Title
Hidden Costs of Using Gamification to Improve Students' Study Behaviors in an Engineering Course

Author
Xu, Luofeng

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Hidden Costs of Using Gamification to Improve Students' Study Behaviors in an Engineering Course

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Mechanical Engineering

by

Luofeng Xu

September 2022

Thesis Committee:
Dr. Thomas F. Stahovich, Chairperson
Dr. Bhargav Rallabandi
Dr. Erfan Nozari

The Thesis of Luofeng Xu is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Hidden Costs of Using Gamification to Improve Students' Study Behaviors in an Engineering Course

by

Luofeng Xu

Master of Science, Graduate Program in Mechanical Engineering
University of California, Riverside, September 2022
Dr. Thomas F. Stahovich, Chairperson

This thesis examines students' homework behaviors and their relationship to academic achievement in an introductory-level mechanical engineering course. Prior work has shown that engaging in effective study behaviors such as distributing study effort over time rather than camming, can have a positive effect on learning outcomes. In the present study, we investigate the use of gamification as a means of motivating students to engage in effective study behaviors such as completing reading assignments, taking notes in lecture, and starting homework assignments early. The study employed a web-based dashboard system that provided students with quantitative measures of study effort. Effort on written work was measured with smartpens that digitize students' work in real time, and an instrumented document viewer that measured reading effort. The dashboard also included games that provided points for completing study tasks as measured by the smartpen and document viewer. This thesis examines how gamification affects students' learning behaviors and learning outcomes. The results demonstrate a strong and

consistent relationship between students' learning behaviors and learning outcomes. Additionally, this work demonstrates that gamification is effective at changing students' study behaviors, but is *ineffective* at improving students' performance in engineering courses.

Table of Content

List of Figures

List of Tables

Chapter 1 Introduction

Research Summary

This study was conducted in a 10-week lower division course "ME010 Statics", in the University of California, Riverside in 2017 (the control group) and 2019 (the experimental group). In the present study, we evaluate the effectiveness of gamification implemented as part of a web-based dashboard system (Figure 1) that provided students with quantitative measures of study effort. Efforts on written work was measured with Livescribe smartpens that digitize students' work in real time, and an instrumented document viewer called DocViewer (Figure 2) that measures reading effort. The games provide points for completing at least 50% of each homework assignment at least 24 hours before the due day, submitting lecture notes, and completing weekly reading assignments. In the experimental group, the gamification was used during the entire course, and students chose whether or not they wished to participate. Students who opted in were offered the web-based dashboard system, document viewer and Livescribe smartpen, which kept track of effort on weekly homework assignments, lecture notes and reading assignments. The dashboard provided students with quantitative feedback about what they had accomplished and provided games points for completing the games. The games included levels, which provided an increasing number of points for successfully completing tasks multiple times in a row. For example, at the first level, students received 30 points for completing at least 50% of a homework assignment more than 24 hours before the deadline. Doing this on two consecutive assignments promoted the student to the second level in which completing at least 50% of the homework early is worth 60 points (30 points multiplied by level 2). Game points were added to students' final course grade. We designed multiple measurements to evaluate students' homework behaviors. For example, because we recorded writing activities with timestamps, we calculated the lead time (how early the writing was done relative to the due date). We computed the average of this measure across all homework assignments to obtain each student's mean lead time. We compared students' performance between the experimental and control group using this and other measures, including scores on pre- and post-tests. The control group were students in an identical course except there was no gamification. Just as for the experimental group, these participants used smartpens and the instrumented document viewer, but the dashboard provided measures effort without games.

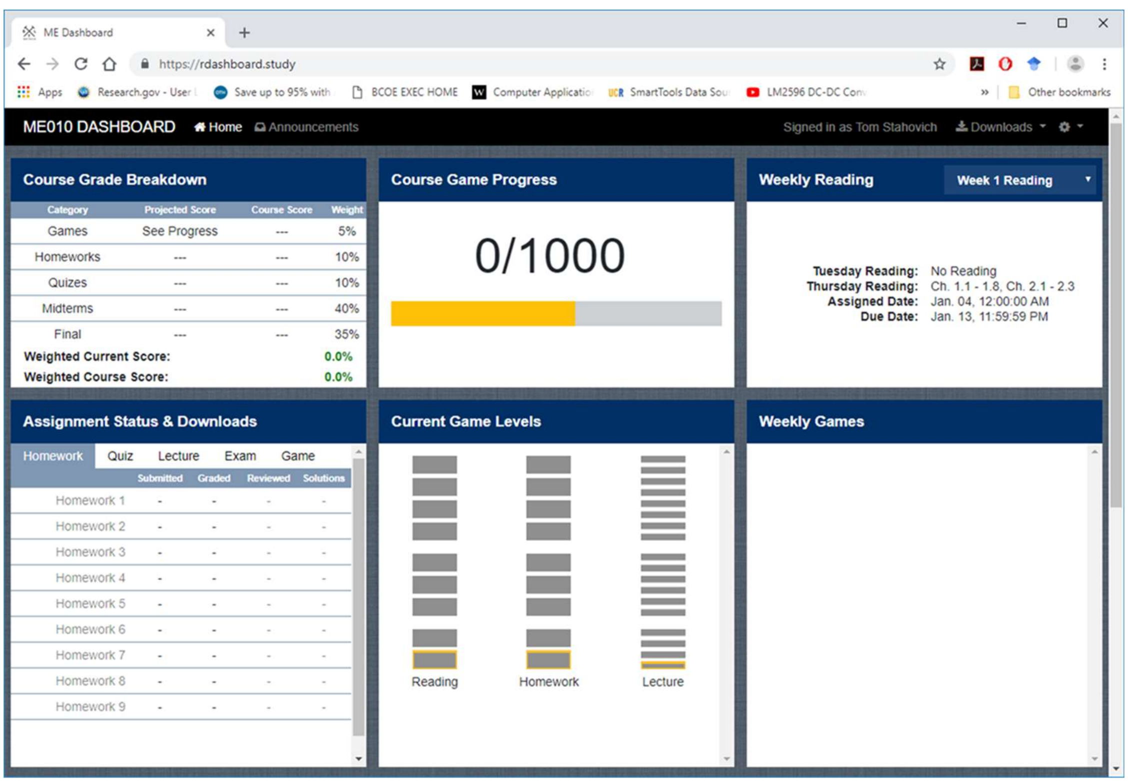
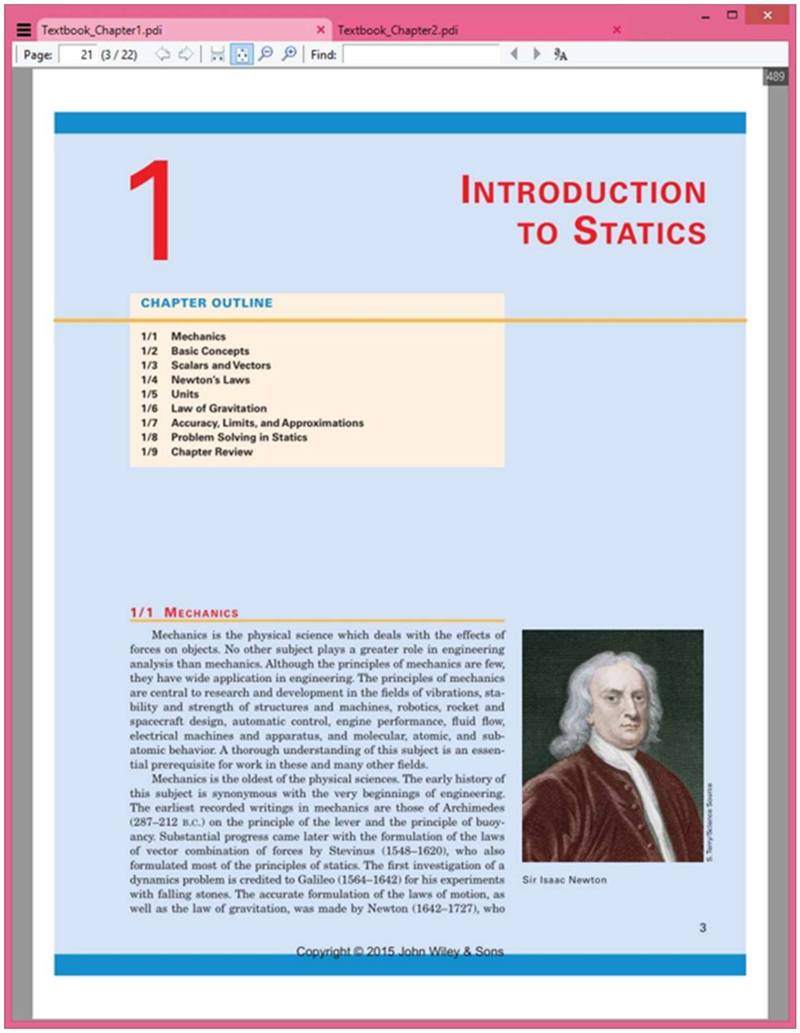


Figure 1 Dashboard interface



1

INTRODUCTION TO STATICS

CHAPTER OUTLINE

1/1 MECHANICS

Mechanics is the physical science which deals with the effects of forces on objects. No other subject plays a greater role in engineering analysis than mechanics. Although the principles of mechanics are few, they have wide application in engineering. The principles of mechanics are central to research and development in the fields of vibrations, stability and strength of structures and machines, robotics, rocket and spacecraft design, automatic control, engine performance, fluid flow, electrical machines and apparatus, and molecular, atomic, and subatomic behavior. A thorough understanding of this subject is an essential prerequisite for work in these and many other fields.

Mechanics is the oldest of the physical sciences. The early history of this subject is synonymous with the very beginnings of engineering. The earliest recorded writings in mechanics are those of Archimedes (287–212 B.C.) on the principle of the lever and the principle of buoyancy. Substantial progress came later with the formulation of the laws of vector combination of forces by Stevinus (1548–1620), who also formulated most of the principles of statics. The first investigation of a dynamics problem is credited to Galileo (1564–1642) for his experiments with falling stones. The accurate formulation of the laws of motion, as well as the law of gravitation, was made by Newton (1642–1727), who

Sir Isaac Newton

3

Figure 2 Document Viewer

Objective and Rationale

Study behaviors (including homework behaviors) have an important role in academic achievement (Rawson, Stahovich & Mayer, 2017). For example, Jones and Ruch (1928) examined the relationship between the amount of time spent studying and first semester grade point average. More recently, Credé and Kuncel (2008) conducted a meta-analysis of 10 study behavior, skill, and attitude inventories (SHSAs) and found that they had incremental validity in predicting academic performance.

Consider an introductory-level college course in engineering in which students must hand in weekly assignments. Previous work has shown that engaging in recommended study behaviors such as working on weekly assignments more than 24 hours before the deadline is related to good learning outcomes (as measured by exam score) in an introductory-level engineering course (Gyllen, Stahovich, & Mayer, 2020). However, many students choose to work on their assignments just before the deadline (Gyllen, Stahovich, & Mayer, 2020). The goal of the present study is to examine techniques for improving student study behavior in such a course. Specifically, the goal is to examine the effectiveness of adding gamification features to an introductory-level engineering course aimed at rewarding students for working on weekly assignments more than 24 hours before the deadline.

The rationale for this study is that we expect students in introductory-level college courses to know how study, but we rarely encourage them to use effective study strategies (Dunlosky et al., 2013; Fiorella & Mayer, 2015; Kiewra, 2022; Mayer, 2019; Miyatsu, Nguyen, & McDaniel, 2018. Study strategies (also called learning strategies) are behaviors that the student engages in during learning that are intended to improve the learning outcome (Fiorella & Mayer, 2015; Mayer, 2019). Thus, part of the hidden curriculum for college students is to learn how to learn. This task is particularly important for students who have not received appropriate study skill training from teachers, parents, or peers along the way. In order to promote student success in college, instructors not only need to present the to-be-learned material, but also need to guide students in how to learn the material effectively.

Chapter 2 Literature Review

Learning Theory

The three dominant models of learning are the behavioral, cognitive and constructivist models (Schunk, 2019). Behaviorist theory focuses on how people learn and form habits. To assess learning outcomes, behavioral models rely on visible changes in the student's behavior. Cognitivism, on the other side, considers people to be mental beings who examine and evaluate data. The major difference is behaviorism focuses on observable phenomena whereas cognitivism focuses unobservable thoughts. Constructivism is founded on the premise that people construct their own understanding and interpretation of new ideas based on their existing knowledge and experience. The theory also asserts that all knowledge and learning exist solely inside the mind. With the new developments, advances, and academic challenges in the fields of engineering and technical sciences, it will not be sufficient to employ only one learning method that is based on one single educational theory. For instance, the behavioristic method of learning for engineering students should include some cognitive aspects (e.g., let the student learn using his/her own way of thinking) and the socio-cultural factor (e.g., in the form of group work, artefacts, project demonstration, etc.) (Hassan 2011).

Data Mining of Homework

Homework is defined as “tasks assigned to students by schoolteachers that are meant to be carried out during non-school hours” (Cooper, 1989). Homework has the potential to improve academic learning, perhaps by extending time to learn beyond the classroom and priming active cognitive processing for learning (Cooper, 1989, 2001; Mayer, 2011).

There is a lot of evidence demonstrating the educational value of homework (Cooper, Robinson, &Patall, 2006; Carter, 2009; Xu, 2013). But most of this research is based on students' self-report and surveys of their study habits, which limits the reliability of the results. A recent study demonstrated that the amount of time students actually spend on homework activities is considerably less thanthe amount of time they report spending on homework activities (Rawson, Stahovich & Mayer, 2017). To obtain an accurate measure of students' effort, in our study we used smartpen techniques to track the details of students' homework behaviors in real time (Herold, Stahovich, Lin, & Calfee, 2011). This technology offers a level of detail about what students are doing and when they are doing it that cannot be obtained through traditional research methodologies.

Gamification

Gamification, which is "the use of game design elements in nongame context (Deterding, Dixon, Khaled & Nacke, 2011)," has been widely applied in promoting learning motivation, engagement, collaboration, and effectiveness (Dicheva, Dichev, Agre & Angelove, 2015; Dichev & Dicheva, 2017). Researchers have investigated the effectiveness of gamification mechanisms in improving collaborative learning. Students in gamified learning environments were shown to be more engaged than those in the comparison group, contributing more posts and replies to online discussion forums (Barata, Gama, Jorge & Goncalves, 2013; Hew, Huang, Chu & Chiu, 2016). Gamification of a college course involves adding game-like features aimed at rewarding desired behaviors (Kapp, 2012; Mayer, 2014). In the present study, we evaluate the effectiveness of gamification implemented as part of a web-based dashboard system that provided students with quantitative measures of study effort. The games provided points for completing at least 50% of a homework assignment at least 24 hours before the due day, submitting lecture notes, and completing a weekly reading assignment. The games included levels, which provided an increasing number of points for successfully completing tasks multiple times consecutively. Effort on written work was measured with smartpens that digitize students' work in real time, and an instrumented document viewer that measured reading effort.

Theoretical Framework and Predictions

We examine two contrasting theoretical frameworks--one that focuses on extrinsic motivation to engage in effective study behaviors and one that focuses on intrinsic motivation to engage in effective study behaviors. Gamification can act as an extrinsic motivation for students to work on assignments far in advance of the deadline because this behavior is externally rewarded (such as adding to the student's point total for the course grade). If extrinsic motivators are effective, then we predict that students in the gamification group will achieve a greater mean lead time than students in the non-gamified group (hypothesis 1) and will achieve a higher score on the final exam which taps the material contained in the weekly assignments (hypothesis 2a), and there will be a substantial correlation between mean lead time and final exam score for both groups (hypothesis 3a).

In contrast, according to the intrinsic motivation view, external rewards can temporarily change study behavior but that does not necessarily translate into better learning. This view suggests that students need intrinsic motivation to exert effort to understand the material, which gamification does not necessarily provide (Deci & Ryan, 1985). In short, there is a hidden cost of gamification analogous to the long-standing research on the hidden cost of reward in which rewarding students for doing something they like doing tends to decrease their willingness to engage in that behavior when it is no longer rewarded (Lepper & Greene, 2015; Lepper, Henderlong, & Gingras, 1999). Students can justify their reason for engaging in a rewarded behavior by saying they are doing it for the reward, but that will not increase their intrinsic motivation to engage productively or over the long term. If intrinsic motivation is the key to academic success, then we predict that students in the gamification group will achieve a greater mean lead time than students in the non-gamified group (hypothesis 1), but this will not lead to better learning outcomes as measured by final exam score (hypothesis 2b) and gamification will diminish the relationship between mean lead time and final exam score as compared to the non-gamification group (hypothesis 2c).

Chapter 3 Design and Method

Participants

The participants were undergraduate engineering students who enrolled in and completed a 10-week lower division course, ME010 Statics, at the University of California, Riverside in the 2017 and 2019 academic years. This course was taught by the same instructor both years. There were a total of 96 participants in 2017 (control group) and 91 in 2019 (experimental group). Most students were from Mechanical Engineering major or other related engineering majors.

Course Materials and Design

Statics is the part of engineering mechanics focused on the equilibrium of objects subject to forces. It is the foundation for many of the branches of engineering. Statics is critical to the engineering curriculum and serves to solidify the student's understanding of other related subjects, such as physics and geometry. The solution to a statics problem typically includes free body diagrams and equilibrium equations. The statics course used in this study covers equilibrium of coplanar force systems; analysis of frames and trusses; noncoplanar force systems; friction; and distributed loads. Figure 3 is an example of a typical statics problem from the course.

The coefficient of static friction between block A and its incline is 0.25. What must the minimum coefficient of static friction between block B and its incline be, if the blocks are in equilibrium? Neglect friction in the pulley.

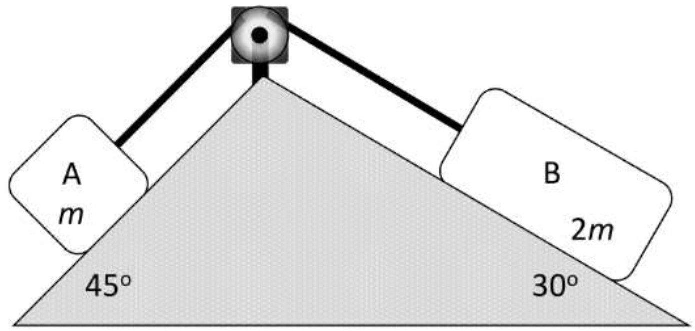


Figure 3 A typical statics problem

During both years, this course was scheduled on Tuesday and Thursday mornings with 80-minute lectures. The 50-minute discussion section was scheduled on Wednesday afternoons. Students from both groups used Livescribe smartpens to complete written work, including lecture notes, quizzes, homework, and exams. Both groups of students had similar homework assignments. A Livescribe smartpens is an ink pen that also digitizes the writing (Herold, Stahovich, Lin, & Calfee, 2011). These devices are used with special dot-patterned paper. As the Smartpens record writing with timestamps, they enable us to analyze the temporal properties of students' learning activities.

Students were also required to submit all their assignments through a software developed in the former research called "InkViewer". Meanwhile, students were provided a web-based visualized system named "dashboard", it automatically uses their ink data and measure their study behaviors. Students from the experimental group were told how the dashboard system works, and were also giving game points for completing three different games. A homework game comprises completing at least 50% of a homework assignment at least 24 hours before the due day. A lecture game comprises taking notes during a lecture and submitting them. A reading came comprises completing a reading assignment by the assigned due date. There are nine possible reading games, nine possible homework games, and 18 possible lecture games. Each reading and homework game is worth 30 points and a lecture game is worth 15 points. As the student advances to higher levels of a game, the possible points are increased with a level multiplier. If a student completes *n* (*2n* for lecture) consecutive games at level *n-1*, then the student advances to level *n*. We refer to this a *streak*. For example, if a student at level 1 of the homework game completes a streak of two homework games in a row (i.e., completes at least 50% of the work early on two consecutive homework assignments), the student advances to level 2. Breaking a streak resets the streak count to 0 but does not reset the level. Once a level is reached, it is locked in.

The number of points achieved for a particular game is the product of the nominal game points (i.e., 15 or 30) and a multiplier equal to the game level. Consider, for example a student who completes the homework game for assignments 1, 2, and 3. The first two games will be at level 1 and, because the student completed a streak of two games, the third game will be at level 2. Thus, the points earned will be:

$$30 \times 1 + 30 \times 1 + 30 \times 2 = 120;$$

Consider another student who completed the homework games for all nine assignments except assignment 5. The points earned will be:

$$30 \times 1 + 30 \times 1 + 30 \times 2 + 30 \times 2 + 0 + 30 \times 2 + 30 \times 2 + 30 \times 2 + 30 \times 3 = 450$$

Notice that by missing the game for assignment 5, the student broke the streak needed to advance to level 3, requiring the student to begin a new streak comprising the games for assignments 6, 7, and 8. This system of points and levels is designed to motivate students to complete the games beginning at the start of the course and to continue completing the games until the end. In total, there were 1800 game points achievable, 600 for each of the three game types.


```
v0.4
1-Subject_Notebook_7--Copy_000--Page_006
1412031991904
2017-01-17 19:15:11 Z
2017-01-17 19:15:42 Z
AYE-AP7-EFP-XM
1
815
27
2606 840 72638029894
2608 837 72638029895
2609 834 72638029896
2611 831 72638029897
2613 828 72638029898
2615 826 72638029899
```


Figure 4 An example of digital ink data

At the beginning of the quarter, students were asked to complete a pre-test through the Blackboard course management system. The pre-test was the Force Concept Inventory (FCI), which is a test measuring mastery of concepts commonly taught in a first semester of physics developed by Hestenes, Halloun, Wells, and Swackhamer (1985). It tests students' knowledge of fundamental concepts of forces taught in physics. We use the pre-test score as a measure of students' average knowledge level before upon starting the statics course.

Every week students were required to attend two lectures and one discussion session. Each week, students were asked to complete 1 homework assignment and attend two lectures and take notes. Students were asked to complete this work with a smartpen and to submit it electronically through InkViewer. Each week, students were also asked to complete a reading assignment using DocViewer. Each week (with some exceptions), students were assigned a quiz during lecture. Quizzes were similar to homework problems from the most recently submitted homework assignment. There were two midterm exams and a comprehensive final exam. All quizzes and exams were closed-book and closed-note. Students solved the quizzes and exams with a smartpen and submitted the work electronically.

After the final exam, students were asked to complete a post-test through the Blackboard course management system. The post-test was the Statics Concept Inventory, which is a quantification of conceptual understanding of students' knowledge of statics (Steif, Dantzler, 2005). We use the post-test scores as a measure of students' knowledge level after they finish the class.

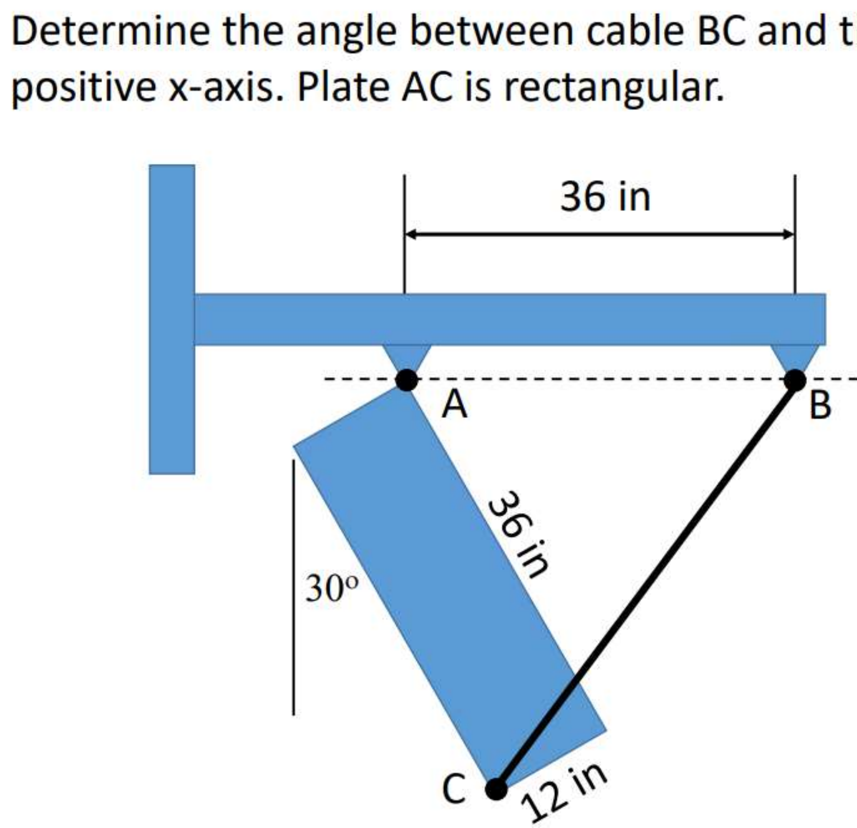


Figure 5 A quiz question from the control group

Chapter 4 Data analysis

Measurements

After students submit their work with InkViewer, the software renders the ink data, enabling students to navigate through the pages of writing. Students use the mouse to select ink for an individual problem and assign it to that problem. In this way, students were asked to label their ink strokes into the individual problems for each assignment, quiz, and exam. In this research, we consider only the writing form homework assignments. Building on prior research (Rawson, Stahovich & Mayer, 2017), we employ six quantitative measurements to characterize a student's homework activity, as summarized in Table 1.

Table 1 Six Measurements derived through smartpen technology

Measure name	Description
Total strokes	Total number of pen strokes written to complete all the assignments.
Mean stroke fraction D0	Proportion of pen strokes written within 24 hours of the due date, averaged across all assignments
Mean stroke fraction D-1	Proportion of pen strokes written between 48 and 24 hours of the due date, averaged across all assignments.
Mean stroke fraction D0 and D-1	Proportion of pen strokes written within 48 hours of the due date, averaged across all assignments.
Stroke lead time	Average time of pen strokes relative to the due date, averaged across all assignments.
Std stroke lead time	The standard deviation of pen strokes time relative to the due date, averaged across all assignments.

Figure 4 shows our data structure for pen strokes. A data point is represented by the triple: (x, y, timestamp). All of the points between when the pen touches the paper and when it leaves the paper represent a single pen stroke. In our analysis, we drop pen strokes with less than 4 points as they may represent a tap of the pen on the paper rather than writing. The timestamp of the first point of a pen stroke is taken to be the timestamp for the stroke. Our first measurement, *total strokes*, is the cumulative number of pen strokes for all assignments.

We use three measurements to characterize the time effort over the assignment period. *Mean stroke fraction D0* is the fraction of the pen strokes written within 24 hours of the due date. Similarly, *mean stroke fraction D-1* is the fraction of the pen strokes written between 48 and 24 hours before the due date. Finally, the *mean stroke fraction D0 and D-1* is simply the sum of previous two measures, and thus is the fraction of the pen strokes written within 48 hours of the due date.

As an additional way of quantifying when a student works on an assignment, we compute the *stroke lead time*, which is the average time of writing relative to the homework assignment due date. Here we define D_j as the due date of homework assignment HW_j and T_i as the timestamp of pen stroke i. We compute the centroid, τ_j, of the distribution of work for HW_j as:

$$\tau_j = \frac{\sum_{i=1}^{N}(D_j - T_i)}{N}$$

where N is the total number of pen strokes written for the assignment. Similarly, *Std stroke lead time* is the standard deviation of $D_j - T_i$. This is averaged over all assignments and measures the distribution effort on homework assignments.

Verifying Data

Students were required to assign their strokes after they finish each assignment, we wanted students to assign all their work related to the assignment, but there is a chance students forgot to label strokes. To avoid that, we manually verified their assignments data by using InkViewer from the host side (Figure 6), which enable us to review all of students' assignment data. The view of pages is sorted by students: if a student did not label some pages, these pages will be placed at the end.

To verify the data, we needed to find the unlabeled strokes and pages. First, we generated the breakdown points of student's homework from database. If a student received zero points for a particular homework assignment (e.g., homework 4), there is a possibility this student is missing some stroked from this assignment. Software already sorted the pages, it the page was in the time window but not labelled, then it does not get flagged. To avoid missing pages, we sorted the pages in the time window from pen stroke data directly and extend for 1 day because of the late work(e.g., The time window for homework 4 is from Jan.1 to Jan.7, we will find the pages written during Jan.1 to Jan.8). Next, we find the last stroke of the previous assignment(e.g., homework 3) in the InkViewer. We flipped through the pages in the InkViewer until we reach the next assignment(e.g., homework 5). At the same time, we compared the pages generated from pen stroke data and the pages from InkViewer, examined the difference and labeled the missing strokes. Once we figured out there were unlabeled strokes in the assignment, we could assign them.

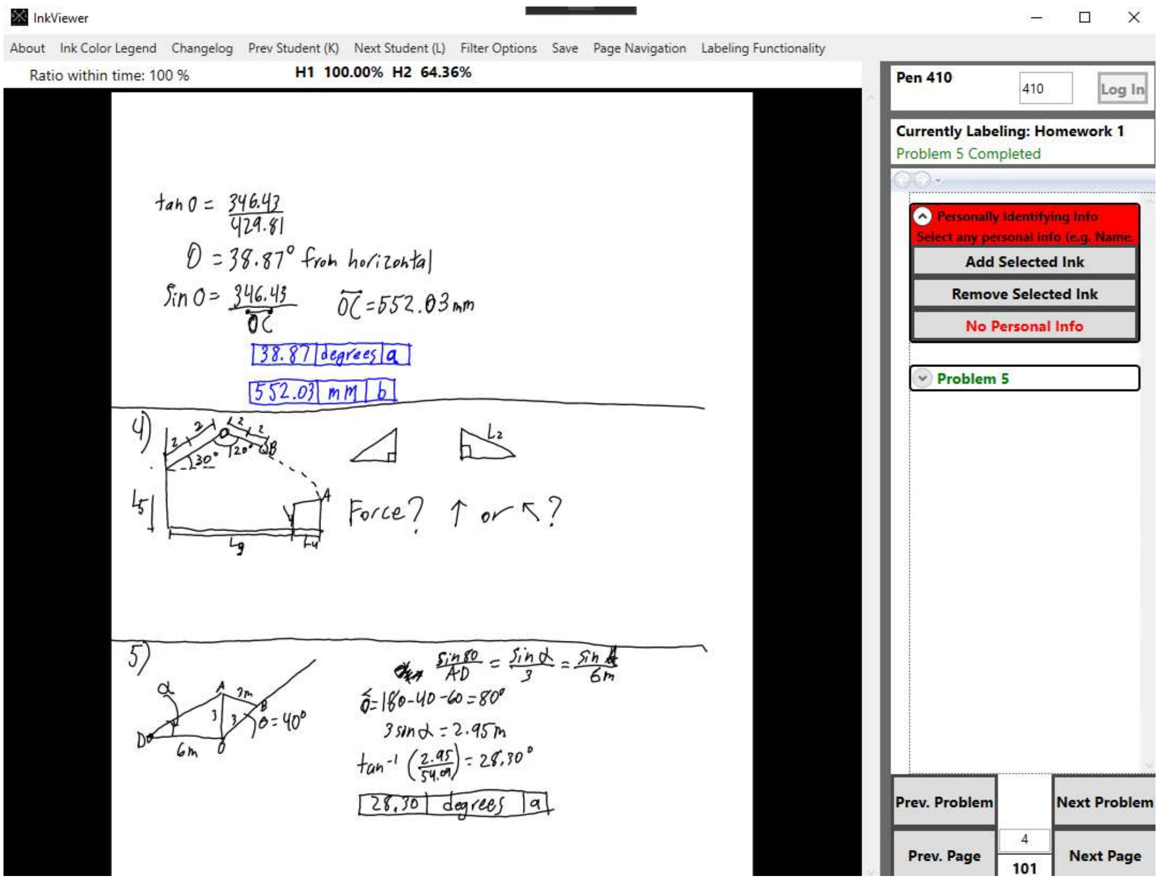


Figure 6 An example of students missing label for question 4 and 5.

Chapter 5 Results

Group Characteristics

Our dataset includes data on 6 measures from a total of 187 students: 96 from Year 2017 (control group) and 91 from Year 2019 (experimental group). All of these students completed the course and received a final course grade. In Table 2, 87 students from the control group and 71 students from the experimental group completed the pre-test. We want to determine if both groups have the same study level before they enter this class. After the tests of normality, the pre-test score is normal distributed in the experimental group (Sig=.094) but non-normal distributed in the control group (Sig=.009), so we apply the Mann–Whitney U test to determine the difference. We define "Con.G" as the control group, "Exp.G" as the experimental group. The results of p=.672 (>0.05) showed that two groups did not differ significantly on prior knowledge on the force concept inventory, so we consider both groups to have the same level of background knowledge.

Table 2 Descriptive statistics of pre-test scores from two groups

Quarter	N	Mean	Std. Deviation	Shapiro-Wilk Sig.
Con.G	87	10.92	4.840	.009
Exp.G	71	11.23	5.524	.094

Table 3 The Mann-Whitney U test result of pre-test scores from two groups

Quarter	Pre-test	Z	p
Con.G	10(7~15)	-0.424	0.672
Exp.G	11(7~16)		

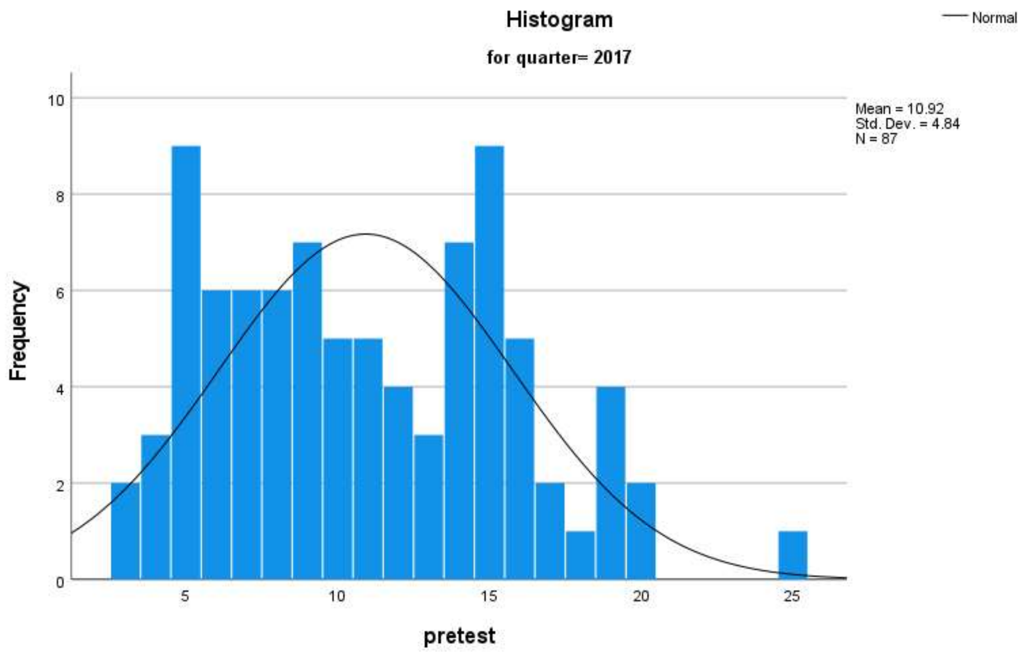


Figure 7 The figure of normality test for the control group

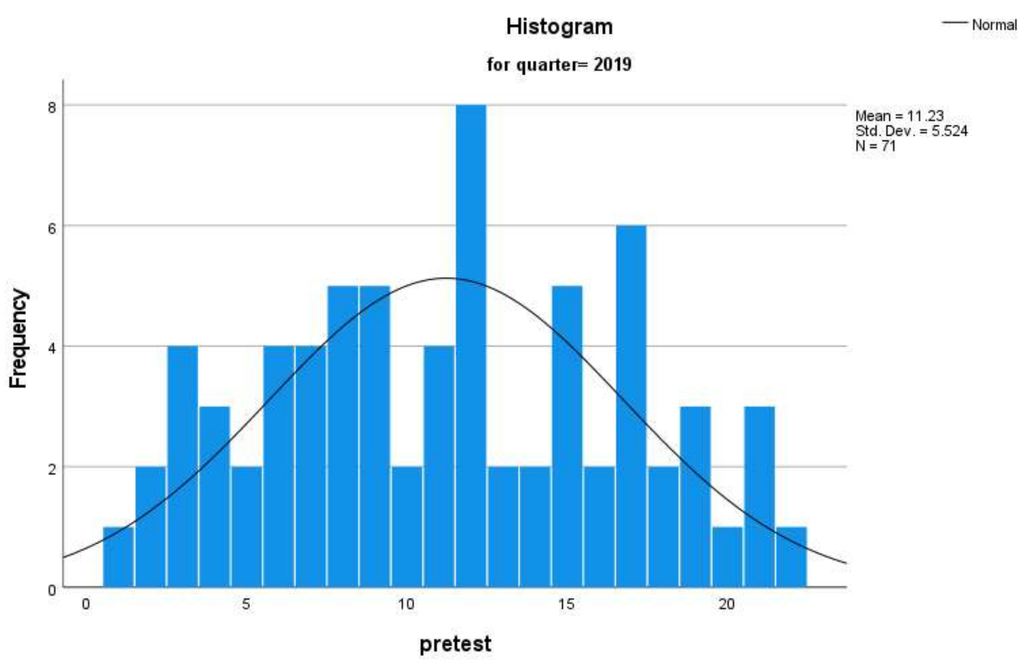


Figure 8 The figure of normality test for the experimental group

Study Behaviors

Our data in Table 4 shows the mean value of the measurements. We proceed the Shapiro-Wilk tests of normality in Table 5, only the total strokes from both groups are normal distributed (sig>0.05), so we will apply independent t-test to it and Mann–Whitney U test to the rests.

Table 4 Descriptive statistics of the measurements for both group

Measurements	Control Group		Experimental group	
	Mean	Std	Mean	Std
Total strokes	24940.26	9959.13	29285.64	11061.691
Mean stroke fraction D0	.6809	.289	.4442	.2697
Mean stroke fraction D-1	.1784	.1563	.3678	.1965
Mean stroke fraction D0 and D-1	.8593	.2038	.812	.2089
Mean stroke lead time	18.71	16.81	25.30	14.265
Std stroke lead time	6.93	5.19	9.69	5.31

Table 5 The test of normal normality of the measurements for both group

Measurements	Control Group		Experimental group	
	N	Sig	N	Sig
Total strokes	96	0.117	91	0.809
Mean stroke fraction D0	96	0	91	0.034
Mean stroke fraction D-1	96	0	91	0.529
Mean stroke fraction D0 and D-1	96	0	91	0
Mean stroke lead time	96	0	91	0.013
Std stroke lead time	96	0	91	0.086

An independent t-test was conducted to explore the difference between the control group and the experimental group in the total number of pen strokes. Both groups were normally distributed. $F = 0.999$, $p > 0.05$. Hence, equal variances were assumed. A statistically significant difference was evident between the two groups, $t(185) = -2.826$, $p < 0.01$. This result indicates that students in the experimental group were writing more strokes than the control group.

Table 6 The independent t-test result of Total Strokes for both group

Measurement	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	N	Sig.
Total Strokes	0.999	0.319	-2.826	185	0.005

A Mann–Whitney U test was conducted to explore the difference between the control group and the experimental group in the Mean stroke fraction D0. Both groups were not normally distributed. A statistically significant difference was evident between the two groups, $p < 0.01$ with a moderate effect size $r = 0.40$. In the control group, students completed 68.09% of their total work in the last 24 hours; in the experimental group, students completed 44.42% of their total work. This result indicates that students in the experimental group were having less work in the last 24 hours before day D0 than students in the control group, which is consist with our expectation. Students were not stacking less work on day D0 because of the gamification.

Table 7 The Mann-Whitney U test result of Mean stroke fraction D0 from two groups

Quarter	Mean stroke fraction D0	Z	p
Con.G	0.7422(0.4697~0.9440)	-5.523	0
Exp.G	0.4105(0.2394~0.6492)		

Next, we continue conducting Mann–Whitney U test to explore the difference between the control group and the experimental group in the Mean stroke fraction D-1 and Mean stroke fraction D0 and D-1. The data for Mean stroke fraction D-1 in the control group was not normally distributed, but the data in the experimental group was normally distributed (Sig > 0.05); both groups were not normally distributed for Mean stroke fraction D0 and D-1. A statistically significant difference was evident between the two groups, $p < 0.01$ with a moderate effect size $r = 0.48$ for Mean stroke fraction D-1 and $p = 0.028$ with a small effect size $r = 0.16$ for late strokes fraction. In the control group, students completed 17.84% of their total work between last 48 hours to last 24 hours; in the experimental group, students completed 36.78% of their total work. Students in the experimental group were having more work stacking one day before the D0. The total work fraction in the last 48 hours is 85.93% in the control group and 81.20% in the experimental group, which is slightly close, but the ratio between last 48 hours is 3.82 in the control group and 1.21 in the experimental group. Although the most of work were still stacking in the last 48 hours before the day D0, compare to the control group, students do separate their work into at least 2 days after we employed the gamification in the experimental group.

Table 8 The Mann-Whitney U test result of Mean stroke fraction D-1 from two groups

Quarter	Mean stroke fraction D-1	Z	p
Con.G	0.1539(0.0496~0.2715)	-6.506	0
Exp.G	0.3566(0.2302~0.5124)		

Table 9 The Mann-Whitney U test result of Mean stroke fraction D0 and D-1 from two groups

Quarter	Mean stroke fraction D0 and D-1	Z	p
Con.G	0.9635(0.7759~1)	-2.201	.028
Exp.G	0.8745(0.7140~0.9903)		

Correspondingly, we processed another Mann–Whitney U test to figure out the difference between two groups in the mean lead time and stroke distribution. A statistically significant difference was evident between the two groups, $p < 0.01$ with a small effect size for both mean lead time ($r = 0.28$) and stroke distribution ($r = 0.26$).

Table 10 The Mann-Whitney U test result of Mean lead time from two groups

Quarter	Mean lead time	Z	p
Con.G	13.19(6.11~28.24)	-3.782	0
Exp.G	23.71(13.61~33.60)		

Table 11 The Mann-Whitney U test result of Stroke distribution from two groups

Quarter	Std stroke lead time	Z	p
Con.G	5.28(2.70~10.59)	-3.579	0
Exp.G	9.10(6.44~13.05)		

The mean stroke lead time is 18.71 hours in control group, less than 24 hours indicated that the majority of work was done in the last 24 hours before the day D0. In contrast, the mean stroke lead time in experimental group is 25.30 hours, which was larger than 24 hours, indicated that the students were not stacking all the work before the day D0, some work was done before the day D-1, which is the day we gave game points.

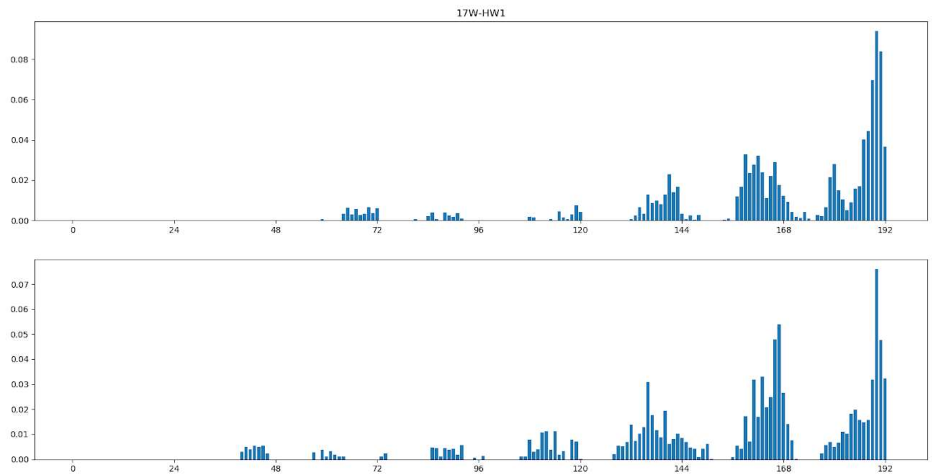


Figure 9 A visual of students' study activities for homework 1 between 2 groups

Figure 9 is a histogram showing the stroke fraction distribution of homework 1 for the control group and the experimental group. We can see that in the control group, there is only 1 peak, which is during day D0 (i.e., within 24 hours of the due date). The stroke fraction before day D0 is significantly lower than the stroke fraction on day D0. In the experimental group, with the gamification, the stroke fraction has 2 peaks, one on day D0 and one on day D-1 (i.e., between 48 and 24 hours of the due date).

Table 12 describes the fraction of strokes complete by day D-1 for both groups and for all nine assignments. The table also describes lead time for completing half of the work for each assignment. The values are averaged across all students in the experimental and control groups respectively. It is clear that students in the experimental group completed a larger fraction of their work by day D-1 than did the students in the control group.

Likewise, students in the experimental group had a large average lead time for completing the first half of an assignment compared to the students in the control group. Figure 10 provides a move visual comparison of the homework behavior of the two groups. Each plot shows the cumulative fraction of work completed for both the experimental and control groups for a particular assignment. For most assignments, the average effort by the experimental group significantly leads the effort of the control group. Figure 11 shows the cumulative homework effort for both groups averaged across all nine assignments. This figure, too, makes it clear that the effort by the experimental group significantly leads the effort of the control group. In the experimental group, students on average finished 57.13% of the total strokes by day D-1. In the control group, students on average finished 36.58% of total strokes by D-1.

Table 12 Total stroke fraction by D-1 and lead time to complete 50% of the work

	Control Group		Experimental group	
Homework assignment	**Average stroke fraction by D-1**	**Average lead time to complete 50% of work**	**Average stroke fraction by D-1**	**Average lead time to complete 50% of work**
1	0.475681	25.85	0.65744	39.03
2	0.425489	25.36	0.451662	21.00
3	0.426592	20.34	0.493695	24.97
4	0.358591	18.5	0.548732	27.54
5	0.298594	15.44	0.667777	28.57
6	0.254557	13.24	0.500144	21.54
7	0.387537	18.89	0.630009	29.01
8	0.248701	13.66	0.548241	23.52
9	0.416815	22.62	0.643839	24.41

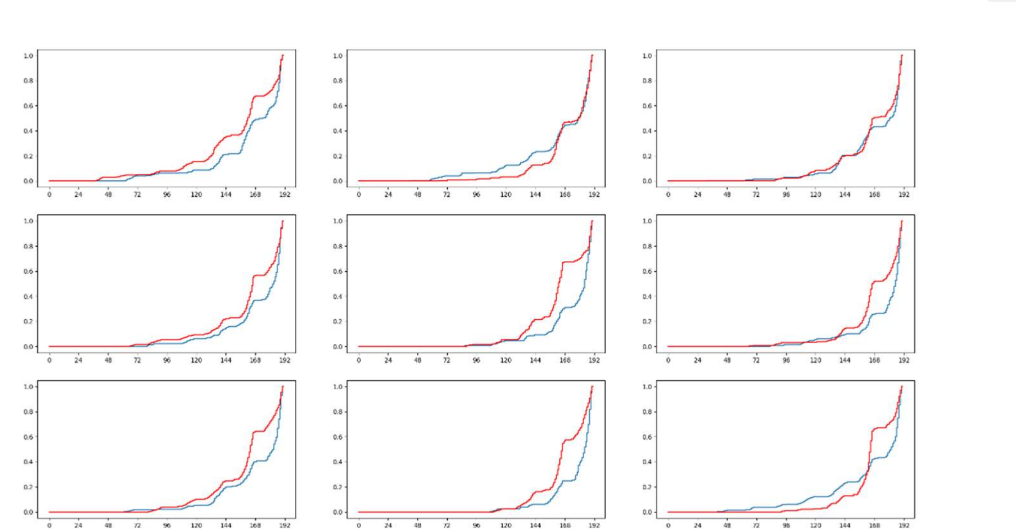


Figure 10 A visual of students' overall study activities for 9 homework assignments between 2 groups.

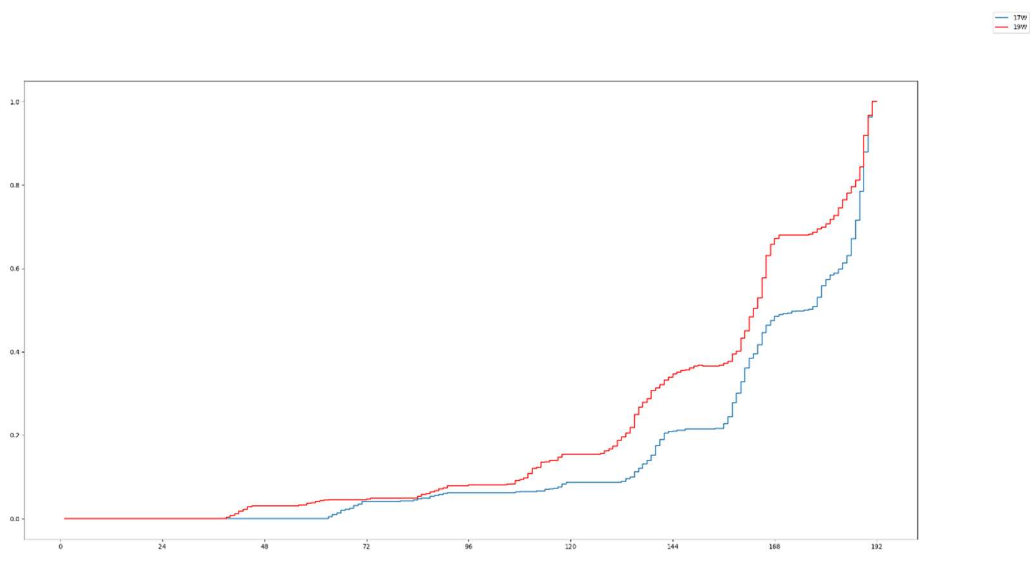


Figure 11 Accumulation figure for the average stroke fraction of all 9 homework

Learning Outcomes

To evaluate the learning outcomes, we use final exam scores and the post-test as the main measurements. Our hypothesis 2a indicated that the gamification group will achieve better learning outcomes. According to the Table, the average final grade of the experimental group is slightly lower than that of the control group, but the post-test results of the experimental group is better than results of the control group.

Table 13 Descriptive statistics of final exam score from two groups

Quarter	N	Mean	Std. Deviation	Shapiro-Wilk Sig.
Con.G	96	60.36	22.69	<0.01
Exp.G	91	58.02	20.34	.603

Table 14 Descriptive statistics of post-test from two groups

Quarter	N	Mean	Std. Deviation	Shapiro-Wilk Sig.
Con.G	84	16.88	6.769	.017
Exp.G	84	17.62	6.032	.027

We continue conducting Mann–Whitney U test to explore the difference between the control group and the experimental group in the final exam score and post-test score. For final exam, there was no statistically significant difference between two groups, $p = 0.170$ with a small effect size ($r = 0.10$). For post-test, there was no statistically significant difference between the two groups, $p = 0.473$ with a small effect size ($r = 0.05$). The results indicate that the learning outcomes did not improve because of gamification. In fact, there was no difference in learning outcomes between the two groups.

Table 15 The Mann-Whitney U test result of Final exam grade from two groups

Quarter	Final exam	Z	p
Con.G	65.56(51.11~75.56)	-1.373	.170
Exp.G	60.00(43.70~72.59)		

Table 16 The Mann-Whitney U test result of Post-test from two groups

Quarter	Post-test	Z	p
Con.G	16.00(12.00~22.75)	-.718	.473
Exp.G	17.00(13.00~22.00)		

Our pre-test results indicated that, on average, the students from both groups have the same background knowledge before they entered this class. However, to consider the effects of prior knowledge for individual students, we employed an Analysis of Covariance (ANCOVA). Final exam score was dependent variable, and pre-test score was covariate.

The covariate, pre-test score, was significantly related to the final score, $F(1, 155) = 31.40$, $p < 0.01$. There was not a significant effect of final score after controlling for the pre-test score, $F(1, 155) = 2.48$, $p = 0.117$. This result is consistent with hypothesis 2b: adding gamification does not improve learning outcomes.

According to the Table 17, the total strokes, the Mean stroke fraction D0, and the Mean stroke fraction D0 and D-1 are significantly correlated to the final grade, which is consistent with previous research. From this, we conclude that the students who put more effort into homework (Total Strokes) will get a better grade. Likewise, we conclude that students who commonly wait until the last day Mean stroke fraction D0 and Mean stroke fraction D0 and D-1) to start working on homework assignments will get worse grade.

Table 17 Correlation between final grade and each measurement for all students and each year separately

Measure	All Students	Con.G	Exp.G
Total Strokes	.317**	.461**	.207*
Mean stroke fraction D0	-.309**	-.394**	-.317**
Mean stroke fraction D-1	.211**	.343**	.208*
Mean stroke fraction D0 and D-1	-.249**	-.296**	-.213*
Stroke Lead Time	.327**	.340**	.355**
Std stroke lead time	.111	.203*	.045

****.Correlation is significant at 0.01 level**
***.Correlation is significant at 0.05 level**

The stroke lead time also has a significant correlation with the final grade. This shows that students who start their assignments early do better. This result confirms our hypothesis 3a. The stroke distribution is significantly positively correlated with the learning outcome in the control group, but not significantly correlated with the learning outcome in the experimental group ($p>0.05$). Adding gamification forced students to distribute their to get game points. As an extrinsic factor, this is only changing students' study behavior but not improving their motivation to understand the material. This is the hidden cost of gamification. Likewise, the correlation of Mean stroke fraction D-1 with final exam score is larger for the non-gamified (control) group than for the gamified (experimental) groups. This again shows that the games led to a changes in behavior which did not translate into a change in learning outcomes.

Effect of Gamification

Here we examine the extent to which students in the experimental group participated in the games. Figure 12 shows the number of students who have achieved particular games levels by the completion of each homework assignment. Homework 3 is the first assignment for which students could reach level 2. 37% students did achieve level 2 by homework 3. Likewise, homework 6 is the first assignment for which students could reach level 3. 26% of students did so. At the completion of homework 9 (the last assignment), only 42% students reached the final level, level 3. 32% students never achieved level 2. Thus, not all students participated fully in the games.

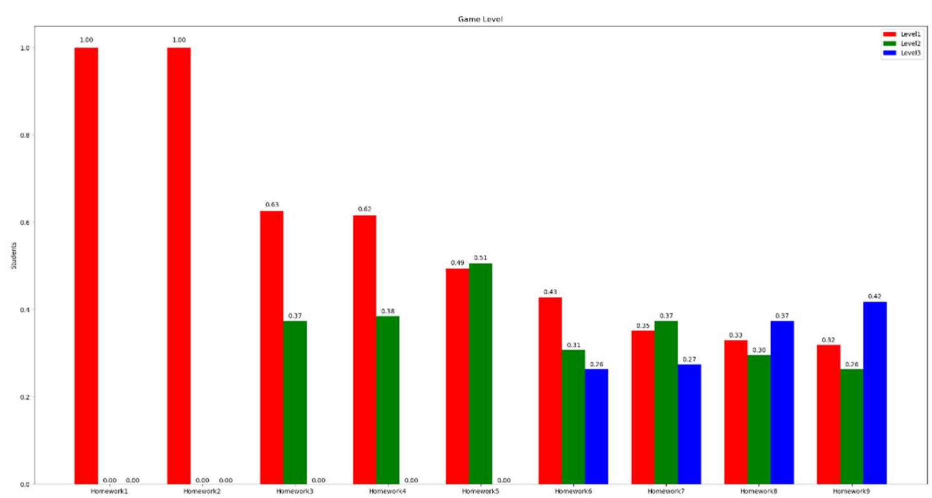


Figure 12 Early homework assignments game level record

Here we examine the relationship between game participation and learning outcomes. Table 18 shows the correlation between total game points achieved and final exam score, homework game points achieved and final exam score, and homework game points achieved and quiz score. The correlations between total game points achieved and final exam score is 0.226, which indicates that gamification corresponded to better overall course performance. The correlation between homework game points achieved and final exam score is 0.281, which indicates participation in the homework games corresponded to better performance. Similarly, the correlation between homework game points achieved and quiz score is 0.362, which again indicates participation in the homework games corresponded to better performance. (All three correlations are significant.) These results suggest that it is possible that participation in the games did cause better performance. However, because on average, there was no difference in outcomes between the two groups, this could also mean that only those students who would otherwise do well in the course are the ones who participated in the games.

Table 18 The correlation between gamification and learning outcomes

Relationship	Correlations	p
Total game points and final exam score	0.226*	0.031
Homework game points and final exam score	0.281**	0.007
Homework game points and quiz score	0.362**	<0.001

Chapter 6 Conclusions

Empirical Contributions

In this research we examined the effect of adding gamification to an engineering course. Our hypothesis was that students in the gamification group will achieve a greater mean lead time than students in the non-gamified group (hypothesis 1) and will achieve a higher score on the final exam which taps the material contained in the weekly assignments (hypothesis 2a), and there will be a substantial correlation between mean lead time and final exam score for both groups (hypothesis 3a). According to our results, the mean lead time was 25.30 hours in experimental group and 18.71 in control group, which confirm our hypothesis 1. With the results of Mann-Whitney U test and ANCOVA, there was not a significant difference of final exam score between the experimental group and control group, which disprove our hypothesis 2a and confirm hypothesis 2b that students in the gamification group will not lead to better learning outcomes. According to table 17, the result confirms our hypothesis 3a, there will be a substantial correlation between mean lead time and final exam score for both groups. For both groups, doing work early did correlate positively with outcomes. This suggests that this behavior -- distributing homework effort -- is beneficial. However, for most measures, the correlations were stronger for the control group than for the experimental group. This suggests that artificially motiving students to engage in this behavior diminishes its benefits. Students who are naturally motivated to employ effective strategies see a benefit, while those who are externally motivated may not see the same benefit. In short, adding external motivation to engage in distributed study behavior appears to water down the overall effect of this behavior. Based on that, we confirm a new hypothesis 3c, that gamification will diminish the relationship between distributed study behavior and final exam score.

Bibliography

[1]K. Rawson, T. F. Stahovich, and R. E. Mayer, "Homework and achievement: Using smartpen technology to find the connection.," *Journal of Educational Psychology*, vol. 109, no. 2, pp. 208–219, Feb. 2017, doi: 10.1037/edu0000130.

[2]L. Jones and G. M. Ruch, "Achievement as Affected by Amount of Time Spent in Study.," in *The Twenty-Seventh Yearbook of the National Society for the Study of Education: Nature and Nurture, Part II--Their Influence upon Achievement.*, Bloomington: Public School Publishing Co, pp. 131–134. Accessed: Aug. 31, 2022. [Online]. Available: http://dx.doi.org/10.1037/11368-009

[3]M. Credé and N. R. Kuncel, "Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance," *Perspectives on Psychological Science*, vol. 3, no. 6, pp. 425–453, Nov. 2008, doi: 10.1111/j.1745-6924.2008.00089.x.

[4]J. G. Gyllen, T. F. Stahovich, R. E. Mayer, N. Entezari, and A. Darvishzadeh, "Priming productive study strategies with preparatory quizzes in an engineering course," *Applied Cognitive Psychology*, vol. 35, no. 1, pp. 169–180, Oct. 2020, doi: 10.1002/acp.3750.

[5]J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques," *Psychological Science in the Public Interest*, vol. 14, no. 1, pp. 4–58, Jan. 2013, doi: 10.1177/1529100612453266.

[6]L. Fiorella and R. E. Mayer, *Learning as a generative activity*. Cambridge: Cambridge University Press, 2015. Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.1017/cbo9781107707085

[7]K. A. Kiewra, *SOAR to college success and beyond*. Cognella Academic Publishing, 2021.

[8]T. Miyatsu, K. Nguyen, and M. A. McDaniel, "Five popular study strategies: Their pitfalls and optimal implementations," *Perspectives on Psychological Science*, vol. 13, no. 3, pp. 390–407, May 2018, doi: 10.1177/1745691617710510.

[9]O. A. B. Hassan, "Learning theories and assessment methodologies – an engineering educational perspective," *European Journal of Engineering Education*, vol. 36, no. 4, pp. 327–339, Aug. 2011, doi: 10.1080/03043797.2011.591486.

[10]D. H. Schunk, *Learning Theories: An Educational Perspective*. Pearson, 2019.

[11]H. M. Cooper, *Homework*. Addison-Wesley Longman Limited, 1989.

[12]H. Cooper, *The battle over homework: Common ground for administrators, teachers, and parents*. Corwin, 2001.

[13]R. E. Mayer, *Applying the science of learning*. Prentice Hall, 2011.

[14]H. Cooper, J. C. Robinson, and E. A. Patall, "Does homework improve academic achievement? A synthesis of research, 1987–2003," *Review of Educational Research*, vol. 76, no. 1, pp. 1–62, Mar. 2006, doi: 10.3102/00346543076001001.

[15]M. Carter, "Visible learning: A synthesis of over 800 meta-analyses relating to achievement," *Educational Psychology*, vol. 29, no. 7, pp. 867–869, Nov. 2009, doi: 10.1080/01443410903415150.

[16]J. Xu, "Homework and academic achievement," *Visible Learning Guide to Student Achievement*, pp. 121–125, Oct. 2019, doi: 10.4324/9781351257848-19.

[17]J. Herold, T. Stahovich, H. Lin, and R. Calfee, "The effectiveness of 'pencasts' as an instructional medium." Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.18260/1-2--18518

[18]S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness," 2011. Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.1145/2181037.2181040

[19]C. Dichev, D. Dicheva, G. Angelova, and G. Agre, "From gamification to gameful design and gameful experience in learning," *Cybernetics and Information Technologies*, vol. 14, no. 4, pp. 80–100, Jan. 2015, doi: 10.1515/cait-2014-0007.

[20]C. Dichev and D. Dicheva, "Gamifying education: What is known, what is believed and what remains uncertain: A critical review," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, Feb. 2017, doi: 10.1186/s41239-017-0042-5.

[21]G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Improving participation and learning with gamification," Oct. 2013. Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.1145/2583008.2583010

[22]K. F. Hew, B. Huang, K. W. S. Chu, and D. K. W. Chiu, "Engaging Asian students through game mechanics: Findings from two experiment studies," *Computers & Education*, vol. 92–93, pp. 221–236, Jan. 2016, doi: 10.1016/j.compedu.2015.10.010.

[23]K. M. Kapp, *The Gamification of Learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons, 2012.

[24]R. E. Mayer, *Computer games for learning*. The MIT Press, 2014. Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.7551/mitpress/9427.001.0001

[25]E. L. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Boston, MA: Springer US, 1985. Accessed: Jun. 10, 2022. [Online]. Available: http://dx.doi.org/10.1007/978-1-4899-2271-7

[26]M. R. Lepper and D. Greene, *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. Psychology Press, 2015.

[27]M. R. Lepper, J. Henderlong, and I. Gingras, “Understanding the effects of extrinsic rewards on intrinsic motivation—Uses and abuses of meta-anlysis: Comment on Deci, Koestner, and Ryan (1999).,” *Psychological Bulletin*, vol. 125, no. 6, pp. 669–676, 1999, doi: 10.1037/0033-2909.125.6.669.

[28]D. Hestenes, M. Wells, and G. Swackhamer, “Force concept inventory,” *The Physics Teacher*, vol. 30, no. 3, pp. 141–158, Mar. 1992, doi: 10.1119/1.2343497.

[29]P. S. Steif and J. A. Dantzler, “A Statics Concept Inventory: Development and Psychometric Analysis,” *Journal of Engineering Education*, vol. 94, no. 4, pp. 363–371, Oct. 2005, doi: 10.1002/j.2168-9830.2005.tb00864.x.