31.17)	66.8 (31.47)
HW4**	78.6 (20.77)	70.6 (28.68)
HW5	68.7 (27.04)	70.7 (20.78)
HW6***	83.7 (18.98)	64.7 (27.63)
HW7***	80.7 (25.48)	62.9 (30.49)
HW8***	79.2 (24.42)	67.2 (34.02)
HW9	75.9 (33.61)	77.2 (27.79)
HW10***	79.8 (21.04)	66.6 (27.10)
HW11***	70.2 (31.73)	53.1 (34.63)
HW12*	73.3 (35.77)	65.2 (33.22)
HW13**	48.6 (31.32)	39.41 (32.71)
HW average	82.7 (12.63)	80.7 (14.29)

Statistical difference of the means \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The mean comparison t-tests suggest that GIN does in fact positively impact student performance on the assignments. I now wish to estimate the magnitude of using GIN versus GAD. I do so by using ordinary least squares (OLS) to estimate a multiple regression in which the dependent variable is the average grade on Aplia assignments (average over all assignments, ignoring the two lowest scores, per student):

$$HW_i = \alpha + \delta GIN_i + \beta X_i + \epsilon_i$$

The model is estimated by controlling for various student and class characteristics such as sex, race, age, GPA, year in college, class size, captured in vector  $X$  and an indicator  $GIN$  which equals one if the assignments were GIN and zero if assignments were GAD. The parameter  $\beta$  is conformable to  $X$  and  $\epsilon$  is the error term. The parameter of interest is of course  $\delta$ , which I am expecting to be positive; the average score on the assignment will be higher when the indicator is turned on - meaning the assignments used GIN rather than GAD.

The results for this regression are reported in Table 6. In the first regression we controlled for numerous student and class characteristics. The OLS estimate for the delta (GIN) coefficient is 4.20 and is statistically significant. This means that the assignment grade is 4.2 percentage points higher when Aplia assignments are using GIN as opposed to GAD version. Many coefficients for control variables are not statically significant (age, race, major, and whether or not this was a morning or an evening class).<sup>5</sup> However, few others are and have expected signs. Class size is statistically significant with a negative coefficient equal to -0.13, which means that for every additional student added to the class, the homework average decreases by 0.13 percentage points.<sup>6</sup> Females are predicted to receive a homework grade that is 3 percentage points lower compared to males, and each additional point in GPA is estimated to increase homework grade by 12 percentage points. Finally, each additional year in school (going from freshmen to sophomore to junior to senior) is estimated to decrease homework grade by 1.5 percentage points, a somewhat surprising result. One possible explanation for this might be because the best students, after they realize their potential, transfer to other universities or that more senior students have other commitments (work, family, etc.) that keeps them away from school work.

In the second regression I drop the statistically insignificant variables and reestimate the regression and obtain similar results. In the second regression we see that the importance of GIN is slightly increased while gender biased decreased. Overall, this results suggest that GIN option in Aplia has a positive impact on assignment grades and has the same magnitude as increasing students' GPA by 0.36 points.

Finally, I further refine my analysis by running three regressions conditional on student's academic achievement. In regression 3, I restrict the sample to include only those students that received a final grade in this class of B or higher. In this case, the delta (GIN) coefficient becomes much smaller and statistically insignificant. This result is quite different from that reported by Lee et al. (2010) where they show that students who received A and B grades and were assigned GIN homework improved their scores by nearly two points over those students that used traditional instructor-assigned and -graded homework.

Another observation of interested in regression 3 (and 4) is that for high achieving students the class size, sex, and year in college also do not matter for their homework assignment grade.

In the fourth regression I restrict our sample to include only those students that had a cumulative GPA at the end of the semester taking this class 3 or higher. Similarly to regression 3, the delta (GIN) coefficient is smaller and statistically insignificant compared to regression 2. In the fifth and final regression I look at the students that had GPA lower than 3. It is here, that the impact of using GIN is the strongest. The delta (GIN) coefficient is now 5.8 and statistically significant. This result suggests that students that stand to benefit the most from having the access to the GIN version of Aplia assignments are academically weaker ones. For students with lower GPA the impact of using GIN assignments is equivalent to increasing student's GPA by almost half a point.



	(1) HW	(2) HW	(3) HW	(4) HW	(5) HW
GIN (0-GAD, 1-GIN)	4.195*** (4.13)	4.316*** (4.28)	0.955 (0.99)	1.626 (1.49)	5.770*** (3.97)
Class Size	-0.131*** (-4.13)	-0.124*** (-4.63)	-0.0225 (-0.93)	-0.00132 (-0.05)	-0.198*** (-5.05)
Sex (0-M, 1-F)	-3.067*** (-2.97)	-2.833*** (-2.83)	-0.591 (-0.59)	0.773 (-0.68)	-4.463*** (-3.11)
GPA	12.16*** (17.36)	12.05*** (17.64)	5.973*** (6.12)	9.530*** (5.35)	12.25*** (9.09)
Year in College	-1.527*** (-2.70)	-1.313** (-2.54)	-0.0683 (-0.14)	-0.57 (-1.08)	2.011*** (-2.60)
Age	0.131 (1.34)				
Race	0.71 (0.99)				
Major	0.17 (0.73)				
Morning Section (0-Yes, 1-No)	-1.913 (-0.99)				
Constant	56.27*** (15.45)	59.18*** (23.07)	71.23*** (20.99)	57.55*** (8.83)	65.03*** (16.41)
Observations	504	504	242	192	312
Adjusted R <sup>2</sup>	0.424	0.424	0.127	0.139	0.317

t statistics in parentheses \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

### GIN and Other Grades:

The results above suggest that, everything else equal, Aplia GIN option is beneficial for students since it positively impacts their assignment grades. Since those scores can only be improved by reviewing submitted answers and redoing questions and as a result practicing with more problems, this should lead to improvement of student's learning and better understanding of the material. In this section, I analyze the effect of using GIN in Aplia on student's midterm exam, final exam, and final grade. Starting again with simple mean comparison t-tests reported in Table 7 we see that the average score on midterm exam, final exam, and the final grade in principles of macroeconomics is higher under GIN compared to GAD, however most of these differences are not statistically different. In principles of microeconomics all of these grades are higher under GAD rather than GIN.

As before I wish to estimate the impact of using GIN versus GAD, but this time using midterm exam grade, final exam grade, and final grade as the dependent variable by estimating the following three regressions:

$$MidtermExamGrade_i = \alpha + \delta GIN_i + \beta X_i + \epsilon_i \quad (1)$$

$$FinalExamGrade_i = \alpha + \delta GIN_i + \beta X_i + \epsilon_i \quad (2)$$

$$FinalGrade_i = \alpha + \delta GIN_i + \beta X_i + \epsilon_i \quad (3)$$

<b>Table 7</b>		
<b>AVERAGE SCORES (STANDARD DEVIATION IN PARENTHESIS)</b>		
Assignment	Grade it Now (GIN)	Grade at Deadline (GAD)
Principles of Macroeconomics Sections		
Midterm Exam	80.9 (13.33)	79.2 (13.82)
Final Exam	70.2 (14.81)	67.8 (16.90)
Final Grade *	80.7 (12.67)	77 (14.60)
Principles of Microeconomics Sections		
Midterm Exam**	77.9 (14.69)	82.8 (13.79)
Final Exam ***	68.7 (15.31)	74.3 (14.04)
Final Grade	79 (12.54)	81.1 (12.72)

Statistical difference of the means \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

As before the model is estimated by controlling for various student and class characteristics captured in vector  $X$  and an indicator  $GIN$  which equals one if the sections used assignments with GIN and zero if assignments were GAD. The parameter of interest is still  $\delta$ , which I again expect to be positive; the average score on the exams and the final grade will be higher when the indicator is turned on meaning the sections used GIN option on the Assignments. The results of these estimations are recorded in Table 8.

As reported in Table 8, the delta (GIN) coefficients are all statistically significant for midterm and final exam, however, the estimated coefficients are now negative. This implies that for the midterm exam grade in sections using GIN option in Aplia, the grade is 2.2 percentage points lower compared to sections where standard GAD assignments were used. Similarly, for the final exam grade, the grade is 2.4 percentage points lower in sections using GIN compared to final exam grade in sections with GAD assignments. Finally, the difference in the impact of GIN vs. GAD on the final grade is very small and statistically insignificant. In our empirical analysis I have also looked at subcategories based on student's academic achievement (results not reported here) but obtained similar results as the ones reported in Table 8.

These results suggest that Aplia's GIN impact on students' performance on exams is negative and does not improve student learning. As suggested by Rhodes and Sarbaum (2013) by providing instantaneous feedback and allowing multiple attempts on assignments students do not improve their learning, but rather learn how to "game" the system. Students are able to more easily improve their grades on assignments and as a result their effort on exams can now be lowered in order to achieve the same outcome - same grade. Since there is no difference in the final grade between sections using GIN vs. GAD, this behavior seems to be supported in the data. In fact, as reported in the subsection above, the biggest impact of Aplia's GIN option on assignment grades was for students that had GPA less than 3. This suggests, that students "gaming" the system the most are the ones with lower levels of GPA (regression 5 in Table 6) while the students with higher levels of GPA did not change their behavior under two different regimes (see regression 4 in Table 6).

It is possible, however, that students did in fact learn more when using GIN option in Aplia. As mentioned, the exams taken by students in each section were similar, but not identical. It is possible that the instructor unconsciously (or endogenously) selected relatively more difficult questions for exams in sections with GIN. The instructor tried to choose questions on

the same topics, but the exact question or the type and difficulty of the questions varied from one exam to the next.

	(1) Midterm	(2) Midterm	(3) Final Exam	(4) Final Exam	(5) Final Grade	(6) Final Grade
GIN (0-GAD, 1-GIN)	-2.079** (-2.04)	-2.197** (-2.22)	-2.414** (-2.21)	-2.224** (-2.06)	-0.336 (-0.41)	-0.306 (-0.38)
Class	0.0091 (0.28)	-4.693*** (-4.70)	-0.108*** (-3.16)	-0.0877*** (-3.06)	-0.0979*** (-3.83)	-0.0963*** (-4.49)
Sex (0-M, 1-F)	-4.546*** (-4.38)	12.30*** (17.95)	-5.224*** (-4.71)	-5.108*** (-4.76)	-4.398*** (-5.29)	-4.328*** (-5.40)
GPA	12.15*** (17.25)	-1.158** (-2.23)	14.00*** (18.62)	13.98*** (19.23)	13.52*** (24.00)	13.52*** (24.68)
Year in College	-1.100* (-1.93)	48.09*** (23.07)	-0.314 (-0.52)	36.94*** (13.71)	-1.116** (-2.45)	-0.999** (-2.41)
Age	-0.0404 (-0.41)	504 0.399	0.148 (1.41)	504 0.437	0.0763 (0.97)	50.11*** (24.34)
Race	-0.674 (-0.94)		0.292 (0.38)		0.152 (0.26)	504 0.562
Major	0.164 (0.70)		0.184 (0.73)		0.179 (0.95)	
Morning Section (0-Yes, 1-No)	2.666 (1.37)		-3.454* (-1.67)		-0.896 (-0.58)	
Constant	48.59*** (13.27)		35.01*** (8.96)		48.40*** (16.53)	
Observations	504		504		504	
Adjusted R <sup>2</sup>	0.398		0.437		0.561	

t statistics in parentheses \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

It is possible that in sections with GIN students were better prepared for class and were able to follow the material more easily. As a result, the instructor covered more complex material. Sensing that students in GIN sections will be able to answer more difficult questions on the exams she increased the difficulty of the exams by choosing more difficult questions. Under this scenario the actual exam grades were not changed (or were slightly lowered as suggested by the data) but the knowledge or the complexity of the knowledge possessed by the students in GIN section was higher.

In order to test this hypothesis, I look at the exams scores adjusted for their difficulty. All the multiple choice questions (MCQ) on the midterm and final exams come from the same test bank that accompanies the textbook, and where all the MCQ are classified according to their difficulty on a 3 point scale (1=easy, 2=intermediate, and 3=difficult). I calculate the adjusted score by multiplying the total number of easy questions on each exam by one, the total number of intermediate questions by two, and the total number of difficult question by three. Further, I divide that sum by the total number of questions on the exam and normalize to 100. This means, that the lower bound for the adjusted scores is 100 (all the questions on the exam are classified as easy) and the upper bound for the adjusted scores is 300 (all the questions are classified as difficult). The adjusted scores are reported in Table 9.

The average adjusted exam score for the midterm exam 1 with GIN option is higher compared to average adjusted exam score in section without GIN (210.9 vs. 207.6). In other

words, the first midterm exams in sections with GIN were on average more difficult compared to the first midterm exams in earlier sections without GIN. This however, is not the case for the second midterm exam and the final exam, where the average difficulty of the exam was higher in sections using GAD rather than GIN. This does not seem to support my hypothesis, that the scores on the exams in sections using GIN are lower due to increased difficulty of the exams. However, it is not clear based on what criteria were those MCQ classified as easy, intermediate, and difficult and who classified them as such. Hence, even these adjusted scores probably do not fully and accurately adjust for exam difficulty.

Semester/Year/Section	Midterm Exam 1	Midterm Exam 2	Final Exam
<b>Sections without GIN</b>			
Spring 08 Micro	206.7	200.0	198.6
Spring 08 Macro	210.0	200.0	198.6
Summer 08 Micro	190.0		175.5
Summer 08 Macro	208.0		208.6
Fall 08 Macro	223.3	166.7	180.0
Average	207.6	188.9	192.3
<b>Sections with GIN</b>			
Fall 08 Micro	213.3	193.3	204.3
Spring 09 Morning Micro	216.6	200.0	191.4
Spring 09 Afternoon Micro	216.6	200.0	191.4
Summer 09 Macro	218.0		208.0
Summer 09 Macro	192.0		156.0
Fall 09 Micro	223.3	193.3	204.3
Fall 09 Macro	196.7	143.3	180.0
Average	210.9	186.0	190.8

Finally, the skills and knowledge obtained by completing Aplia assignments might be different from the one needed in order to do well on the exams. Recall that Aplia offers fairly complex, multistage questions that require not only numerical calculations but also manipulation of graphs or derivations of curves. All the questions on the exams, however, are MCQ and may require a different type of preparation. As a result, the skills learned in Aplia (either under GIN or GAD) do not translate well to MCQ based exams.

### CONCLUDING REMARKS

The purpose of this paper was to study the effect of Grade It Now (GIN) option in Aplia on student learning. Based on the data used in this study I conclude that Aplia's GIN positively impacts students' grades on Aplia assignments. This impact is especially strong for academically weaker students and has the same impact on students' grade as does increasing students' GPA by almost half a point.

However, in sections where Aplia' GIN option was selected, students' performance on midterm exams and final exam was either not statistically different from sections with Grade At

Deadline (GAD) option or was actually worse. In addition, using OLS regression, I show that Aplia's GIN impact on students' performance on exams is negative and does not improve student learning.

One possible explanation for this might be due to students' trying to "game" the system by increasing their grades on assignments and lowering their efforts on exams. Since there is no difference between using GIN vs. GAD on the final grade, this behavior seems to be supported with the data.

A drawback of this study stems from the fact that the exams were not identical from section to section. For future research, I thus recommend including a consistent evaluation mechanism such as a pre- and post-test and administering the Test of Understanding in College Economics (TUCE) as used by Lee et al. (2010). In addition, the type of questions asked on the exams could also be diversified, so that the students can more easily showcase their skills (such as graph manipulations and curve derivations) that are greatly emphasized in Aplia, but not directly tested with MCQ.

Finally, all the sections that include the GIN option were taught *after* the sections with GAD. As a result, it is possible that with the passage of time the teaching efficiency of the professor might have also changed. For example, with more practice, teaching skills might improve or with more repetition teaching fatigue and indifference could creep in. In order to more convincingly test the impact of GIN on students' performance and learning the sections with GIN and GAD should be taught simultaneously in the same semester.

## APPENDIX

The textbook used by the instructor in Principles of Macroeconomics (ECO 201) was *Brief Principles of Macroeconomics*, 4<sup>th</sup> by N. Gregory Mankiw (2007a). Below is the list of assignments and corresponding book chapters that were assigned in this class.

### Assignments in Principles of Macroeconomics

- HW1 Introduction to Using Aplia Problem Sets
- HW2 Thinking Like an Economist (Chapter 2)
- HW3 Interdependence and the Gains from Trade (Chapter 3)
- HW4 The Market Forces of Supply and Demand (Chapter 4)
- HW5 Measuring a Nation's Income (Chapter 5)
- HW6 Measuring the Cost of Living (Chapter 6)
- HW7 Unemployment (Chapter 10)
- HW8 The Monetary System (Chapter 11)
- HW9 Aggregate Demand and Aggregate Supply (Chapter 15)
- HW10 The Influence of Monetary and Fiscal Policy on Aggregate Demand (Chapter 16)
- HW11 Production and Growth (Chapter 7)

### Assignments in Principles of Microeconomics

The textbook used by the instructor in Principles of Microeconomics (ECO 202) was *Principles of Microeconomics*, 4<sup>th</sup> by N. Gregory Mankiw (2007b). Below is the list of assignments and corresponding book chapters that were assigned in this class.

- HW1 Introduction to Using Aplia Problem Sets
- HW2 Thinking Like an Economist (Chapter 2)
- HW3 Interdependence and the Gains from Trade (Chapter 3)

- HW4 The Market Forces of Supply and Demand (Chapter 4)
- HW5 Elasticity and its Application (Chapter 5)
- HW6 Supply, Demand, and Government Policies (Chapter 6)
- HW7 Consumers, Producers, and the Efficiency of Markets (Chapter 7)
- HW8 Application: The Costs of Taxation (Chapter 8)
- HW9 Application: International Trade (Chapter 9)
- HW10 Externalities (Chapter 10)
- HW11 Public Goods and Common Resources (Chapter 11)
- HW12 The Costs of Production (Chapter 13)
- HW13 Firms in Competitive Markets (Chapter 14)

## END NOTES

1. Other academic publishers have developed their own online homework and learning solutions; Pearson with MyEconLab, McGraw-Hill with Connect, and Wiley with WileyPlus. In addition, some online learning tools, like Sapling and TopHat, work independently from textbook publishers.
2. I dropped 87 students from our original data that was obtained from the registrar and merged with Aplia gradebook. These students were registered for these classes but either did not take any exams and/or completed fewer than two assignments in Aplia. I regard these students as officially enrolled but they unofficially dropped the class.
3. These t-test assumed equal variances. This assumption was also tested and verified with variance comparison tests.
4. It should also be mentioned that a small fraction of students took both courses with the same professor (in either order) and who were thus familiar with Aplia in their second class. Unfortunately, I am unable to isolate those students in the data and control for this.
5. Race: 0 - White, 1 - Black, 2 - Asian, 3 - Other, Major: Business Administration and Urban Affairs - 0, Education and Human Services Education - 1, Engineering - 2, Liberal Arts and Social Sciences - 3, Sciences and Health Professions - 4, Other (undecided, transient, nondegree) - 5.
6. All of the assignments *followed* class discussions and lectures. Class size might be less important in the situation where the assignments *precede* class discussion.

## REFERENCES

- Ball, S.B., Eckel, C. & Rojas, C. (2006). Technology improves learning in large principles of economics classes: Using our WITS. *The American Economic Review*, 96(2), 442-446.
- Bianco, C.A., Lindblom, C. & Nixon, M. (2014). The impact of online assignments on student performance. *Financial Education Association*, 12(3),
- Bonham, S., Beichner, R. & Deardorff, D. (2006). Online homework: Does it make a difference? *The Physics Teacher* 39, 293-296.
- Butler, M.M. & Zerr, J. (2005). The use of online homework systems to enhance out-of class student engagement. *International Journal for Technology in Mathematics Education*, 12(2), 51-58.
- Flannery, D., Kennelly, B. & Considine, J. (2013). Paper and online assignments in economics: A comparison with panel data. *International Review of Economics Education*, 13, 32-43.
- Kennelly, B., Considine, J. & Flannery, D. (2011). Online assignments in economics: A test of their effectiveness. *The Journal of Economic Education*, 42(2), 136-146.
- Lee, W., Courtney, R.H. & Balassi, S.J. (2010). Do online homework tools improve student results in principles of microeconomic courses? *American Economic Review: Papers and Proceedings* 100, 283-286.
- Mankiw, G.N. (2007a). *Brief Principles of Macroeconomics* (4th ed.). Thomson South-Western, Cengage Learning.
- Mankiw, G.N. (2007b). *Principles of Microeconomics* (4th ed.). Thomson South-Western, Cengage Learning.
- Rhodes, M.T. & Sarbaum, J.K. (2013, October). Online homework management systems: Should we allow multiple attempts? Working paper, University of North Carolina, Greensboro.
- Richards-Babb, M., Drelick, J., Henry, Z. & Robertson-Honecker, J. (2011). Online homework, help or hindrance? What students think and how they perform. *Journal of College Science Teaching*, 40(4), 81-93.

- Self, S. (2013). Utilizing online tools to measure effort: does it really improve student outcome? *International Review of Economics Education*,14, 36-45.
- Taylor, M. (2013). *Improving student learning using aplia: A study of the impact of engage learning's aplia on student learning in developmental english*. Cengage Learning. Peer Reviewed White Paper.