USING A GAMIFIED POINTS-BASED GRADING SYSTEM IN TECHNOLOGY COURSES FOR PRE-SERVICE TEACHERS

A Doctoral Thesis Presented to the
Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Philosophy

By

George Zhao

August 2019

Acknowledgment

I want to thank Dr. Sara McNeil, Dr. Bernard Robin, Dr. Susie Gronseth, Dr. Jie Zhang, and Dr. Erwin Handoko for helping me complete my dissertation, for everything that I have learned and accomplished in my doctoral career, and for being the best. I want to thank my parents, Dr. Heping Zhao and Dr. Yan Joy Zhao for conceiving me, raising me, and for supporting me on my life's journey. I want to thank my Sister, Brother-in-Law, and Niece for being great siblings. Finally, I would like to thank myself for finally finishing my studies and completing a long-sought life objective.

USING A GAMIFIED POINTS-BASED GRADING SYSTEM IN TECHNOLOGY COURSES FOR PRE-SERVICE TEACHERS

An Abstract
of a Doctoral Thesis Presented to the
Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Philosophy

By

George Zhao

August 2019

Abstract

Background: Current research in grading practices suggest that there are connections between effective grading practices and students' motivation. Researchers have argued that traditional grading systems, such as letter grades, are not indicative of students' real abilities and hinder authentic learning by punishing students for their work, rather than rewarding them. Alternative grading systems that are points-based, such as those used in games, have the potential to motivate students and foster higher-order thinking. Gamification is a relatively new field in education, and there have been few research studies on how educators can best use game elements in instruction. Most studies on gamification in educational settings have tested multiple game mechanics at the same time with mostly positive results. However, individual elements of game mechanics have not been adequately studied in isolation. As a result, it is difficult for educators to make informed decisions about which, if any, game elements to incorporate in their courses. **Purpose:** The purpose of this study was to investigate the effects of a points-based system on students' perceptions of their motivation and their class performance and to compare those results to students in a traditional letter-graded course. **Research Questions**: The research questions were: 1) How does a class taught using a points-based grading system compare to a class taught using a traditional letter grading system in terms of intrinsic motivation? 2) How does a class taught using a points-based grading system compare to a class taught using a traditional letter grading system in terms of class performance? 3) How do students perceive their grade at the beginning of a course before the submission of any assignments? **Methods:** Four sections of an undergraduate course served as the participants. Two sections of the class were randomly assigned as the control group in which a traditional letter-grade system was used to display progress

in the course, and the other two class sections formed the treatment group in which a points-based system was used. At the beginning and the end of the semester, each participant's intrinsic motivation level was measured using the Intrinsic Motivation Inventory. Information was also collected about how familiar the participants were of games and their ultimate opinions of the points-based grading system. Lastly, the final grades of all participants were collected at the end of the semester. **Results:** Analysis of the students' post-semester motivation levels were performed using a one-way multivariate analysis of variance (MANOVA). To compare treatment and control group differences in final grades, independent sample T-Test was used. The results indicated that participants' motivation and class performance was largely unchanged by using the points-based system when compared to the traditional letter-grading system. Students reported that they felt mostly neutral about the points-based grading system, although most preferred it over traditional letter grading. **Conclusions:** There is insufficient empirical evidence to begin gamifying education. Further research is needed to identify whether or not this type of game mechanic would be useful in the classroom.

Table of Contents

Chapter I Introduction 1
- Problem Statement 3
- Research Questions 4
- Significance 5
- Definitions 5
- Summary 6

Chapter II Literature Review 7
- Traditional Letter Grading 7
- Gamification Background 11
- Gamification 12
- Pointsification 13
- Individual Game Mechanics 15
  - Badges 15
  - Leaderboards 18
  - Points 20
  - Isolated game mechanics 24
- Summary 25

Chapter III Method 27
- Research Questions and Hypotheses 27
- Participants 28
- Research Design 29
- Measurement 33
- Data Analysis 36
- Summary 38

Chapter IV Results 39
- Participant Demographics 39
- Game Experience 41
- Intrinsic Motivation 43
  - Intrinsic motivation inventory 44
  - Pre-test MANOVA 44
  - Post-test MANOVA 46
- Class Performance 48
  - Independent samples t-test 49
- Grade Perception 51
- Points Impressions 52
- Summary 53

Chapter V Conclusion 54
- Overview of Results 54
- Discussion 55
- Points Impressions 57
- Game Familiarity 61
- Combined Game Mechanics 62
- Relation to Other Research 63
- Future Research 65

Conclusion 67
References 70
Appendix A Questionnaires 79
Appendix B Consent to Take Part in a Human Research Study Materials 88
Appendix C Data Analysis Assumptions Testing 97
Pre-Test Assumptions Analyses 98
Post-Test Assumptions Analyses 104
Class Performance Assumptions Analyses 114
Appendix D Points-Based Grading Impressions Figures and Tables 118
Appendix E IRB Approval 136

List of Tables

1. Badges created in ClassBadges.com used by de Rocha Seixas et al…….....17
2. Frequency table for Game Experience……42
3. Pre-test descriptive statistics……...45
4. Multivariate tests for the pre-test MANOVA……46
5. Descriptive statistics for the post-test IMI subscales……..47
6. Multivariate tests for the post-test IMI subscales……48
7. Descriptive statistics for class performance……50
8. T-test comparing the Class Performance of the Control and Treatment groups……..51
9. Frequencies table for participant's perceived grade at the start of the class……52
10. Descriptive statistics for the post-test points-based grading impressions questions….52
11. Results of tests of between-subjects effects for the post-test IMI subscales……..57
12. Pearson correlation table for the pre-test IMI subscales……..102
13. Pearson's test for homogeneity of variances……104
14. Pearson's r correlation matrix table of competence subscales……109
15. Levene's Test of Equality of Error Variances for post-test IMI Subscales…….111
16. Homogeneity of variances testing for the post-test IMI subscales……..112
17. Results of tests of between-subjects effects……113
18. Descriptive statistics for class performance……114
19. Levene's test for equality of variances for the class performance variable…….117
20. Descriptive statistics for the post-test points-based grading impressions questions…120
21. Descriptive statistics for all points questions combined……..121
22. Earning points and levels motivated me to learn more than just traditional letter grades…...…….....123
23. I prefer a traditional letter grading system over the points and levels system used in this particular class…….125
24. I felt more engaged in this particular class using points and levels than in other classes using traditional letter grades…….127
25. I felt that I was rewarded more for my performance on activities in this particular class……129
26. I felt that points and levels provided milestones that I was encouraged to reach…….131
27. I found the points grading system frustrating…….133
28. The points grading system was easy to understand…….135

List of Figures

1. A portion of a leaderboard used in a graduate level instructional design course at the University of Houston (2017). 19
2. An example of a simple point system that could be applied to a course with a 200-point total.. 22
3. Assignment Point distribution and level thresholds for the treatment class sections. 31
4. Example Pointcard with section descriptions. 32
5. Histogram of participant ages……..……...41
6. Bar chat for the game experience question on the pre-test questionnaire……..…….43
7. Pre-Test Control Group Questionnaire 82
8. Pre-Test Treatment Questionnaire 84
9. Post-Test Control Questionnaire..... 85
10. Post-Test Treatment Group Questionnaire 87
11. Boxplots of the pre-test IMI subscales: Interest / Enjoyment, Perceived Competence, Effort / Importance, and Pressure Tension. 99
12. Histogram graph of the Pre-Test IMI Subscales with a normal curve overlay..... 101
13. Scatterplot matrix of the pre-test IMI subscales: interest/enjoyment, perceived competence, effort/importance, and pressure/tension..... 103
14. Boxplots for assessing outliers for all of the post-test IMI subscales..... 106
15. Distribution histograms for the IMI post-test subscales. 108
16. Scatterplot matrix of the Post-test IMI subscales: interest/enjoyment, perceived competence, effort/importance, and pressure/tension. 110
17. Boxplots to assess outliers for the final grade variable..... 115
18. Histograms of the class performance variable to determine distribution normality. 116
19. Line graph of the combined points question scores..... 121
20. Bar graph of the results of points question 1 122
21. Bar graph of the results of points question 2. 124
22. Bar graph of the results of points question 3 126
23. Bar graph of the results of points question 4 128
24. Bar graph of the results of points question 5 130
25. Bar graph of the results of points question 6 132
26. Bar graph of the results of points question 7 134

Chapter I

Introduction

Literature suggests that the academic grading system in America is problematic. Specifically, the problem stems from the traditional top-down letter grading system which has been the standard in American educational institutions since 1897 (Durm, 1993). While the letter grading scale is great as a categorization tool, the possible issues with this system lie within the potentially negative psychological impact that it has on learners subjected to it (Dueck, 2014; Kohn, 1999). The rising sentiment about letter grading is that it is fear-based, demotivational, and not reflective of the professional world for which schools are preparing youths (Kohn, 1999; Schinske & Tanner, 2014; Zichermann, 2012). It would be abnormal in a professional environment for new employees to begin with the highest position available and only get demoted as time passes. Instead, most companies take a bottom-up approach, one in which employees begin at the bottom and work their way up through the ranks as their experience and expertise accumulate. The majority of students, when they enter a new class for the first time, before any assignments have been completed, believe that they have an A+ in that class (Sheldon, 2011). As they begin to submit assignments, each mistake or failure that they make on their assignments, no matter how slight, only serves to reduce their grade. The best that a student can hope to accomplish in a letter-graded environment is to maintain their grade with no real opportunity to improve, with the notable exception of extra credit assignments. Students, as they learn and complete assignments, are not rewarded; they are at best not punished.

There is a new grading system that has been proposed as part of the influx of *gamification* practices in education generally known as points-based grading (also known as XP or experience points) (Attali & Arieli-Attali, 2015; Papp, 2017). A points-based grading system is a bottom-up system that has been refined by decades of video game development to draw from a human being's powerful, natural desire for progress and efficiency (Deterding, 2012; Sheldon, 2011). Furthermore, experience points have already been shown to be effective in businesses around the world to inspire customer loyalty and employee engagement with highly publicized success (Hamari & Koivisto, 2015; Kim, 2015; Robson, Plangger, Kietzmann, McCarthy, & Pitt, 2015; Seaborn & Fels, 2015; Zichermann & Linder, 2013). The primary difference between the letter grading system and the points-based grading system lies within the student's perception of their grade as they progress through a course. With the points-based grading system, students are told when they begin each class that they have a failing grade, and with each subsequent assignment submission, they will gain points-based on their performance, effectively raising their grade as they progress. As a result, the learner receives a motivational boost with each reception of points, theoretically encouraging them to advance and continue building up their points – ultimately resulting in a course completion grade (Kim, 2015; Kuo & Chuang, 2016).

There have been studies on gamified classrooms which include a points-based grading system in addition to other game mechanics (Flores, 2015; Hanus & Fox, 2015; Kuo & Chuang, 2016; da Rocha Seixas, Gomes, & Filho, 2016; Sheldon, 2011). However, while these preliminary studies have shown mostly promising results, there remains a dearth of authentic information on the topic regarding how it pertains to

education (Cohen, 2011; Hamari, 2015; Hanus & Fox, 2015; Landers & Armstrong, 2015). Furthermore, there are very few studies that separate and investigate the elements of gamification individually (Attali & Arieli-Attali, 2015; Chen, Burton, Mihaela, & Whittinghill, 2015; Mekler, Bruhlmann, Tuch, & Opwis, 2017). If education is truly moving towards a gamified approach, it becomes critical to isolate the various elements of gamification and determine which aspects of gamification are worthy of implementation. The purpose of the study was to isolate the game element of points from the overall gamified education approach by applying it within a real classroom setting and to measure the impact that this grading system had on overall class performance. Additionally, this study contributes to the growing literature on educational gamification and may help to determine whether a points-based grading system should or should not be implemented in our schools.

Problem Statement

Research suggests that the traditional top-down letter grading scale used in America, as well as many other countries in the world, is potentially harming student motivation by emphasizing punishment as opposed to rewarding student achievement. A points-based grading system has been proposed as an alternative grading system that would remove this potential barrier to student achievement. Under this grading system, students will instead begin a new class with zero points and, as they submit assignments for the course, they will be awarded points which serve to raise their grade. Points-based grading is, therefore, a shift in how students perceive their advancement in a course. This change in perception is intended to provide achievement-based motivation through positive reinforcement rather than fear-based motivation through punishment (Skinner,

1953). Furthermore, the research done to date on gamification fails to investigate the effects of individual game elements in isolation. This study was focused on identifying the impact of just a points-based grading system in a classroom setting.

Another problem is that the body of literature on the topic of educational gamification, while growing, exhibits a lack of focused research on the many elements of gamification (Hamari, 2015; Hanus & Fox, 2015). The majority of studies performed on educational gamification test the effectiveness of a combination of many game mechanics in the classroom. However, studies on the individual game mechanics associated with gamification have not yet been adequately measured in isolation. This study measured the impact of one game mechanic, points, and has contributed a semester-long study to the small but growing body of literature on points-based grading in isolation.

Research Questions

Given the problem statement, the following research questions were posed:

(1) How do undergraduate educational technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of intrinsic motivation?

(2) How do undergraduate educational technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of class performance?

(3) How do students in undergraduate educational technology courses for preservice teachers perceive their grade at the beginning of the course prior to receiving grades for any class assignments?

Significance

This research, when combined with the ever-growing body of work on gamification in education, can bring educators one step closer to understanding how best to apply game mechanics to educational settings. Among all of the game mechanics that compose gamification, points may hold the greatest promise since it has the potential to improve our grading systems. If points-based grading has the capacity to positively influence student intrinsic motivation, then it may be able to transform education. It is towards this possible outcome that this research has investigated points-based grading.

Definitions

Educational gamification. Refers to the use of game design elements in academic settings (Deterding, Dixon, Khaled, & Nacke, 2011).

Game element. Used interchangeably with the term "game mechanic."

Game mechanic. Rule-based systems found within games, especially video games, which dictate or influence how game players interact with the game world.

Gamification. Refers to "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 1).

IMI. An acronym for the Intrinsic Motivation Inventory questionnaire developed by Deci and Ryan (2005)

Pointcard. The personalized document used in this study to deliver a points-based assessment to participants in the treatment group.

Points-based grading system. Refers to a game mechanic used in gamification in which a student's graded assignments result in points being given to the student.

Throughout the course, the points accumulate and at the end of the course the amount of points determines the grade. Points are also known as XP, or experience.

Pointstification. A subcategory of gamification that consists of the three most popular gamification mechanics: badges, points, and leaderboards.

Traditional letter grading system. Refers to the standard grading system used in America since 1987 (Durm, 1993). Utilizes letters from the Roman alphabet (A, B, C…) to represent a student's level of success in a course or on an assignment.

Summary

The letter grading system utilized in America may be harmful to student motivation. Educational gamification utilizes a point-based grading system that has the potential to improve the motivation of students. Unfortunately, research performed on the individual game mechanics in the gamification process is sparse and is insufficient as evidence for the effectiveness of points-based grading. This research project studied just the game mechanic of points so that educators can have more insight into this mechanic's potential in the classroom.

Chapter II

Literature Review

This chapter will cover studies that are associated with and have influenced this project. The chapter begins with studies that have analyzed the effects of the American letter grading system on student motivation and performance before tackling gamification. This chapter then discusses research that has been performed on educational uses of gamification and highlights potential problems with the current state of educational research on gamification. This chapter also breaks down the most popular game mechanics utilized in gamification and analyzes the research performed on each mechanic individually.

Traditional Letter Grading

Prior to the implementation of Roman letters as measures of course performance, which has been the standard in American education for over a century, feedback was given regarding class performance without actually assigning grades (Durm, 1993). This would change as enrollment increased throughout the years, which resulted in the need for a standardized grading system in order to be able to categorize and aggregate student achievement (Marzano, 2000). Universities began assigning grades using a four-point scale, which later would evolve into the familiar 100-point scale (Tocci, 2010). In the 1850s, the practice of using the 100-point scale as a percentage-based grade began, and in 1897, the first evidence of using letters to denote a range of the 100-point grading point scale was first developed into the system with which Americans are now familiar (Marzano, 2000).

Grades served to satisfy a number of needs for educational organizations. Due to the expansion of educational institutions and the swelling of student numbers, it became necessary to designate, label, and organize students by their academic rank (Callahan, 1962). Additionally, grades served to scientifically measure student mastery of course content through a nationwide academic standards system (Tocci, 2010). Grades are also used as a form of feedback to students on their ongoing class performance and also to teachers about their instructional performance, thereby influencing their instructional planning (Marzano, 2000). While, the organizational benefits of the grading systems for institutions are many, the rise of letter grading and grades, in general, were the result of a nation-wide need to categorize, rank, and organize students into a standardized system. While grades do benefit students as a form of distributing performance feedback, the development of our current grading system was born more to serve the needs of teachers and administrators, over the needs of the students subject to them.

Research since the advent of letter grading has consistently produced results that suggest that they have a detrimental impact on the intrinsic motivation of students. Traditional grading has been shown to produce performance-avoidance goals in students (Pulfrey, Buchs, & Butera, 2011). When students set performance-avoidance goals, they are motivated to perform in the class to avoid performing more poorly than their peers rather than in an effort to master the course content intrinsically, which in turn is associated with producing poor performance and low interest in the course content (Elliot & Church, 1997). In contrast, in the same study, the researchers discovered that students who were not graded demonstrated performance-approach goals. Performance-approach goals generate motivation in the desire to achieve positive outcomes and result in greater

investment in the content and better overall performance (Darnon, Harackiewicz, Butera, & Quiamzade, 2007). This study suggests that the current grading system may not be beneficial and could actually have some detrimental impacts on the academic success of students.

Researchers have attempted to explain why the grading system of the United States is potentially harmful to students (Maehr & Midgley, 1996). It has been shown in many studies that the traditional grading system negatively affects intrinsic motivation in students (Butler, 1998; Darnon et al., 2007; Kohn, 2011; Pulfrey et al., 2011). The root of the problem seems to stem from the top-down nature of our grading system. In the United States, the majority of students perceive that they begin a new class with an A, effectively a 100% score in the class. As a result of this belief, assignment submissions could never raise a student's grade; it would only reduce their grade (extra credit assignments are an exception). For example, a student submits an excellent paper that solidly demonstrates her mastery of the course content with only minor errors. If she receives a score of 48 out of 50 points for the assignment, which most would consider being an outstanding grade, the student might be thrilled and may feel recognized for her mastery (Guskey, 1994). However, the student's happiness stems not from gaining 48 points, but instead from the relief that she avoided losing 48 points from her total class grade. In other words, she perceives the result of her assignment as a loss of only 2 points, and not as a gain of 48. The practice of submitting assignments then becomes a fear-based cycle of avoiding punishment. If they perceived that they instead gained points for their assignment, then it would be a form of positive-reinforcement or reward.

In the previous example, the punishment by loss of points was minor, and so the overall impact on the student would be negligible. However, in situations where a student performs poorly on an assignment and receives a significant loss of points, the consequences can become severe. In an ideal scenario, when a student performs poorly on an assignment, they should be motivated to increase their efforts in the future. However, studies have found that the inverse often occurs. Students who receive low grades often respond to the loss of points by removing themselves from learning, believing that they will be unable to recover their overall grade (Guskey, 2004). Furthermore, the impact of our traditional grading system has an overall negative impact on intrinsic motivation, where the grade becomes more important than learning (Kohn, 2011). The grading system furthermore encourages students to find the easiest path to an A, rather than encourage them to perform at the best of their ability for the sake of achievement (Maehr et al., 1996).

Since there are many studies on the subject of the harmful effects of grading on students, it seems like the logical choice would be to remove our traditional grading system from our academic institutions. Unfortunately, more than 76 million students enrolled in schools in the United States in 2017 and consequently, the original need for grades as a tool for categorizing and ranking continues, and it is more needed today than it was during its inception over a century ago (U.S. Census Bureau, 2018). Perhaps, removing grading altogether would be unreasonable and would add further burden on teachers. Instead, the best option may be to adjust our current grading system to alleviate some of these inherent detrimental issues by adopting elements from other assessment

systems that have found success outside of education such as gamified points-based grading.

Gamification Background

Two decades ago in the seminal article "Computers as Mindtools for Engaging Learners in Critical Thinking," Jonassen, Carr, and Yueh, (1998) stated that computer microworlds were "perhaps the ultimate examples of active learning environments" (p. 27). "Microworlds" was a blanket term defined within the article as computer applications that have the users navigate virtual environments such as in simulations and, more importantly, video games. However, during that time, the common perception of games was that they mainly appealed to and, therefore, mostly affect younger learners (Jonassen et al., 1998). As time passed and video games continued to increase in popularity and mainstream acceptance, this perception has been proven false and, furthermore, was likely not entirely accurate at the time the article was written. In 2016, the average gamer in America can be described as a middle-aged 35-year-old (Electronic Software Association, 2016). Approximately 63% of American households currently contain a resident who games regularly (ESA, 2016). Additionally, about 41% of gamers are female (ESA, 2016). It is heavily implied that gaming has the power to impact people regardless of their age and gender (ESA, 2016; Thomas & Brown, 2011).

Presently, the effectiveness of gaming as a tool for generating motivation and promoting learning has been well researched (Cheong, Filippou, & Cheong, 2014; Flores, 2015; Hamari & Koivisto, 2015; Shute, Ventura, & Ke, 2015; Thomas & Brown, 2011; Zichermann & Linder, 2013). While games themselves have been used to moderate success in the classroom through educational video games such as the *Oregon Trail* and

Where in the World is Carmen Sandiego, some researchers recognized that many of the mechanics and systems that the video game creators were using could be applied outside of the microworlds in which they were originally presented (Hamari, 2015; Renauld & Wagoner, 2011; Shute et al., 2015; Zichermann, 2012). This led to an important question for the field: How do educators harness the motivational power found within games for use to increase student engagement within our classrooms?

Gamification

The answer indicates *gamification*. Gamification is a concept that is new to education and still is being defined (Deterding et al., 2011; Hakulinen, Auvinen, & Korhonen, 2015). Currently, the most accepted definition of gamification is "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 1). Still, this definition continues to be challenged, with a range of modified and expanded interpretations of the term used in publications (Kim, 2015; Seaborn & Fels, 2015). Furthermore, additional labels and subcategories pertaining to the individual game mechanics found within gamification have recently developed, including *pointsification* and *exploitationware*, further complicating the field (Chen et al., 2015, Kim, 2015; Seaborn & Fels, 2015).

Even though *gamification* is a fairly recent term, game elements and mechanics have an extensive and lengthy history (Deterding, 2011; Urh, Vukovic, Jereb, & Pintar, 2015; Zichermann, 2012). Prior to its use in education, gamification was used effectively in business and marketing (Hamari, 2015; Robson et al., 2105; Zichermann, 2012; Zichermann & Linder, 2013). It was in these areas that it was discovered that game mechanics had the potential to make even the most mundane of tasks fun and engaging to

users (Thomas & Brown, 2011; Hamari, 2015; Zichermann, 2012). Evidence suggests "that merely labeling a task as 'play' or 'game' changes its perception and subsequent performance" (Deterding, 2011, p. 4). This is quite possible, considering the success of a game such as *Jeopardy*, which, when broken down, is essentially just a gamified quiz.

The way that games create such an engaging environment is through how they manifest intrinsic motivation within players (Banfield & Wilerson, 2014; Deterding, 2011; Deterding, 2012; Lapp, 2012; Kim, 2015). Extrinsic goals, such as incentives, are generally considered weak reinforcements that can only motivate in the short term and could become harmful in the long term (Benabous & Tirole, 2003). Extrinsic motivation in education would be the student's desire to obtain a good grade. Intrinsic goals, on the other hand, compels students to learn through their desire to progress, achieve, and satisfy curiosities (Banfield & Wilkerson, 2014). Intrinsic motivation in education would be the student's desire to learn because there are interested in the subject matter. Gamification creates a game structure within a traditional non-game environment. The game structure results in a system of attainable non-incentivizing goals which theoretically trigger students' intrinsic motivation by allowing students to concretely visualize how they are improving (Deterding, 2012; Hamari, 2015; Kapp, 2012; Kim, 2015; Kuo & Chuang, 2016; Zichermann, 2012). Many began to see that through gamification "everyday drudgery, dull learning experiences, and stressful tasks can be ameliorated with the application of game dynamics and mechanics" (Kim, 2015, p. 20).

Pointsification

So, what are the game elements that educators are considering for use in their classrooms? Most studies have focused on three primary game elements - points, badges,

and leaderboards (Hakulinen et al., 2015, Hamari, 2015, Kuo & Chuang, 2016, Robson et al., 2015, de Rocha Seixas et al., 2016). These three elements have often been combined into the gamification subcategory called *pointsification*, which will be the focus of the remainder of this literature review (Kim, 2015; Seaborn & Fels, 2014). There are, however, many more game mechanics, such as narratives, challenges, and avatars, which are also considered a part of classroom gamification (Chen et al., 2015; Kapp, 2012; Kuo & Chang, 2016). Points, badges, and leaderboards are possibly garnering the most attention, because unlike many of the other game elements, these mechanics do not require that educators change their instructional content for effects to be seen (Zichermann, 2012).

However, empirical research on gamification's effectiveness is rather limited (Cohen, 2011; Hamari, 2015; Hanus & Fox, 2015; Landers & Armstrong, 2015). Initial studies have shown mostly promising results with students exposed to gamification practices exhibiting greater intrinsic motivation (Banfield & Wilkerson, 2014; Chen et al., 2015; Cheong et al., 2014; Hakulinen et al., 2015; Kuo & Chuang, 2016; Landers & Armstrong, 2015; Marin, Lopez, & Maldonado, 2015; de Rocha Seixas et al., 2016; Turan, Avinc, Kara, & Goktas, 2016). However, not all research performed on gamification has produced positive results. For example, Hanus & Fox (2015) conducted a semester-long study on a gamified classroom of 80 adults and measured learning using motivational, psychological, and behavioral measures. Their results suggested that some gamification elements in the classroom may have possibly harmed the achievement of the students in comparison to a non-gamified course, specifically badges and leaderboards. These findings are surprising to gamification researchers and go against the majority of

empirical literature currently available. However, while this study involved game mechanics such as badges and leaderboards, a points-based grading system was not used. This leaves readers questioning whether or not the lack of a points system or one of the other game elements caused this harm.

Individual Game Mechanics

The great majority of gamification studies utilize a combination of game mechanics to test gamification's effectiveness. There are very few articles that have been committed to analyzing these game elements individually, allowing educators to understand which mechanics are truly beneficial and which are possible hindrances (Attali & Arieli-Attali, 2015; Chen et al., 2015; Mekler et al., 2015). As Hanus and Fox (2015) demonstrated, there is a need for these elements to be tested separately from the rest of the gamification elements. The following sections will each focus on a single element and will analyze some of the studies that attempted to isolate the effectiveness of that element. The last section looks at a study that has isolated and tested gamification elements separate from each other.

Badges. Perhaps, the most widely recognized and utilized game element being introduced in classrooms is the *badge* mechanic. Badges function as an aesthetic award for students when they reach academic milestones in their class. Typically, once a badge is obtained, the awarded student can display the badge either physically in the classroom or, more likely, on a shared online platform to be viewed by others as evidence of their mastery of certain topics. Badges applied in an instructional environment provide goals for students to aspire to, making it so that students set their performance expectations

higher (which will ultimately also increase their performance) and increasing student self-efficacy and satisfaction (Hamari, 2015, Mekler et al., 2017).

De Rocha Seixas et al. (2016) observed, interviewed, and gave questionnaires to 61 elementary students in a gamified geometry classroom in an eighth-year elementary school in Brazil (the equivalent of a Junior High seventh grade in the United States). The researchers utilized two free online gamification tools (which they refer to as *badging platforms*) to apply the badges, ClassDojo.com and ClassBadges.com. Upon completing specific learning objectives, students would be rewarded with a badge on the badging platform which would then be displayed on the class Facebook page. You can see examples of the badges that they used in the study in Table 1. They found that students wanted to be recognized for doing their class activities, not only by their peers but also by their teachers. The existence of the badge on their account demonstrated that their teacher knew their students' capabilities. They concluded that the gamification of the class had an overall positive effect on student engagement. The researchers primarily emphasized badges as being the game mechanic that most prominently played a role in this success. However, other game mechanics, such as points, leaderboards, and narrative, were also applied to this class, complicating the results.

Table 1

Badges Created in ClassBadges.com used by de Rocha Seixas et al

Badge	Title	Message
	*Lord of Calligraphy*	Congratulations! You dominate the legendary art of the ancient ages: calligraphy.
	*Lord of technical drawing line*	Wow! You know how to handle the drawing sacred weapons (Set Square, Compass, etc.). You know how to differentiate the drawing lines and your tasks are organized and clean.
	*Sense 8 – Geometric Thinking*	You've reached the highest level … You went beyond the other senses. You have a good abstract and geometric reasoning and can make connections between content studied.
	*Master of Triangles*	You understand well the properties of triangles.
	*Master of Angles*	You understand the properties of angles and know how to build them and their divisions.
	*Master of Quads*	You understand the properties of angles and know how to build them and their divisions.
	*I'm your biggest fan!*	Congratulations! You have won the most desired badge in the world e You've got all the other badges.

Note. From de Rocha Seixas et al. (2016). Effectiveness of gamification in the engagement of students. *Computers in Human Behavior, 58* (p. 55).

Other studies on badges include one performed by Hakulinen, et al. (2015), who experimented with using badges for 281 adult online learners. Using observation and student self-reports, the badged group demonstrated positive changes in their behaviors. However, similar to the study performed by de Rocha Seixas et al. (2016), the success of Hakulinen's study may not be exclusively the result of badges. The researchers admitted that they also included a gamified point-grading system that could have also contributed to the impact on the students.

Hamari (2015) tested badges on a peer-to-peer marketing service. While this study was not performed in an educational setting, this research is notable as it was a two-year longitudinal study with a sample size of nearly three thousand. Furthermore, Hamari measured the impact of only badges and no other element of gamification. The study discovered that the members of the group that was rewarded for their marketing efforts with badges had a dramatic increase in overall motivation and effort. Hamari's work is a strong indication that badges have legitimate motivational application. Still, he believed that more efforts by researchers should be made to study game mechanics in isolation, and he urged other researchers to follow in his footsteps.

Leaderboards. It is believed that competition, either against one's self or versus peers is one of the primary forces behind gamification (Banfield & Wilkerson, 2014; Kapp, 2012; Hanus & Fox, 2015). Nearly as prominent as badges, the game element of *leaderboards* emphasizes social competition and has also seen success in educational contexts. A leaderboard is a publicly posted list of the total grades (or total points) of all students in the class in order of best performing to worst. Oftentimes, the names of the students are hidden under aliases in order to avoid potential humiliation. The idea is that

by seeing one's position in the class, a student's natural competitive instincts will activate, resulting in a stronger performance. In Figure 1, you can see an example of a real leaderboard used in a gamified graduate course.

	Astronaut	Total Hours in Space	Level	Hours Needed to Reach Next Level
 snø	snø	214.2	Level 5	You are at the highest level
 Hayat	Hayat	212.5	Level 5	You are at the highest level
 Zaphod Beeblebrox	Zaphod Beeblebrox	211.9	Level 5	You are at the highest level
 Crunch	Crunch	211.0	Level 5	You are at the highest level
 Gigi	Gigi	210.4	Level 5	You are at the highest level
 Augustina	Augustina	206.0	Level 5	You are at the highest level

Figure 1. A portion of a leaderboard used in a graduate level instructional design course at the University of Houston (2017). This example also exhibits the game elements: avatars, narrative, and points. Adapted with permission from the University of Houston Learning, Design, and Technology Program Area Coordinator Dr. Sara McNeil.

Banfield and Wilkerson (2014) applied gamified leaderboards to a lesson used in two sections of a university level introductory computer networking course. One section was the gamified treatment group, and the other section was left unchanged as the control group. In the article, they described the leaderboard as the *scoreboard*, but conceptually they are the same. Ninety-six total students participated in this study. In both sections, the students had to individually complete an activity with seven objectives. However, in the gamified section, an ongoing, publicly-displayed leaderboard tracked each student's progress through the objectives, while they worked on the activity. Students were able to look at the leaderboards and see their progress in relation to other students in the class. Afterward, the students in both sections were interviewed regarding their intrinsic motivation, extrinsic motivation, and self-efficacy. Ninety-two percent of the treatment group was intrinsically motivated whereas only thirty percent of the control group was. Eight percent of the gamified group was extrinsically motivated, whereas the control group was seventy percent. Ninety percent of the gamified students also demonstrated high self-efficacy, whereas the control group was only twenty-eight percent. The results were strongly in favor of the gamified course. The leaderboard, in this case, was a great success, and the study was another good example of a single game element being tested in isolation.

Points. *Points* is the game mechanic that is most pertinent to this paper. Points are also referred to as experience points, experience, EXP, and XP, and they serve as a measure of how much a student has mastered the content of a course. Students in a gamified course begin the class with zero points and, as they complete assignments, they gain points based on their performance. Furthermore, most applications of points divide

the total possible points into different *levels*. Levels provide obtainable goals for students to reach and are also a quick reference for how well they are performing. For example, consider the simple point scale shown in Figure 2. In a class using this point scale, a new student would begin with 0 points as a Level 1 student in the course. After a month, the student has completed six 10-point assignments and has earned 53 total points. Since the teacher set the threshold for level advancement at 20 points per level, the student is now a Level 3 student in the course. She knows that she only needs seven more points to reach Level 4. Even though, at this moment, she has not yet reached a point total that would translate to a grade higher than F, because of the levels, she can see that she is already making progress and reaching learning objectives through the point-grading system.

Figure 2 shows that at the course's conclusion, points are translated into a letter grade. However, the psychological effect of gradually gaining points is what theoretically triggers students' motivation (Mekler et al., 2013; Zichermann, 2012). Unfortunately, whether this theoretical motivation is enough to result in a significant improvement in student performance remains unclear. Studies focused solely on a points-based grading system applied to an educational setting are extremely limited.

Levels	XP required	Letter Grade
Max Level	200xp	A+
Level 10	180xp	A
Level 9	160xp	B
Level 8	140xp	C
Level 7	120xp	D
Level 6	100xp	F
Level 5	80xp	F
Level 4	60xp	F
Level 3	40xp	F
Level 2	20xp	F
Level 1	0xp	F

Figure 2. An example of a simple point system that could be applied to a course with a 200-point total. In this example, the point thresholds for levels are equidistant. However, many point systems increase a student's point requirement to level up for each level that a student obtains.

Ahn, Johnsen, and Ball (2019) tested points on a three-day field study on children aged 9 to 13 (n = 67). In the study, the children were presented with a virtual dog on a monitor that would encourage the child to perform physical activity. In the treatment group (n = 39), the children performed physical activity tasks that earned points that they could then use to *pay* the dog to perform tricks. In the control group (n = 28), the digital dog had all of the tricks unlocked by default and would perform them automatically as soon as physical activity goals were met without the need for children to provide points. In both cases, if physical activity goals were not met, the digital dog would encourage them to exercise. At the end of the three-day experiment, the participants took a modified version of the Intrinsic Motivation Inventory (IMI). The results showed that the children in the points group performed significantly more physical activity on the 2nd

day of the experiment than the control group. However, the physical activity decreased sharply on the last day, far below that of the control group. Additionally, the children in the points condition exhibited less strenuous physical activities, effectively figuring out how to do the minimum required to obtain points. In terms of intrinsic motivation, the points group perceived a higher relatedness than the control group, which means that they developed a closer relationship with the digital dog (Ryan & Deci, 2000). However, there was no significant difference in the other IMI subscales. Ultimately, Ahn et al. (2019) felt that a three-day experimental period may have been too short to properly determine the effectiveness of points and that further research should be performed at longer durations.

Chen et al. (2012) used a gamification system called Cogent with 32 students. One of the key features of the Cogent system is the "virtual economy" that it creates within the classroom. Assignments were not worth a grade, but instead, students were given a fictional virtual currency for completing tasks. At the end of the semester, the students must *purchase* their passing grade from their professor using the virtual currency that they have accumulated. In practice, this currency parallels a gamified points system with one difference. In addition to featuring a point mechanic, the use of fantasy money meant that this approach also incorporated the game mechanic of narrative (Kapp, 2012; Sheldon, 2011). However, Chen et al. argued that Cogent's use of the narrative element is what makes this gamified system meaningful whereas a pointsified class without this element would not be. Using interviews and focus groups, Chen at al. found that their Cogent system succeeded in motivating and engaging students, though it is impossible to determine if it would have been successful without the inclusion of a game narrative.

Attali and Arieli-Attali (2015) tested the effects of points that were isolated from the other elements of gamification. This study applied a point score onto the computer screen of a mathematics exam taken by 693 middle school students. As the middle school students took the exam, they would see the points on the screen increase as they correctly answered questions. The results of this study showed that the application of points had no significant effect on student performance. This is not surprising, as this study has a number of serious flaws, at least in relation to how experience points are believed to take effect. The main issue is that the points were assigned to just a single assignment. This does not provide motivation-associated progress and growth to manifest (Kapp, 2012; Zichermann, 2012). Instead, since it was only a single assignment, the point score was realistically perceived no differently to those students than the letter grade they would have received. A more robust study on the effectiveness of a points-based grading system would need to be a longitudinal study, one that spans the length of an entire course at minimum, so that students would be able to gradually accumulate points through a multitude of assignments and therefore be able to receive the motivational benefits of such a system.

Isolated game mechanics. Recently, a number of researchers have recognized that there was a gap in the research surrounding gamification. While the majority of gamification studies use a combination of game elements, researchers agree that it is necessary to study the mechanics of gamification in isolation from each other. Only one published study attempted to test multiple gamification mechanics in isolation against a control group.

Mekler et al. (2017) separated the individual elements of gamification and studied their effects on intrinsic motivation and performance on an online image annotation activity with 172 adult participants. In the study, they applied three game mechanics in isolation: points, levels, and leaderboards. Note that Mekler et al. consider points and levels to be separate game mechanics while the research of this paper does not. After the activity concluded, participants took a modified version of the Intrinsic Motivation Inventory (Deci & Ryan, 2005) in order to assess their intrinsic motivation levels after the intervention. Though Mekler et al.'s study was not conducted in an educational context, the activities were a learning and assessment process, and the results can be connected to education. Mekler's study determined that the gamification of the task did not statistically significantly increase the intrinsic motivation of the participants, did not increase the perceived competence of the participants, and did not significantly increase the performance of the participants. This was a surprise to the researchers. However, the gamified groups did significantly increase the quantity of image tags produced when compared to the control group, which implies an increase in effort. Yet, the quality of the image tags was found to be poorer in the gamified groups as opposed to the control groups. Still, due to the short duration of the study and the limitations associated with the single image-annotating activity, they ultimately concluded that further research must be performed on isolated game mechanics.

Summary

Games have long been considered beneficial to educators. Gamification applies game mechanics into non-game contexts in an effort to apply the benefits in real-world scenarios. While in recent years, there has been a surge of research surrounding this

topic, studies of the individual game mechanics offered by gamification, are inadequate. This is especially true of the game mechanic of points. Published studies showed insignificant results or were affected by a factor that did not allow the game mechanic to take effect properly. Ultimately, there remains a need for additional, more specific, and longer studies into the core elements of gamification.

Chapter III

Method

This chapter covers the procedures that were used to determine (a) the impact of points-based grading on the intrinsic motivation of undergraduate classes, (b) the effectiveness of points-based grading on undergraduate class performance, and (c) how undergraduate students perceive their grade at the beginning of their courses prior to any assignment submission. Additionally, this chapter describes the hypotheses, research design, participants, measurement instruments, procedures, and the limitations of the study.

Research Questions and Hypotheses

The following are the research questions and hypotheses for this study.

(1) How do undergraduate educational technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of intrinsic motivation?

(2) How do undergraduate educational technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of class performance?

(3) How do students in undergraduate educational technology courses for preservice teachers perceive their grade at the beginning of the course prior to receiving grades for any class assignments?

The hypotheses for research question (1) were:

H0: There is no difference in the change in intrinsic motivation towards the course content between students graded using the points-based grading system and students graded using the traditional letter grading system throughout the semester.

H1: There is a difference in the change in intrinsic motivation towards the course content between students graded using the points-based grading system and students graded using the traditional letter grading system throughout the semester.

The hypotheses for research question (2) were:

H0: There is no difference in class performance between the traditional letter grading system and the gamified points-based grading system in educational technology classes for preservice teachers.

H1: There is a difference in class performance between students graded using the points-based grading system and students graded using the traditional letter grading system in educational technology classes for preservice teachers.

No hypothesis testing was performed for research question (3).

Participants

The participants of this study were students enrolled in a face-to-face undergraduate educational technology course for pre-service elementary teachers at a large public university in the southwestern United States. The participants were enrolled in four sections of the course during the spring 2018 semester. All students were asked to participate in the study by the researcher's supervisor. Students were given the option to opt in or opt out of the study via a written consent form (see Appendix B). All students had the option to refuse to participate or to indicate that they were not yet 18 years old.

Consent forms were distributed in the fourth week of class during the spring 2018 semester.

The students of two randomly chosen class sections served as the control group, and the students of the remaining two class sections served as the treatment group. The class sections had twenty students each, resulting in a total of 80 students enrolled in the combined class sections and therefore 80 total potential participants. Of these, 52 students agreed to participate in the study.

Research Design

This study used a quasi-experimental quantitative design. The control group consisted of two of the class sections, chosen at random, and was taught using the traditional letter grading system, as it has been used in that class in the past. The treatment group consisted of the remaining two class sections and had the gamified points-based grading system applied to it, replacing the traditional letter grading.

One drawback in some prior studies on points-based assessment was that the system was applied to just a single assignment (e.g., Attali & Arieli-Attali, 2015). For points generate engagement, a study of greater duration is needed (Mekler et al. 2015; Zichermann, 2012). Therefore, this study took place over eleven weeks of a fifteen-week semester and took into consideration all of the assignments. The duration of this term was intended to hopefully provide more time for points-based grading to have an effect.

On the fourth week of the course, students were asked to consent to participate in the study by the researcher's supervisor. Directly afterward, the students who consented received an online pre-test questionnaire electronically in class to gather their gender, age, familiarity with games, perception of their grade at the beginning of a new class, and

motivation levels in terms of schoolwork at that point in the course. On the last day of the course, the research supervisor returned to the classrooms to administer a post-test questionnaire to determine whether the students' motivation levels changed. The treatment group received additional questions specific to the points-based grading system to determine their impressions about their experience with the points-grading system. Both questionnaires were created and distributed using Qualtrics.

Between these two questionnaires, grades were shared with the students differently depending on whether they were in the treatment or control group. Additionally, both the treatment group and the control group had their separate blackboard shells so that there would be no cross contamination of grading styles. Students in the control group were able to view their grade normally on Blackboard through the "My Grades" menu option found on the course navigation panel. In the class sections that used the points-based system, students were told before any assignments were graded that they had zero points (denoted as "XP" which is short for "experience points"), which the students recognized as a failing grade. With each assignment they submitted, the students gained points and had the chance to level up when reaching pre-determined point thresholds. Students were told that, upon reaching Level 5, they would have effectively obtained an A- in the class relative to the traditional letter grading scale (see Figure 3).

Assignment Title	**Possible XP**
Attendance/Participation	15xp
Google Site Homepage Edit and URL	5xp
Web Resource Review	10xp
Google Collaborative Presentation	10xp
ShareMyLesson Resource Discussion	5xp
Google Form Survey	5xp
Classroom Blueprint	5xp
Newsletter	10xp
Digital Story Script/Picture Plan	6xp
Digital Story Script/Picture Plan Peer Review	4xp
Digital Story Video Project	15xp
Google Site Final Portfolio	10xp
Total XP Available	**100xp**
XP Required for Level 5	**90xp - 100xp**
XP Required for Level 4	**70xp - 89xp**
XP Required for Level 3	**50xp - 69xp**
XP Required for Level 2	**30xp – 49xp**
XP Required for Level 1	**10xp – 29xp**
XP Required for Level 0	**0xp – 10xp**

Figure 3. Assignment Point distribution and level thresholds for the treatment class sections.

Students in the treatment group had the "My Grades" menu option disabled and instead received personal messages through the internal Blackboard messaging system directly from the instructor after each assignment has been graded. These messages were linked to a private online document called a *pointcard* that was viewable only by the instructor/researcher and the student. Each pointcard informed the student of their points for newly submitted assignments, total accumulated points, level, the required amount of points needed to reach the next level, assignment feedback, a history of previously scored assignments, and a celebratory graphic image that only appears if they passed a new point threshold and reached a new level since their previous pointcard (see Figure 4). At the end of the semester, after the post-test questionnaire was administered, the treatment group students' points were converted into a traditional letter grade and the "My Grades"

menu option was made available and viewable within the treatment group's Blackboard course shell.

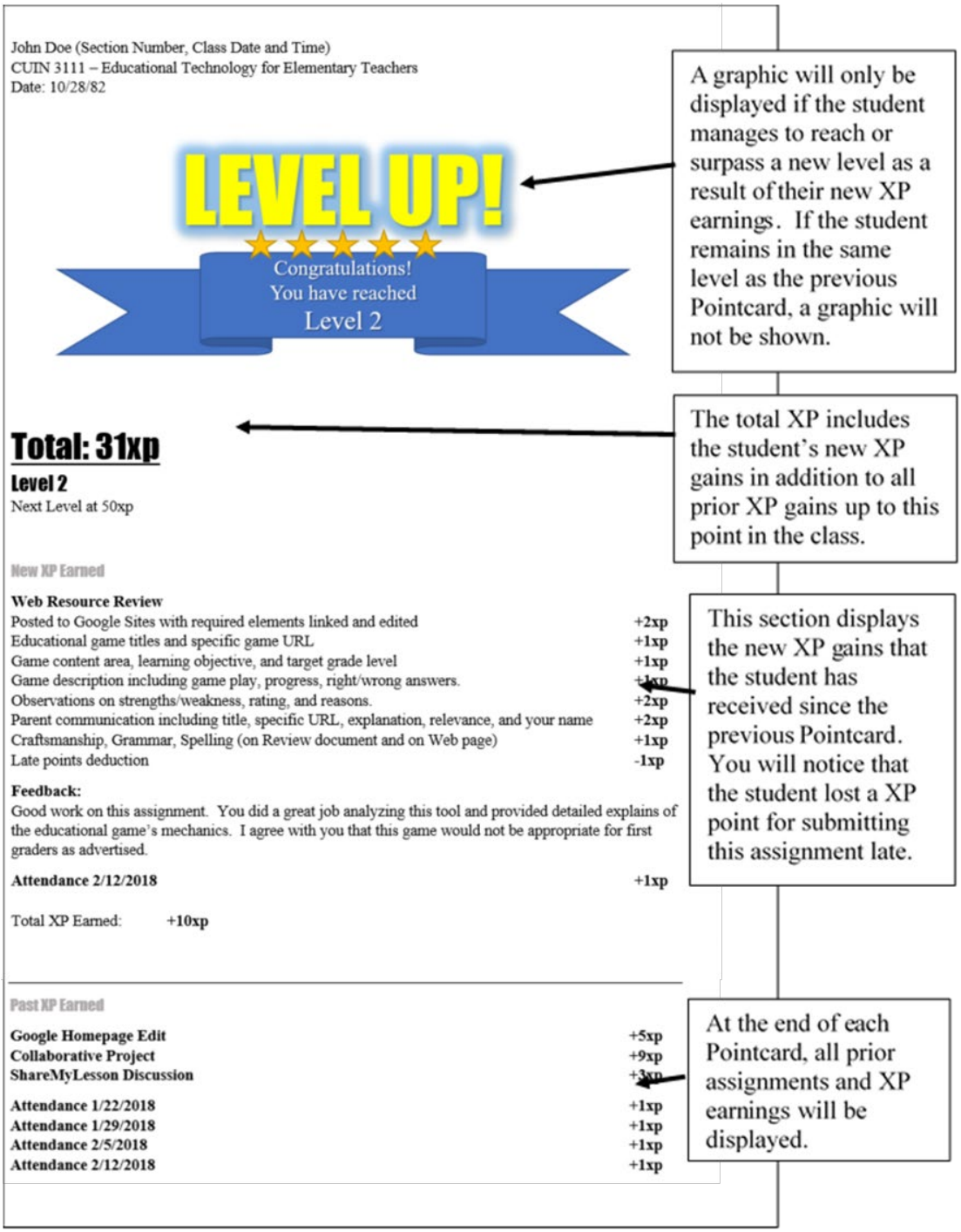


Figure 4. Example Pointcard with section descriptions.

Upon the submission of final grades to the University, data analysis began. The questionnaire data was compiled using SPSS statistics software and grades of the participating students were collected through Blackboard's instructor view for analysis. However, the students' identities were not collected with the grades so that they could not be traced back to the student.

Measurement

Data about student's perceptions of their motivation was gathered using pre- and post-test questionnaires. Both the pre-test and post-test questionnaires were used to determine the intrinsic motivation levels of the students before and after the application of the grading system. Furthermore, the pre-test questionnaire had additional question items to gather demographic information such as age and gender, as well as the student's perception of their grade when first starting the class. The post-test questionnaire included additional question items exclusive to the treatment group to investigate the students' general impressions of the points-based grading system. Lastly, the final grades of all participating students were collected once the semester concluded.

The majority of the four of the questionnaires (pre-test and post-test treatment and pre-test and post-test control) pertain to the measurement of intrinsic motivation. These questions were adapted from the Intrinsic Motivation Inventory (IMI) created by Deci and Ryan (2005). The IMI was selected for this research because it has been used extensively in many research studies to measure intrinsic motivation (Choi, Mogami, & Medalia, 2009; Monteiro, Mata, & Peixoto, 2015). Additionally, the IMI has been used for a number of published gamification studies (Lieberoth, 2015; Mekler et al., 2017; Sailer et al., 2017; Van der Kooji et al., 2019), including gamification studies that

emphasized points-based game mechanics (Ahn et al., 2019; Feger et al., 2019; Mekler et al., 2013). The use of the IMI to measure intrinsic motivation provides this study with a valid instrument that replicates the methods used by the authors of existing gamification literature.

The questionnaire items come from the 45 items post-experimental IMI, though not the entire inventory. The 45 items are broken down into seven separate subscales, with each subscale containing 5 to 8 items. Each IMI questionnaire item was presented as a statement that participants were asked to indicate how true the statement is for them using a 7-point Likert scale. Deci and Ryan (2005) recommend using only the portions of the IMI that are needed for the target study. Of the seven subscales, the Interest/Enjoyment, Perceived Competence, Effort/Importance, and Pressure/Tension subscales were used in the questionnaire design. The other three subscales were not used as these did not apply to the student participants (Perceived Choice), their assignments (Relatedness), or were not influenced by a grading system (Value/Usefulness). The items were adapted so that they related to the content of the course and could function as pre-test and post-test question items (see Appendix A). The pre-test items reference the participant's past academic experiences whereas the post-test items reference the participant's experience with the specific course in which they are enrolled. As per the IMI's recommendation, the question order was randomized in the final distributed questionnaires. Finally, all questionnaire items contained a seven-point Likert scale.

Both pre-tests were identical for both the treatment and control groups. However, in addition to the IMI, the pre-tests had additional question items. The pre-tests began with a section intended to capture demographic data (gender and age) as well as a

question to determine each participant's perception of their familiarity with games. Additionally, a multiple-choice question was included in the pre-tests to measure what grade the student subjects believed that they had at the beginning of a new class before any assignments have been graded (PerceivedGrade). Up to the point in which the pre-tests were distributed, no assignments had been graded, and no grading system or gamification had yet been applied to the class. The question item asked what grade they believed they had at the time using the traditional letter grades (A, B, C, D, and F).

In addition to questionnaire data on the IMI and demographic information, the students in the treatment group received additional questions on their impressions of the points-based grading system. These questions were only provided to the treatment group because the control group did not receive the implementation of the points-based grading system in their class. Additionally, these questions were only found on the post-test questionnaires since participants would not have been familiar with the grading system when they received the pre-test questionnaire. The points-based grading impressions questions were placed at the end of the treatment post-test questionnaires (see Appendix A).

The points impressions questions were the following:

1. Earning points and levels motivated me to learn more than just traditional letter grades. (PointsQ1)
2. I prefer a traditional letter grading system over the points and levels system used in this particular class. (PointsQ2R)
3. I felt more engaged in this particular class using points and levels than in other classes using traditional letter grades. (PointsQ3)
4. I felt that I was rewarded more for my performance on activities in this particular class. (PointsQ4)

5. I felt that points and levels provided milestones that I was encouraged to reach. (PointsQ5)
6. I found the points grading system frustrating. (PointsQ6R)
7. The points grading system was easy to understand. (PointsQ7)

These questions were designed to be similar in structure to the IMI question items. Like the IMI, for these items, participants were asked to evaluate the truth of each statement. Also like the IMI, students responded using a 7-point Likert scale (1 = very untrue; 4 = somewhat true; 7 = very true). Finally, the questions were randomized when the students took the questionnaire, again matching the recommended IMI implementation. The reasoning behind this design was to make these questions easy for participants to grasp as students will have, by that point, become accustomed to the format of the IMI questions.

The pre-test questionnaire was distributed after receiving consent from the participants, prior to any assignment in the class being graded. The post-test questionnaire was distributed on the last day of the semester. All questionnaires were presented and distributed to the participants by the researcher's supervisor. While the questionnaires were being administered, the researcher (who was also the teacher) left the classroom. The questionnaires can be found in Appendix A.

This study has been approved by the Institutional Review Board.

Data Analysis

A one-way multivariate analysis of variance (MANOVA) method was used for the pre-test and post-test IMI subscales scores to analyze the data for research question (1). The independent variable was the Grading System (Treatment & Control), which is a between-subjects factor. The dependent variables were the four IMI subscales:

Interest/Enjoyment, Perceived Competence, Effort/Importance and Pressure/Tension. Additionally, since the intrinsic motivation was measured using a pre-test and a post-test questionnaire, a MANOVA was first performed on the pre-test IMI subscales to determine if there was no significant difference between the pre-test scores. The subscales of the IMI questions are broken down into four separate subscales, resulting in four dependent variables. The subscale variables are Interest/Enjoyment, Perceived Competence, Effort/Importance, and Pressure/Tension. The dependent variables are the mean scores of the combined items within each of the IMI subscales that are used in the questionnaires. The intention was to shed light on whether or not a point-based grading system is able to have any influence over student intrinsic motivation in any of the subscales over the fifteen-week semester.

For the second research question, the Grading System was used as the independent variable. Since the dependent variable, Class Performance (as determined by the participants' final grades), is not a continuous variable since each students' final grade could only be between 0 and 100, time was not a factor, and no pre-testing occurred. As a result, an independent samples t-test was performed with Class Performance as the dependent variable and the Grading System as the sole independent variable.

Research question (3) asked whether or not students perceive that they possess an A or an A equivalent grade at the start of each new class before any assignments have been submitted. Since the result of this question was determined by only a single question item (in the pre-test questionnaires given to both groups), descriptive statistics (frequency) was used to determine whether or not a significant portion of both groups of

student participants believe that they have an A or an A equivalent grade at the start of a new class.

Summary

This chapter contained detailed information on the many steps that took place during the study. Additionally, the researcher's hypotheses were shared along with the procedures set in place for the data analysis. The participants were described, and the measurement questionnaires were explained and shown in full. Lastly, the limitations were explored. The next chapter will present the results of the study.

Chapter IV

Results

This chapter covers the findings that resulted from the study. The study used a quantitative research design of a points-based grading system through statistical analysis of student Intrinsic Motivation Inventory (IMI) scores, class performance, and perceived grade recorded using pre-treatment and post-treatment questionnaires. The participants were students in an educational technology course for pre-service elementary teachers. Four class sections were used for this study and randomly divided into two treatment courses and two control courses. No identifiers were used in the data analysis.

A total of 52 student participants gave consent to have their final grades collected and to take part in the study. However, only 49 participants took the pre-test motivation questionnaire (Treatment $n = 21$, Control $n = 28$) and 47 took the post-test questionnaire (Treatment $n = 19$, Control $n = 28$). This was a result of participants not being present on the class day when the post-test questionnaire was available.

Participant Demographics

The participants for the study were 52 undergraduate students that were studying to become elementary teachers. The study took place at a major research university in the United States. On the pre-test questionnaires, participants were asked two preliminary questions to determine demographic information. These questions were on student gender and student age.

Participants were asked a multiple-choice question on their gender. Students were only given the ability to answer male or female. They also had the option to leave the question unanswered in the case that neither option appropriated defined them, though

none of the students opted to do so. The overall majority of the participants (n = 46) were female, and there was a small number of male participants (n = 3).

As for participant age, participants were provided a slide bar with a range of 0 - 100 that they could manually adjust to provide their age (n = 49). The expected grade level for students of the educational technology course for pre-service elementary teachers was the sophomore or junior undergraduate level. The age demographics reflect the expected age typical of sophomore or junior undergraduates, as the students were primarily within the 19 – 26 year old young adult age range, with an average overall age of 22.5 years. However, this average is skewed by two older students in their mid-40s (44 and 46) that also attended the course and enrolled in the study (n = 2), which made the overall age range 19 to 46 (n = 44). Like the gender question above, students were given the ability to skip this question, and a handful of participants opted to not provide their age on the pre-test questionnaire (n = 5).

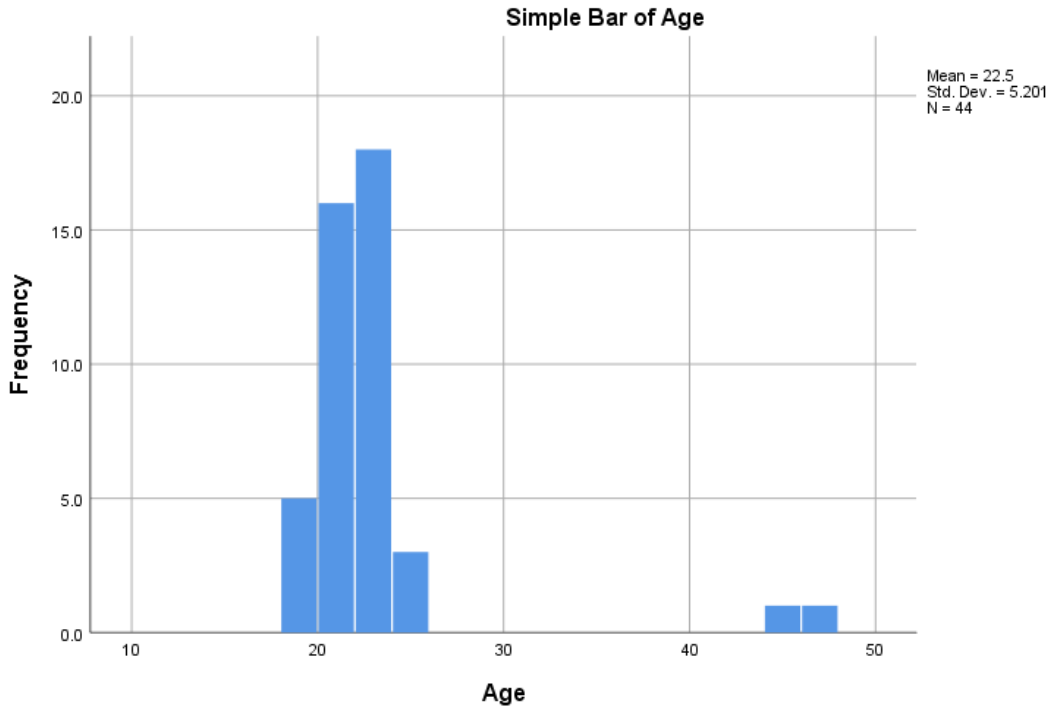


Figure 5. Histogram of participant ages.

Game Experience

In addition to preliminary question items on participant demographics, an additional question was asked to determine the participants' familiarity with games on the pre-test questionnaire (n = 49). Using a Likert scale from 1 (not at all true) to 7 (very true), they were asked to indicate to what extent that they consider themselves gamers, which included mobile games, web-games, and video games. All 49 pre-test participants answered this question (M = 2.55). The majority of the students placed themselves on the non-gamer side of the scale. As evidence, the most selected answer was 1, which meant that those students did not consider themselves gamers at all and most likely have little to no experience with any form of video game (n = 19). While, the majority of the participants exhibited some form of familiarity with gaming, as the other 30 participants

all indicated that they at least have some familiarity with the gamer identity, the participants did not identify as gamers.

Table 2

Frequency Distribution of Game Experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	19	36.5	38.8	38.8
	2	11	21.2	22.4	61.2
	3	3	5.8	6.1	67.3
	4	11	21.2	22.4	89.8
	5	2	3.8	4.1	93.9
	7	3	5.8	6.1	100.0
	Total	49	94.2	100.0	
Missing	System	3	5.8		
Total		52	100.0		

Note. 1 = no experience, 7 = highly experienced.

Figure 6. Bar Chart for the Game Experience Question on the Pre-Test Questionnaire.

Intrinsic Motivation

This section of the chapter covers the data analysis process and results for research question (1): How do undergraduate technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of intrinsic motivation?

The hypotheses for research question (1) were:

H0: There is no difference in intrinsic motivation towards the course content between students graded using the points-based grading system and students graded using the traditional letter grading system throughout the semester.

H1: There is a difference in intrinsic motivation towards the course content between students graded using the points-based grading system and students graded using the traditional letter grading system throughout the semester.

Intrinsic motivation inventory. A one-way multivariate analysis of variance test (MANOVA) was first conducted on the pre-test questionnaire data (PreIMIG) and then on the post-test questionnaire data to determine whether the grading systems affected the intrinsic motivation levels of the preservice teachers (PostIMIG). The questionnaire questions created using the Intrinsic Motivation Inventory (IMI), created by Deci & Ryan in 1982, as a base. The IMI is broken down into several subscales that are intended to be analyzed individually. As such, the MANOVA test was selected for this analysis since the questionnaire items contained multiple dependent variables for the IMI subscales that are also interrelated.

Pre-test MANOVA. In order to be able to apply the MANOVA on the post-test data to make the comparison, tests first needed to be performed on the pre-test data to determine whether the base IMI scores at the beginning of the treatment period were statistically significantly different between the groups. If the pre-test data is not statistically significantly different, then the post-test data can be directly compared between the two groups to determine if any change in intrinsic motivation occurred as a result of the experiment. A MANOVA was first conducted on the pre-test questionnaire IMI scores (PreIMI) between the two groups: pre-test treatment (n = 19) and pre-test control (n = 28). Assumption tests for univariate and multivariate outliers, multivariate normality, multicollinearity, linear relationships, homogeneity of variance-covariance matrices, and homogeneity of variances were performed. Several univariate outliers were found in the data, but were not removed. No other violations of assumptions were found. Detailed analysis of pre-test assumptions testing can be viewed in Appendix 3 in the section Pre-Test Assumptions Analyses.

The multivariate tests illustrated that there was no statistically significant difference between the two groups as shown by Wilk's Lambda, F (4, 42) = 2.217, Wilks' $\Lambda = .174$; partial $\eta^2 = .128$ (see Table 4). Wilk's Lambda was selected for the multivariate statistics as it is commonly recommended for the MANOVA procedure (Field, 2017). While not significant, the control group exhibited higher Interest/Enjoyment, Perceived Competence and Effort/Importance over the treatment group and the treatment group reported that they felt less Pressure/Tension than the control group (see Table 3).

Table 3

Pre-Test Descriptive Statistics.

	Treatment	Mean	Std. Deviation	N
IntPreG	Control Group	4.1633	.85953	28
	Treatment Group	3.9023	.57351	19
	Total	4.0578	.76099	47
CompPreG	Control Group	4.8393	1.02688	28
	Treatment Group	4.7544	.65113	19
	Total	4.8050	.88691	47
EffPreG	Control Group	4.6286	.38764	28
	Treatment Group	4.5263	.35409	19
	Total	4.5872	.37394	47
PresRPreG	Control Group	4.0929	.70864	28
	Treatment Group	4.6316	.64726	19
	Total	4.3106	.72808	47

Note. a higher Pressure/Tension score equates to less pressure/tension.

Table 4

Multivariate Tests for the Pre-Test MANOVA

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.996	2411.661[b]	4.000	42.000	.993	.996
	Wilks' Lambda	.004	2411.661[b]	4.000	42.000	.993	.996
	Hotelling's Trace	229.682	2411.661[b]	4.000	42.000	.993	.996
	Roy's Largest Root	229.682	2411.661[b]	4.000	42.000	.993	.996
Treatment	Pillai's Trace	.174	2.217[b]	4.000	42.000	.128	.174
	Wilks' Lambda	.826	2.217[b]	4.000	42.000	.128	.174
	Hotelling's Trace	.211	2.217[b]	4.000	42.000	.128	.174
	Roy's Largest Root	.211	2.217[b]	4.000	42.000	.128	.174

a. Design: Intercept + Treatment

b. Exact statistic

Post-test MANOVA. In order to conduct the MANOVA test on the post-test IMI subscales, preliminary analyses were first conducted to test several assumptions. Assumption tests for univariate and multivariate outliers, multivariate normality, multicollinearity, linear relationships, homogeneity of variance-covariance matrices, and homogeneity of variances were performed. Several mild univariate outliers were identified, but were not removed. No other violations of assumptions were found. These assumption analyses can be found in Appendix 3 in the section Post-Test Assumptions Analyses.

Having tested the various assumptions for performing the MANOVA test, the test was then conducted on the participant post-test IMI data. Participants were separated into two groups: post-test treatment (n = 16) and post-test control (n = 29). The intrinsic motivation levels were further broken down into four focused subscales: Interest/Enjoyment (PostIntG), Perceived Competence (PostCompG), Effort/Importance

(PostEffG), and Pressure/Tension (PostRPresG). The multivariate tests for the IMI subscales were conducted. Wilk's Lambda was selected for the multivariate statistics, as it is commonly recommended for the MANOVA procedure (Field, 2017). Again. Wilk's Lambda showed that there was no statistically significant difference between the treatment and control groups on the combined dependent variables on the post-test IMI subscale scores, $F (4, 40) = 1.465$, Wilks' $\Lambda = .231$; partial $\eta^2 = .128$.

Table 5

Multivariate Tests for the Post-Test IMI Subscales

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.993	1476.378[b]	4.000	40.000	.000	.992
	Wilks' Lambda	.007	1476.378[b]	4.000	40.000	.000	.992
	Hotelling's Trace	147.638	1476.378[b]	4.000	40.000	.000	.992
	Roy's Largest Root	147.638	1476.378[b]	4.000	40.000	.000	.992
Treatment	Pillai's Trace	.128	1.465[b]	4.000	40.000	.231	.128
	Wilks' Lambda	.872	1.465[b]	4.000	40.000	.231	.128
	Hotelling's Trace	.146	1.465[b]	4.000	40.000	.231	.128
	Roy's Largest Root	.146	1.465[b]	4.000	40.000	.231	.128

a. Design: Intercept + Treatment, b. Exact statistic

Though not significant, the points-based treatment group scored higher in their Interest/Enjoyment, Perceived Competence, in the amount of Effort/Importance they put into the class and felt less Pressure/Tension than the control group graded using the traditional letter grading scale (see Table 6).

Table 6

Descriptive Statistics for The Post-Test IMI Subscales.

	Treatment	Mean	Std. Deviation	N
IntPostG	Control Group	4.6453	1.16097	29
	Treatment Group	4.8393	.83605	16
	Total	4.7143	1.05111	45
CompPostG	Control Group	4.7701	1.22265	29
	Treatment Group	5.2083	.60093	16
	Total	4.9259	1.05801	45
EffPostG	Control Group	4.2897	.51154	29
	Treatment Group	4.6250	.54589	16
	Total	4.4089	.54265	45
PresRPostG	Control Group	4.2483	.70641	29
	Treatment Group	4.3375	.72927	16
	Total	4.2800	.70762	45

Note. The Pressure/Tension subscale is reversed.

Interestingly, tests of between-subjects effects showed that there was a statistical difference between the control and treatment groups specifically on the Effort/Importance IMI subscale (PostEffG), $F(1) = 4.227$, $p = .046$, partial $\eta^2 = .089$ (See appendix 3). However, since Wilk's Lambda takes precedence over the test of between-subjects effects, the ultimate result was that there was no statistically significant difference in any of the IMI subscales between the traditional grading group and the points-based grading group.

Class Performance

This section of the chapter covers the data analysis process and results for research question (2): How do undergraduate educational technology courses for

preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of class performance undergraduate technology courses for preservice teachers?

The hypotheses for research question (2) were:

H0: There is no difference in class performance between the traditional letter grading system and the gamified points-based grading system in educational technology classes for preservice teachers.

H1: There is a difference in class performance between students graded using the points-based grading system and students graded using the traditional letter grading system in technology classes for preservice teachers.

In addition to collecting data through the pre-test and post-test questionnaires, the final grades for each consenting student were also collected. This data was gathered from the Blackboard course shell for the four sections of the class after the conclusion of the semester term. The purpose was to determine whether the different grading systems had any impact on overall class performance. The grades for the students in both the treatment and control groups were separated and compared to each other. In total, 52 final grades were collected (Control $n = 29$, Treatment $n = 23$).

Independent samples t-test. To determine whether there was a difference in class performance between the two groups, an independent samples t-test was performed on the finals grades of the participants between the treatment group and the control group. The independent samples t-test was chosen because there is only one dependent variable (FinalGrade) and one independent variable (Treatment). Furthermore, there is independence of observations between the independent variable and dependent variable.

For these reasons, the independent samples t-test was the preferred choice for analyzing the data for research question (2). Tests for assumptions of no outliers, of normality, and of homogeneity of variances were conducted. One mild outlier was identified, but not removed from the data. No other assumption violations existed. A detailed analysis of the assumptions testing can be found in Appendix 3 in the section Class Performance Assumptions Analyses.

Table 5

Descriptive Statistics for Class Performance by Group

	Treatment	N	Mean	Std. Deviation	Std. Error Mean
FinalGrade	Control Group	29	92.10	5.205	.967
	Treatment Group	23	93.61	7.785	1.623

The t-test showed that the difference between the two variables being compared was not statistically significantly, t (50) = -.833, p = .409, r = .12 (see Table 8). Overall, the treatment group had a slightly higher average final grade than the control group. The treatment group had an average final grade of 93.61 out of 100, whereas the average final grade for the control group was 92.10 out of 100. The lowest grade also belonged to the treatment group, with one student receiving a grade of 73 out of 100. However, this was balanced by a student in the treatment group also having the highest grade, with a 103 out of 100 due to extra credit. In comparison, the control group minimum grade was 81 out of 100 and the maximum was a 100 out of 100.

Table 6

T-Test Comparing the Class Performance of the Control and Treatment Group

		t-test for Equality of Means		t-test for Equality of Means			t-test for Equality of Means 95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Diff	Std. Error Diff	Lower	Upper
Final Grade	Equal variances assumed	-.833	50	.409	-1.505	1.806	-5.133	2.122
	Equal variances not assumed	-.797	36.737	.431	-1.505	1.889	-5.334	2.324

Grade Perception

This section of this chapter covers the data analysis and results of research question (3): How do students in undergraduate educational technology courses for preservice teachers perceive their grade at the beginning of the course prior to receiving grades for any class assignments? No hypothesis testing was conducted.

To evaluate student grade perception, a question item was placed in the pre-test questionnaires for all participants. Participants were asked a multiple-choice question: "When you started the class, what grade did you believe you had?" and given the ability to select A, B, C, D, or F in response. The question used traditional letter grading terminology, because it would have been the grading system that students were most familiar with at that point and since the pre-test questionnaire was given prior to any treatment.

Forty-nine participants answered this question (n = 49). Most (97.95%; n = 48) of the students reported that they believed that they had an A grade in the course when they started the class. Only 2.04% (n = 1) of the students reported that they believed that they had a B grade when they started the class. None of the participants reported that they believe they had a C, D, or F grade when they started the class

Table 9

Frequencies Table for Participant's Perceived Grade at the Start of the Class

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Grade A	48	92.3	98.0	98.0
	Grade B	1	1.9	2.0	100.0
	Total	49	94.2	100.0	
Missing	System	3	5.8		
Total		52	100.0		

Points Impressions

Twenty students completed the impressions section of the treatment group's post-test questionnaire (n = 20). The mean results of the points impressions questions can be found in Table 10. A full breakdown of all seven points impressions questions in available in Appendix D.

Table 10

Descriptive Statistics for the Post-Test Points-Based Grading Impressions Questions

	PointsQ1	PointsQ2R	PointsQ3	PointsQ4	PointsQ5	PointsQ6R	PointsQ7
N	20	20	20	20	20	20	20
Mean	3.60	5.20	3.10	4.05	4.10	3.95	3.50
Median	3.50	6.00	3.00	4.00	4.00	4.50	3.50
Mode	1[a]	7	2[a]	5	4	6	4

Note. The scores for questions two and six are reversed.

Summary

This chapter covered the data analysis procedures and the results of the experiment. For research question (1), a one-way multivariate analysis of variance test (MANOVA) was performed on the post-test IMI subscales comparing the two groups. There was no statistically significant difference between the results of the traditional grading groups and the points-based grading groups in any of the IMI subscales. For research question (2), an independent samples t-test was performed comparing the final grades of the participants in both groups. There was no statistically significant difference between the class performance of the students in either group. For research question (3): descriptive statistics were run on participant's perceptions of their grade at the start of the course and 97.95% of the students reported that they believed that they started the course with an A. The next chapter will discuss the results in relation to the wider literature, analyze the impressions that participants in the treatment group had of points, and suggest future studies.

Chapter V

Conclusion

The purpose of this quantitative study was to apply a points-based grading system in a class to determine whether it had any impact on the intrinsic motivation and class performance of the students. In this chapter, the results of the study are discussed and considered in relation to the broader literature surrounding gamification. Furthermore, additional data collected on how the treatment group felt about the points-based grading system was analyzed and considered for how it may inform the results of the experiment. Also included in this chapter are suggestions for future studies. Limitations of the study are discussed throughout the chapter.

Overview of Results

(1) How do classes taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of intrinsic motivation in undergraduate educational technology courses for preservice teachers?

A MANOVA was conducted on the post-test IMI subscale scores between the control and treatment groups. Wilk's Lambda determined that the multivariate analysis was not statistically significant (Wilks' $\Lambda = .231$). Ultimately, there was no statistically significant difference between the points-based grading group and the traditional letter grading group in terms of intrinsic motivation.

(2) How do classes taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of class performance in undergraduate educational technology courses for preservice teachers?

An independent samples t-test was conducted comparing the final grades between the control and treatment groups. The results of the experiment showed that there was no statistically significant difference between the two groups in terms of class performance.

(3) How do students in undergraduate educational technology courses for preservice teachers perceive their grade at the beginning of courses prior to the submission of any assignments?

Of the students, 97.95% (n = 48) reported that they believed that they had an A grade in the course when they started the class. The only other participant selected a B grade.

Discussion

From the results of the study, the points-based grading system did not significant impact student intrinsic motivation or class performance in comparison to traditional letter grading. However, the results of research question (3) strongly supports the theory that traditional letter grading could be a demotivational system in which students might perceive to be punished for submitting assignments with grade loss for which this study sought to find an alternative (Kohn, 1999). A gamified-points based grading system was tested, but the results were shown to be insignificant, given the body of empirical research that educators have on gamified points. Points-based grading cannot be recommended at this time for consideration as a replacement for the grading systems that is currently in place.

However, it would be a hasty to abandon the points-based grading system entirely from these results alone. Even with the results of this study, educators cannot yet conclude that points-based grading has no impact on students' intrinsic motivation or

class performance. Taking a second look at the data in this study shows that there may have been a small impact on intrinsic motivation and class performance. The MANOVA test on the post-test IMI score resulted in the following: the IMI subscale scores where higher for the treatment group for Interest / Enjoyment (PostIntG), Perceived Competence (PostCompG), and Pressure/Tension (PostRPresG) but the differences between the two groups was not statistically significant ($p = .056$; $p = .187$; $p = .690$; see Table 5). The IMI subscale for Effort / Importance (PostEffG) was also higher for the treatment group and was statistically significantly ($p = 0.46$, see Table 11) but could not be accepted as a result due Wilk's Lambda determining that the multivariate analysis was shown not to be statistically significant (Wilks' $\Lambda > .05$). Additionally, the points-based grading group had a slightly higher overall average final course grade (Treatment $M = 93.61$; Control $M = 92.10$; see Table 7) as well. The descriptive statistics alone show that the gamified group scored higher on all subscales of intrinsic motivation and in terms of class performance. The gains were marginal and, possibly due to the small sample size (treatment $n = 23$; control $n = 19$), could not be deemed statistically significant.

Table 11

Results of Tests of Between-Subjects Effects for the Post-Test IMI Subscales

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Treatment	IntPostG	.388	1	.388	.346	.560	.008
	CompPostG	1.980	1	1.980	1.801	.187	.040
	EffPostG	1.160	1	1.160	4.227	.046	.089
	PresPostG	.082	1	.082	.161	.690	.004

a. R Squared = .008 (Adjusted R Squared = -.015)

b. R Squared = .040 (Adjusted R Squared = .018)

c. R Squared = .089 (Adjusted R Squared = .068)

d. R Squared = .004 (Adjusted R Squared = -.019)

Points Impressions

The treatment group received additional questions on their impressions of the points-based grading system. Analyzing the responses to the post-test questionnaire may provide insight into why the points-based grading system had no significant effect. Additionally, the analysis could provide useful information for researchers who wish to investigate the points-grading system in future studies. There are three main points that we can extract from the points-impressions data from the post-test questionnaires: 1) participant impressions of the points-based grading system varied greatly and were both positive and negative, 2) a large proportion of the treatment group found the gamified points-based grading to be difficult to understand, and 3) participants who experienced the points-based grading system preferred it over the traditional grading system. In this section, the implications of these three points will be discussed regarding the experiment as a whole.

1. Participant impressions of the points-based grading system varied greatly and were both positive and negative.

For all points-impressions questions, except for question 2 (which is discussed in the third main point in this section), the participants reported a wide range of impressions. This is shown by the mean scores for each for those questions, which all fall within one point from the middle ($3 < x < 5$). In many of the questions, students were nearly divided equally between agreeing and disagreeing with the statement and all questions featured substantial responses favoring each side. This suggests that the gamification applied in this study was influential to some, but not all students in the group. This invites the possibility that points-based grading is not a system that influences students equally and that there likely is a divide between those that are affected by gamification and those that are not. One factor that might influence the effectiveness of gamification on an individual is their familiarity with games and the game environment. Game familiarity and its potential influence on the results will be discussed in the next section. What factors determine whether gamification can influence a person is not known given the current research and was beyond the scope of this study. Future studies on gamification could capture additional data on the participants to try to understand this phenomenon.

These mixed opinions imply that the gamified grading system may still have value. Considering that on question 4, a majority of the treatment group stated that they felt more rewarded in the gamified class and on question 5, a majority of the class claimed that having points and levels as milestones were a source of encouragement, it seems that the points-based system had some influence on at least some students. As

stated, current research has not explained what personal attributes determine the effectiveness of gamification on individuals. Still, it is clear that the points-based grading system has value for those susceptible to its influence. There is a large body of empirical evidence has been collected about the different learning styles that individuals possess; styles that the best teachers need to accommodate (Cassidy, 2004). Likewise, it stands to reason that different individuals have different preferences (or reactions) to how they are graded. Towards this end, points can serve as a separate grading style that can be applied to students who respond favorably to its bottom-up approach. The researcher believes that the gamified points-based grading system may be better suited, not as a replacement but as an added alternative, for traditional letter grading.

2. A large proportion of the treatment group found the gamified points-based grading system to be difficult to understand.

Question 6 asked if the students found the points-based grading system to be frustrating and question 7 asked if the grading system was easy to understand. Overall, students did not find it frustrating, but there were some students that reported frustration (n = 8). Furthermore, half of the students (n = 10) reported that the grading system was hard to understand. The portion of the treatment group that had difficulty comprehending the points-based grading system was noteworthy.

One possible explanation for why so many of the students had difficulty processing the points-based grading system could be because these students were not familiar with the game mechanics upon which the points-based grading system was based. As stated in Chapter 4, the majority of the participants did not consider themselves gamers, with 67.8% of participants reporting that they had low familiarity

with games on the pre-test questionnaire. Thus, the high rate of confusion about the grading system may have been a result of the overall low game familiarity of the study sample. Upon accessing the pointcards, these non-gamer students may have had trouble comprehending the information. This potential barrier in understanding the points-based system could have negatively contributed to the statistically insignificant results of this study. In future implementations of points-based grading, more effort should be put towards informing and reminding students how to understand their points.

Another possible reason that students had difficulty with the study was how the graded pointcards were delivered to students. As stated in Chapter 3, participants in the treatment group received their grades via the internal messaging system through the Blackboard learning management system course page. Unfortunately, this system is not very popular with instructors or students at the study site and is rarely used. Far more often, traditional email is used by faculty at this university to communicate with students. Unfortunately, instructors at this university are not permitted to send grades via email. Since the pointcards were equivalent to grades for the treatment group, the pointcards were digitally distributed through the LMS's internal messaging system. This was stated in class, but students in the treatment group may have experienced difficulty locating and accessing their pointcards, which would have also negatively influenced the results of this study and the exposure that the treatment participants may have had with the points-based grading system as a whole. In future studies on points-based grading, more effort should be put towards instructing and reminding students on how to access their points.

3. Participants who experienced the points-based grading system preferred it over the traditional grading system.

Curiously, only one of the participant impressions questions was skewed favorably towards one grading system (PointsQ2R). That question asked the treatment students whether they preferred the traditional letter grading system or the points-based grading system used in their class. The majority of the students selected responses favoring the points-based grading system (n = 14). Additionally, the most selected response to this question item was *very untrue*, which signifies favor for the points-based grading system (mode = 1, n = 1). Even though there was no statistically significant difference in intrinsic motivation or class performance, the students still preferred the gamified approach over the traditional. This result is unexpected since in questions 1 and question 3, students reported that they did not feel particularly more motivated or engaged in the class as a result of the points-based system. This may imply that students are aware that the traditional letter grading system is problematic and are eager to find a new grading system.

Game Familiarity

Another potential reason that the points-based grading in this study resulted in no statistical significance could be related to the level of game experience possessed by the participants. In the pre-test questionnaires, all participants were also asked how much they identified themselves as gamers using a Likert scale from 1 (very untrue) to 7 (very true), like the IMI. In this self-evaluation, participants were asked to include mobile games, web-games, and video games in their consideration. All 49 pre-test participants answered this question (M = 2.55, Median = 2, Mode = 1, SD = 1.733). For this study, the majority of the students were on the less experienced side of gaming. As evidence, the most selected answer was 1: *very untrue* (n = 19), which meant that those students did

not consider themselves gamers at all and most likely had little to no experience with any form of video game. Only five (10.204%) of the 49 pre-test participants favorably identified themselves as gamers on the top half of the Likert scale (see Figure 6).

Being familiar with the game mechanics in gamification may be a predictor of gamification's success on a student. However, research is mixed on this aspect. For instance, Papp (2017) applied a combination of game mechanics, including points (denoted as XP), leaderboards, and a narrative to an undergraduate course on business communications. Using qualitative methods, students reported being more motivated to learn in the gamified classes. However, 75.5% of students also reported that they do not play games regularly, yet they still perceived that they benefited from gamification. As the relationship between student game familiarity and the impact of gamification cannot be confirmed, more studies need to be done on this topic.

Combined Game Mechanics

Another possible explanation for why the points-based grading system did not have statistically significant results may be the result of isolating the game mechanic from the others. If educators remove the perception of how points are presented, realistically points are not very different from traditional grades. A similar study that focused primarily on gamified points in isolation was by Ahn et al. (2019), who tested the impact of points on children's physical activity exercises. Their study on points also ended with insignificant results on the IMI subscales. Ahn et al. cite other gamified physical activity studies that used a combination of game elements that were successful in increasing physical activity (Peng, Crouse, & Lin, 2013; Pope, Lewis, & Gao, 2015)

and concluded that the points-based system alone could not significantly influence the amount of physical activity in children.

Not including other elements of gamification meant that the gamification performed in this study was missing the social constructivist aspect of gamification (Cheong et al., 2014). As mentioned in Chapter 2, points are one of the three most popular gamification mechanics. The two other mechanics, badges and leaderboards, function with a social purpose. Badges provide an aesthetic award that represents evidence of a student's achievements that they can share with others on social media (de Rocha Seixas et al., 2016). Leaderboards are a public list of all students' progress in order of best performing to worst. This public comparison of peers is meant to incite social-competition (Banfield & Wilkerson, 2014; Hanus & Fox, 2015). These two game mechanics have a direct social aspect to them, which points lack on its own. Perhaps, this social aspect of gamification is a strong contributor to its successes in other studies, and without it, the game mechanic's impact is far less pronounced.

Relation to Other Research

Attali and Arieli-Attali (2015) performed a study that focused solely on the points-based grading system. In Chapter 2, this study was analyzed and the researcher considered that the research design insufficiently measured points as an alternative grading system. The primary critique was that points-based grading was only applied to a single assignment, a math exam, and it did not allow adequate time for points to take effect. However, the present study was a longitudinal design that took place over a semester and yet the results were the same – no statistically significant difference in performance. This can be interpreted in a number of ways. Either the study of points-

based grading needs to be even longer, points need to be combined with other game mechanics to take effect, or points do nothing at all. In this section, the results of this study will be compared to that of the existing empirical studies performed on points and gamification as a whole.

Mekler et al.'s (2017) conducted one of the only other gamification studies that attempted to analyze the effectiveness of game mechanics individually. Mekler et al. applied gamification mechanics to an abstract image-tagging exercise. In addition to applying points and levels, leaderboards were also included in the gamification. However, the results were similar to the research results found in this study. Also using the Intrinsic Motivation Inventory (IMI), Mekler's research determined that the gamification of the task did not significantly change the intrinsic motivation of the participants, did not significantly change the perceived competence of the participants, and did not significantly change the performance of the participants. All of the IMI and performance results of Mekler's study mirror and reinforce the results of this study.

Mekler's study did have another interesting finding that could inform future studies on points-based grading. Unlike the present study, Mekler did not use the IMI to measure the Effort/Important subscale. Instead, the quantity and the quality of the image tags were tracked and compared across the treatment and control groups. They discovered that the gamified groups did significantly increase their image tag quantity compared to the control group, which may suggest an increase in effort or feelings of importance in the task. However, the overall quality of the tags was inferior in the gamified group when compared to the control group, which suggests a decrease in Effort/Importance by another metric. As compared to the present study, the

Effort/Importance subscale results were nearly statistically significant, which implies that a points-based grading system may have some influence over this subscale.

This specific interaction is echoed in another study by Ahn et al. (2019). Ahn applied a points system to a three-day long exercise activity for children. The researchers observed that, after the first day, participants quickly identified what actions they needed to complete to gain the most points and would only exercise to the extent that rewarded the most points. This was in contrast to the control group in which the participants engaged in more elaborate and complete physical activity. Ahn's experiment again suggests that there is a discrepancy in how points motivate student effort. The Effort/Importance subscale may be too broad to adequately evaluate the impact that points have on intrinsic motivation. Going forward, it may be beneficial to break down the Effort/Importance subscale into two more specific subscales – one to determine the quantity of work and one to determine quality.

Future Research

The studies on educational gamification are limited at best and additional research is needed to investigate gamification's effectiveness in the classroom (Hamari, 2015; Hanus & Fox, 2015; Landers & Armstrong, 2015). Efforts are already being made to gamify courses, programs, and even schools around America, yet educators do not yet possess the empirical evidence to confirm its effectiveness (Cohen, 2011; Hamari, 2015; Hanus & Fox, 2015). This is especially true for the game element of points, as unfortunately, research performed analyzing the effects of just a points-based grading system is less common than those performed on badges and leaderboards (Attali &

Arieli-Attali, 2015). Researching individual game mechanics can provide useful insights in which elements to apply and which elements to leave behind (Hanus & Fox, 2015).

This study has identified a number of additional paths for researchers to build upon. Firstly, additional studies on the individual game mechanic of points are necessary. An ideal route would be to test points in a study that is both longitudinal and large in sample size. Currently, the study in this paper is the longest gamification study that has tested only the points mechanic in isolation. However, the sample size in this study was not large enough to determine any statistical significance, even though the participants showed increases in all subscales of intrinsic motivation and class performance. A handful of other studies have tested points on larger sample sizes but did not run longitudinal studies, with the second longest points-only study being three days in duration. These studies have also found insignificant results (Attali & Arieli-Attali, 2015; Mekler et al., 2017). A next step would be to have a study on points that is both long in duration and large in sample size in an attempt to better understand the impact of points on intrinsic motivation and task performance. Hamari (2015) conducted a study on badges in a peer-to-peer marketing platform with a sample size of nearly three thousand over two-years. Perhaps, this could inspire researchers to do for points in education what Hamari did for badges in peer-to-peer marketing (see Chapter 2).

The second research path is to take a closer look at the IMI subscale for Effort/Importance. The IMI has been used by many researchers investigating gamification, and it continue offers consistency among studies (Ahn et al., 2019; Mekler et al., 2017). However, multiple studies have found conflicting results with the Effort/Importance subscale. Mekler et al. (2017) and Ahn et al. (2019) both discovered

that points seem to impact participant effort in two different ways. In essence, work quantity seems to increase while work quality often tends to decrease. The current theory is that participants deduce how to most efficiently earn the most points and may be applying minimal effort to obtain the points. If this is true, it could have strong negative implications on the underlying effects of how points, and of how gamification as a whole, might activate motivation. Going forward, it may be beneficial to modify the Effort/Importance IMI subscale to accommodate, measure, and attempt to explain what is actually happening to participant effort while undergoing gamification.

Further exploration of points with and without other game mechanics is needed. One of the possible reasons that this study on points-based grading and other studies on points in isolation resulted in insignificance may be caused by the missing social element that other gamification mechanics produce. Other popular gamification elements such as badges and leaderboards bring a social aspect to the classroom. While, some studies have determined that these social aspects may actually harm student motivation (Hanus & Fox, 2015), many other studies performed on badges and leaderboards suggest that there are benefits to these game mechanics (2016 Banfield & Wilkerson, 2014; de Rocha Seixas et al.,). The social aspect might be the element that was missing from the points-based grading system to make it function as intended. Testing a points-only group against a group with points in combination with a social game mechanic would make great strides towards greater understanding.

Conclusion

Student motivation is one of the greatest challenges that instructors face in education today (Cheong et al., 2014). This is particularly worrisome as the traditional

letter grading system used in American educational institutions since 1897 may be contributing towards harming student motivation (Durm, 1993; Zichermann, 2012). Gamification, or the application of game elements to non-game contexts, is a means to assist in increasing student engagement (Deterding, 2011; Hamari & Koivisto, 2015; Kapp 2012; Robson et al., 2015). Unfortunately, the studies on gamification in the classroom are limited (Hamari, 2015; Landers & Armstrong, 2015). Additional research must be done to investigate gamification's educational effectiveness, in particular how the individual mechanics function in a classroom setting (Hanus & Fox, 2015). Schools around the United States have already begun gamifying their programs and courses, even though educators do not have sufficient empirical evidence to back its effectiveness (Cohen, 2011; Hamari, 2015; Hanus & Fox, 2015).

The primary message to take from this study should be that it is too early to begin gamifying education with the limited information that we currently have.

The results of this research suggest that points-based grading has no significant effect on motivation in comparison to traditional letter grading and educators should be cautious about applying gamification to the classroom until the individual effects of gamification are known. This study, along with other recent studies are bringing to light the possibility that gamification may not be the motivational cure-all that many believe that it is (Attali & Arieli-Attali, 2015; Ahn et al., 2019; Humari, 2015; Mekler et al., 2017; Lieberoth, 2015). At this stage, the general opinion of gamification may be beginning to shift, as more and more empirical studies put individual mechanics of gamification under scrutiny.

The purpose of the study was to isolate the game element of points from the overall gamified education approach by applying it within a real classroom setting and to measure the impact that this grading system had on overall class performance. The results of this study did not find any significance. However, the results of this study should not completely reject a possible future for gamification in education since several new avenues of research have been described here. Many recent studies continue to shine a positive light on the effects of gamification, and there are many more aspects of gamification that should be studied before any conclusions on its educational impact can be made (Van der Kooij, van Dijsseldonk, van Veen, Steenbrink, de Weerd, & Overvliet, 2019).

It is also possible that gamification is as good as it sounds. In terms of points-based grading, the theories that explain why it works are sound in logic. Only time and further research can determine whether or not gamification can endure under the magnifying glass. Until gamification mechanics can be sufficiently investigated, educators can only hope that gamification is the motivational tool that researchers have been anticipating.

References

Ahn, S. J., Johnsen, K., & Ball, C. (2019). Points-based reward systems in gamification impact children's physical activity strategies and psychological needs. *Health Education & Behavior.* 1 – 9. https://doi.org/10.1177/1090198118818241

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education, 83,* 57-63. https://doi.org/10.1016/j.compedu.2014.12.012

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*,1-14. https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Banfield, J. & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research, 7(*4*),* 291-298. https://doi.org/10.19030/cier.v7i4.8843

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies, 70*(3), 489-520. https://doi.org/10.1111/1467-937x.00253

Callahan, R., E. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools.* Chicago: University of Chicago Press.

Cassidy S., (2004). Learning styles: An overview of theories, models, and measures. *Educational Psychology, 24*(4), 419–444. https://doi.org/10.1080/0144341042000228834

Choi, J., & Mogami, T., & Medalia A. (2009). Intrinsic motivation inventory: An adapted measure for schizophrenia research. *Schizophrenia Bulletin, 36*(5), 966-976. https://doi.org/10.1093/schbul/sbp030

Cheong, C., Filippou, J., & Cheong, F. (2014). Towards the gamification of learning: Investigating student perceptions of game elements. *Journal of information systems education, 25(*3), 233-244. https://www.learntechlib.org/p/157984/

Chen, Y., Burton, T, Mihaela, V., & Whittinghill, D. M. (2015). Cogent: A case study of meaningful gamification in education with virtual currency. *International Journal of Emerging Technologies in Learning 10*(1), 39-45. https://doi.org/10.3991/ijet.v10i1.4247

Cohen, A. M. (2011). The gamification of education. *Futurist*. 45. 16-17.

Darnon, C., Harackiewicz, J. M., Butera, F., & Quiamzade G. M. A., (2007). Performance-approach and performance avoidance goals: When uncertainty makes a difference. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167207301022

Deci, E.L., & Ryan, R.M. (2005). Intrinsic motivation inventory (IMI). Retrieved from http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness: Defining "gamification"*. Paper presented at the meeting of the eleventh annual academic Mindtrek conference at Tampere, Finland. 1 – 7. https://doi.org/10.1145/2181037.2181040

Deterding, S. (2011). *Gamification: Using game elements in non-gaming contexts*. Paper presented at that CHI Workshop 2011, Vancouver, BC, Canada.

Deterding, S. (2011). *Situated motivational affordances of game elements: A conceptual model.* Paper presented at that CHI Workshop 2011, Vancouver, BC, Canada.

Deterding, S. (2012). Gamification: Designing for motivation. *Interactions*, *19*(4), 14-17. https://doi.org/10.1145/2212877.2212883

De Rocha Seixas, L. R., Gomes, A. S., & De Melo Filho, I. J. (2016). Effectiveness of gamification in the engagement of students. *Computers in Human Behavior, 58,* 48-63. https://doi.org/10.1016/j.chb.2015.11.021

Durm, M. W. (1993). An A is not an A is not an A: A history of grading. *The Educational Forum*, *57*, 294-297. Retrieved from http://www.indiana.edu/~educy520/sec6342/week_07/durm93.pdf

Dueck, M. (2014). *Grading smarter, not harder: assessment strategies that motivate kids and help them learn.* Alexandria, VA: ASCD.

Elliot, A.J., & Church, M.A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social P*sychology, 72, 218-232. https://doi.org/10.1016/b978-012372545-5/50005-8

Entertainment Software Association (2016). Essential facts about the computer and video game industry. Retrieved from http://essentialfacts.theesa.com/Essential-Facts-2016.pdf

Feger, S. S., Dallmeier-Tiessen, S., Woźniak, P. W., & Schmidt, A., (2019). Gamification in science: A study of requirements in the context of reproducible research. *Conference on Human Factors in Computing Systems (CHI 2019).* https://doi.org/10.1145/3290605.3300690

Field, A. P. (2009). Discovering statistics using SPSS (and sex and drugs and rock 'n' roll). Thousand Oaks, CA: SAGE Publications. https://doi.org/10.1024/1012-5302/a000397

Flores, J. F. F. (2015). Using gamification to enhance second language learning. *Digital Education Review, 27,* 32-54. Retrieved from https://files.eric.ed.gov/fulltext/EJ1065005.pdf

Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership, 52*(2), 14–20. Retrieved from https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1012&context=edp_facpub

Guskey, T. R. (2004). Zero alternatives. *Principle Leadership, 5*(2), 49-53. Retrieved from https://tguskey.com/wp-content/uploads/Grading-5-0-Alternatives.pdf

Hamari, J., & Koivisto, J. (2015). Why do people use gamification services? *International Journal of Information Management, 35,* 419-431. https://doi.org/10.1016/j.ijinfomgt.2015.04.006

Hamari, J. (2015). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior, 71,* 1-10. https://doi.org/10.1016/j.chb.2015.03.036

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152-161. https://doi.org/10.1016/j.compedu.2014.08.019

Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The effect of achievement badges

on students' behavior: An empirical study in a university-level computer science course. *International Journal of Emerging Technologies in Learning. 10*(1), 18-28. https://doi.org/10.3991/ijet.v10i1.4221

Jonassen, D.H., Carr, C., & Yueh, H-P. (1998). Computers as mindtools for engaging learners in critical thinking. *TechTrends, 43*, 24-32. https://doi.org/10.1007/bf02818172

Kapp, K. M. (2012). The gamification of learning and instruction: Game-based methods and strategies for training and education. San Francisco, CA: Pfeiffer.

Kim, B. (2015). Understanding gamification. *Library Technology Reports, 51(2)*, 1-36. Retrieved from https://journals.ala.org/index.php/ltr/article/view/5629

Kohn, A. (1999). From degrading to de-grading. *High School Magazine, 6*(5), 1999, 38-43. Retrieved from https://www.alfiekohn.org/article/degrading-de-grading/

Kohn, A. (2011). The case against grades. *Educational Leadership, 69*(3), 28-33. Retrieved from https://eric.ed.gov/?id=EJ963095

Kuo, M., & Chuang, T. (2016). How gamification motivates visits and engagement for online academic dissemination: An empirical study. *Computers in Human Behavior, 55,* 16-27. https://doi.org/10.1016/j.chb.2015.08.025

Landers, R. N. & Armstrong, M. B. (2015). Enhancing instructional outcomes with gamification: An empirical test of the technology-enhanced training effectiveness model. *Computers in Human Behavior, 71*, 499-507. https://doi.org/10.1016/j.chb.2015.07.031

Lieberoth, A. (2015). Shallow gamification: Testing psychological effects of framing an

activity as a game. *Games and Culture*, *10*(3), 229-248 https://doi.org/10.1177/1555412014559978

Lister, M. C. (2015). Gamification: The effect on student motivation and performance at the post-secondary level. *Issues and Trends in Educational Technology 3*(2), 2-23. https://doi.org/10.2458/azu_itet_v3i2_lister

Maehr, M.L., & Midgley, C. (1996). *Transforming school cultures.* Boulder, CO: Westview.

Marin, V., Lopez, M., & Maldonado, G. (2015). Can gamification be introduced within primary classes? *Digital Education. 27,* 55-58. http://dx.doi.org/10.1344/der.2015.27.55-68

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD

Mekler, E. D., Bruhlmann, F., Opwis, K. & Tuch, A. N. (2013). Disassembling gamification: The effects of points and meaning on user motivation and performance. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. 1137 – 1142. https://doi.org/10.1145/2468356.2468559

Mekler, E. D., Bruhlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior, in press,* 1-10. http://dx.doi.org/10.1016/j.chb.2015.08.048

Monteiro, V., Mata, L., & Francisco, P. (2015). Intrinsic motivation inventory: Psychometric properties in the context of the first language and mathematics learning. *Psicologia: Reflexão e Crítica. 28*(3), 434-443. http://dx.doi.org/10.1590/1678-7153.201528302

Renauld, C. & Wagoner, B. (2011). The gamification of learning. *Principal Leadership, 12(2),* 56-59. Retrieved from https://eric.ed.gov/?id=EJ965250

Papp, T. A. (2017). Gamification effects on motivation and learning: Application to primary and college students. *International journal for Cross Disciplinary Subjects in Education*, 8(3), 3193-3201. Retrieved from https://infonomics-society.org/wp-content/uploads/ijcdse/published-papers/volume-8-2017/Gamification-Effects-on-Motivation-and-Learning.pdf

Peng, W., Crouse, J., & Lin, J.-H. (2013). Using active video games for physical activity promotion: A systematic review of the current state of research. *Health Education & Behavior*, 40, 171-192. http://dx.doi.org/10.1177/1090198112444956

Pope, Z. C., Lewis, B. A., & Gao, Z. (2015). Using the transtheoretical model to examine the effects of exergaming on physical activity among children. *Journal of Physical Activity and Health*, 12, 1205-1212. http://dx.doi.org/10.1123/jpah.2014-0310

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. Journal of *Educational Psychology, 103*(3), 683-700. http://dx.doi.org/10.1037/a0023911

Robson, K., Plangger, K., & Kietzmann, J. H., McCarthy, I., & Pitt, L. (2016). Game on: Engaging customers and employees through gamification. *Business Horizons*, *59,* 29-36. https://doi.org/10.1016/j.bushor.2015.08.002

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, *58,* 411-420. https://doi.org/10.1016/j.bushor.2015.03.006

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68. https://doi.org/10.1037/0003-066x.55.1.68

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychology need satisfaction. *Computers in Human Behavior, 69*, 371-380. http://dx.doi.org/10.1016/j.chb.2016.12.033

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74,* 14-31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Sheldon, L. (2011). *The multiplayer classroom: Designing coursework as a game*. Boston, MA: Cengage Learning.

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE Life Sciences Education, 13*(2), 159-166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Free Press.

Shute, V. J., Ventura, M., & Ke F. F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers and Education, 80*, 58-67. https://doi.org/10.1016/j.compedu.2014.08.013

Thomas, D., & Brown, J. S. (2011). *A new culture of learning: Cultivating the imagination for a world of constant change*. Lexington, KY: CreateSpace. https://doi.org/10.1080/00393541.2015.11518969

Tocci, C., (2010). An immanent machine: Reconsidering grades, historical and present. *Educational Philosophy and Theory, 42*(7), 762-778, http://dx.doi.org/10.1111/j.1469-5812.2008.00440.x

Turan, Z, Avinc, Z., Kara, K., & Goktas, Y. (2016). Gamification and education: Achievements, cognitive loads, and views of students. *International Journal of Emerging Technologies in Learning, 11*(7). 64-69. https://doi.org/10.3991/ijet.v11i07.5455

Urh, M., Vukovic, G., Jereb, E., & Pintar, R. (2015). The model for introduction of gamification into e-learning in higher education. *Procedia – Social and Behavioral Sciences, 197,* 388-397. https://doi.org/10.1016/j.sbspro.2015.07.154

U.S. Census Bureau. (2018). More than 76 million students enrolled in U.S. schools, census bureau reports. (Report No. CB18-192). Retrieved from https://www.census.gov/newsroom/press-releases/2018/school-enrollment.html

Van der Kooij K., van Dijsseldonk R., van Veen M., Steenbrink F., de Weerd C. & Overvliet K. E., (2019). Gamification as a sustainable source of enjoyment during balance and gait exercises. *Frontiers in Psychology, 10*(294). https://doi.org/10.3389/fpsyg.2019.00294

Zichermann, G. (2012). *Changing the game in education: Gabe Zichermann at TEDxBerlin*. Retrieved from http://tedxtalks.ted.com/video/TEDxBerlin-Gabe-Zichermann-Chan;search%3Agabe%20zichermann

Zichermann, G. & Linder, J. (2013). *The gamification revolution: How leaders leverage game mechanics to crush the competition*. New York, NY: McGraw Hill Education.

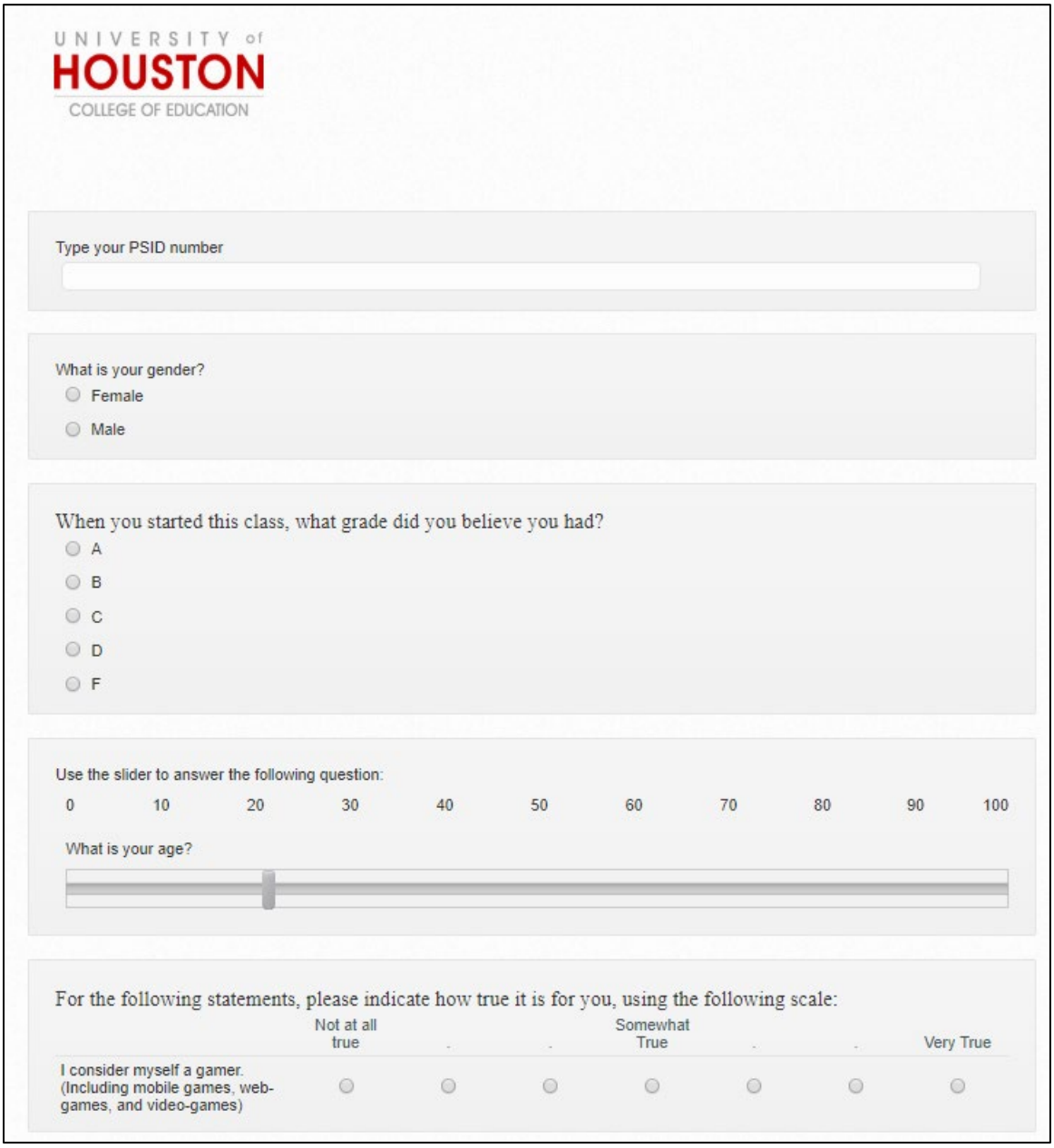
Appendix A

Questionnaires

In this appendix you will find the four questionnaires that were be used as the measurement instruments for the Intrinsic Motivation Inventory portion of the study. Furthermore, the links to the original questionnaires in the Qualtrics experience management system are provided. The pre-tests are identical between the two groups. In the post-test the treatment group's questionnaire has additional questions on the student's opinion of the points based grading system.

Pre-Test Control Group Questionnaire

Qualtrics link: https://coeuh.co1.qualtrics.com/jfe/form/SV_cBHvDWi4B8Cl6rb



UNIVERSITY of HOUSTON
COLLEGE OF EDUCATION

Type your PSID number

What is your gender?
- Female
- Male

When you started this class, what grade did you believe you had?
- A
- B
- C
- D
- F

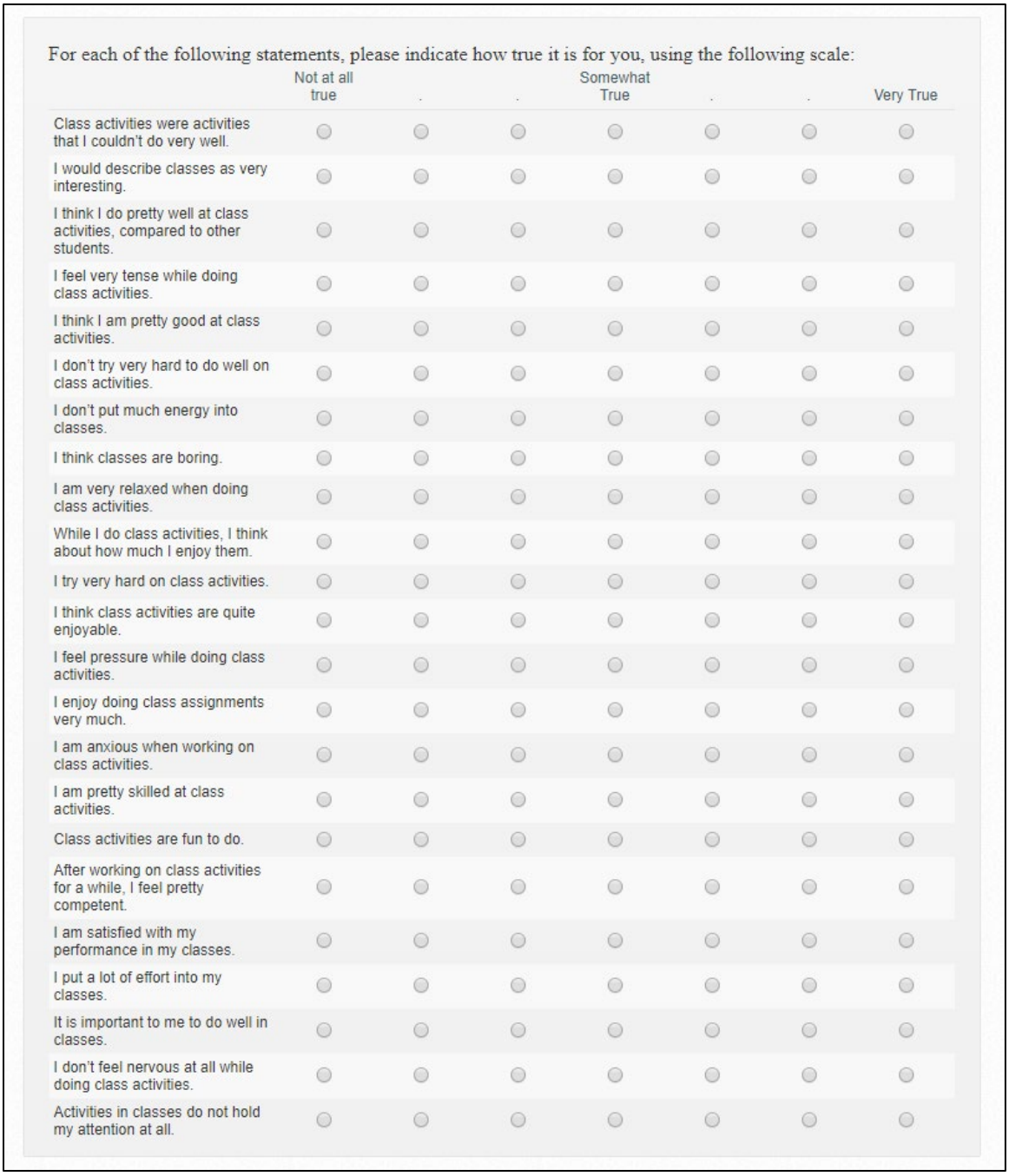
Use the slider to answer the following question:

0 10 20 30 40 50 60 70 80 90 100

What is your age?

For the following statements, please indicate how true it is for you, using the following scale:

	Not at all true			Somewhat True			Very True
I consider myself a gamer. (Including mobile games, web-games, and video-games)							



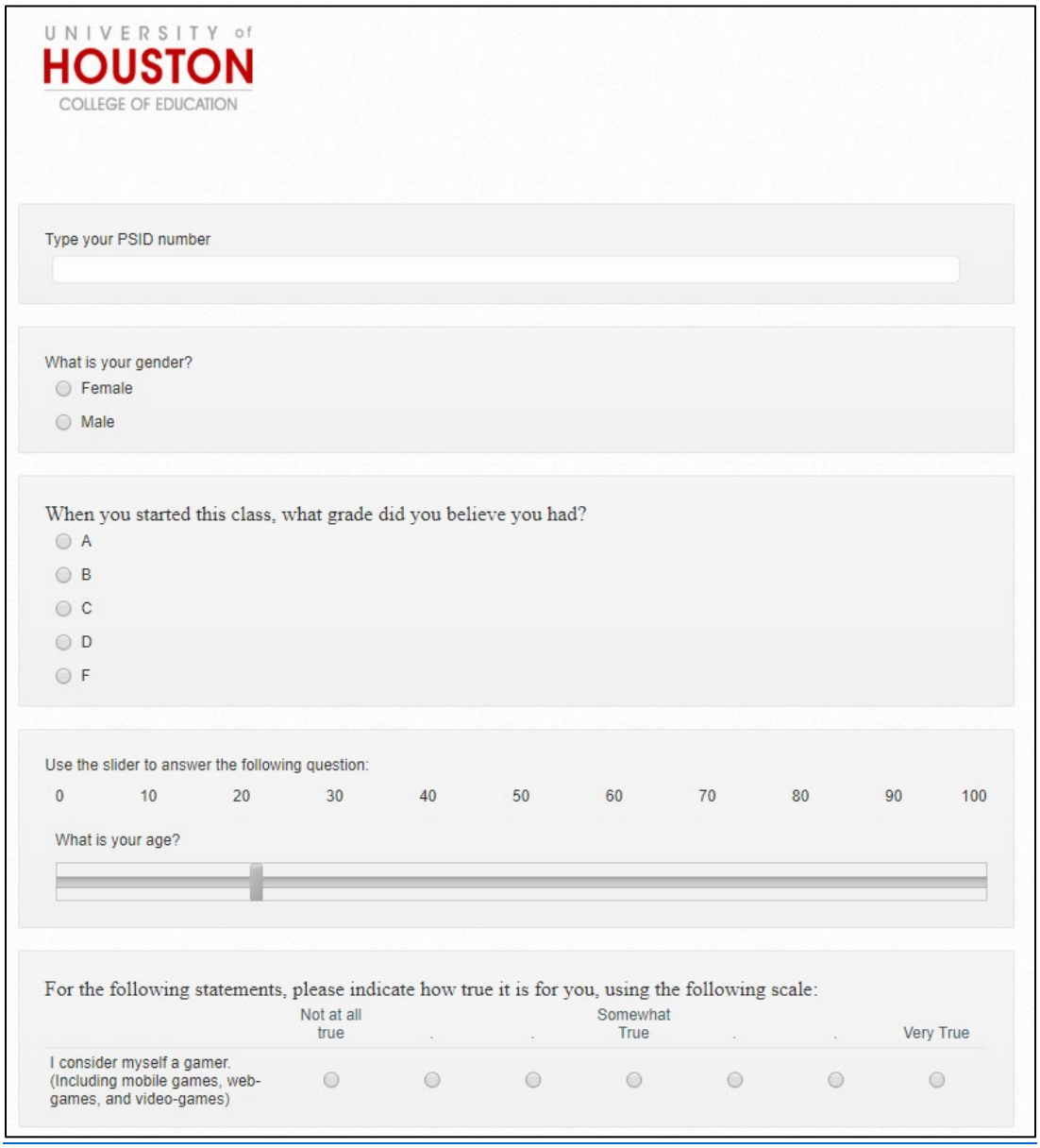
For each of the following statements, please indicate how true it is for you, using the following scale:

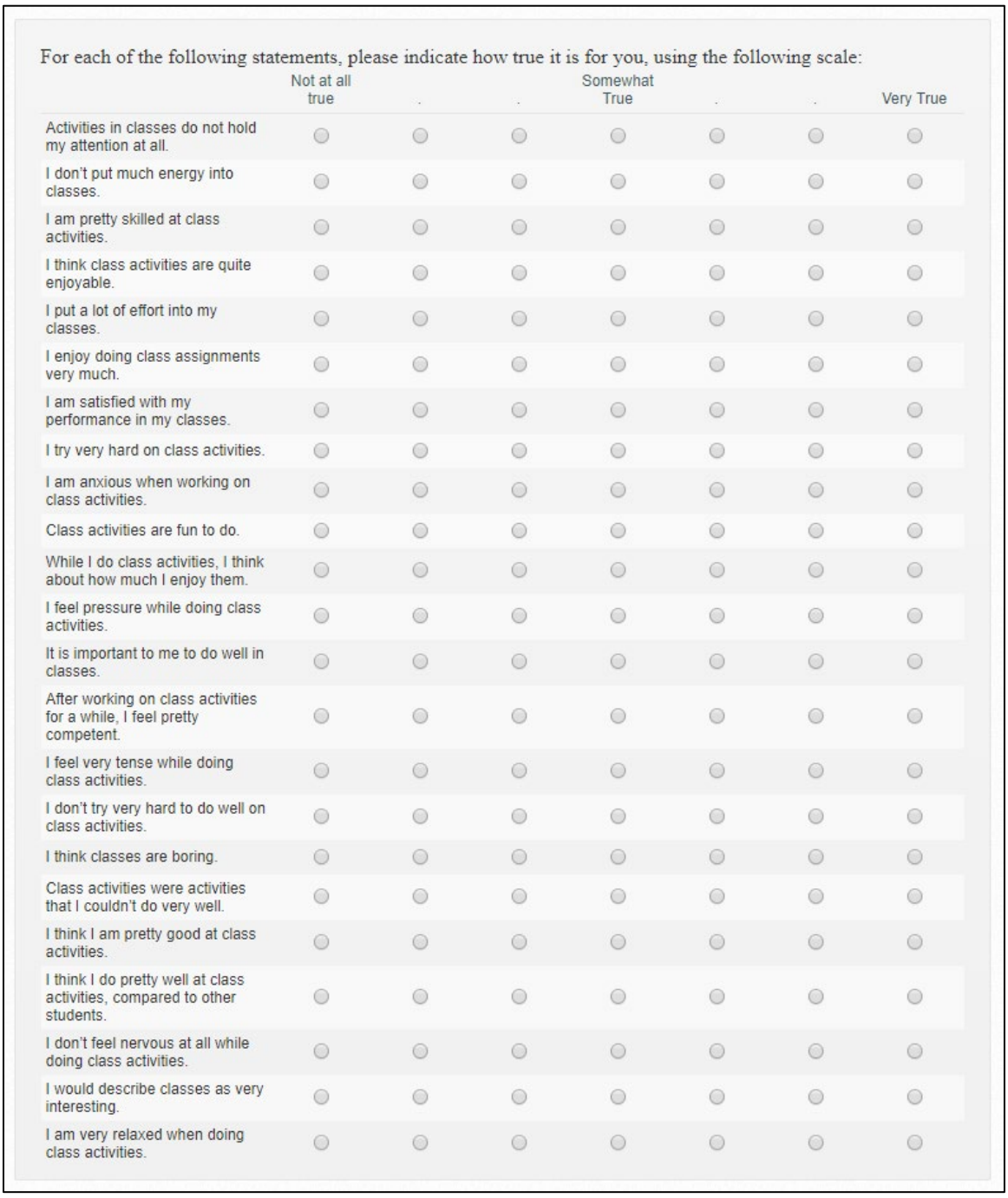
	Not at all true	.	.	Somewhat True	.	.	Very True
Class activities were activities that I couldn't do very well.	○	○	○	○	○	○	○
I would describe classes as very interesting.	○	○	○	○	○	○	○
I think I do pretty well at class activities, compared to other students.	○	○	○	○	○	○	○
I feel very tense while doing class activities.	○	○	○	○	○	○	○
I think I am pretty good at class activities.	○	○	○	○	○	○	○
I don't try very hard to do well on class activities.	○	○	○	○	○	○	○
I don't put much energy into classes.	○	○	○	○	○	○	○
I think classes are boring.	○	○	○	○	○	○	○
I am very relaxed when doing class activities.	○	○	○	○	○	○	○
While I do class activities, I think about how much I enjoy them.	○	○	○	○	○	○	○
I try very hard on class activities.	○	○	○	○	○	○	○
I think class activities are quite enjoyable.	○	○	○	○	○	○	○
I feel pressure while doing class activities.	○	○	○	○	○	○	○
I enjoy doing class assignments very much.	○	○	○	○	○	○	○
I am anxious when working on class activities.	○	○	○	○	○	○	○
I am pretty skilled at class activities.	○	○	○	○	○	○	○
Class activities are fun to do.	○	○	○	○	○	○	○
After working on class activities for a while, I feel pretty competent.	○	○	○	○	○	○	○
I am satisfied with my performance in my classes.	○	○	○	○	○	○	○
I put a lot of effort into my classes.	○	○	○	○	○	○	○
It is important to me to do well in classes.	○	○	○	○	○	○	○
I don't feel nervous at all while doing class activities.	○	○	○	○	○	○	○
Activities in classes do not hold my attention at all.	○	○	○	○	○	○	○

Figure 7. Pre-Test Control Group Questionnaire.

Pre-Test Treatment Questionnaire

Qualtrics Link:
https://coeuh.co1.qualthttps://coeuh.co1.qualtrics.com/jfe/form/SV_b79CrUyHM3vxOZ





For each of the following statements, please indicate how true it is for you, using the following scale:

	Not at all true	.	.	Somewhat True	.	.	Very True
Activities in classes do not hold my attention at all.	○	○	○	○	○	○	○
I don't put much energy into classes.	○	○	○	○	○	○	○
I am pretty skilled at class activities.	○	○	○	○	○	○	○
I think class activities are quite enjoyable.	○	○	○	○	○	○	○
I put a lot of effort into my classes.	○	○	○	○	○	○	○
I enjoy doing class assignments very much.	○	○	○	○	○	○	○
I am satisfied with my performance in my classes.	○	○	○	○	○	○	○
I try very hard on class activities.	○	○	○	○	○	○	○
I am anxious when working on class activities.	○	○	○	○	○	○	○
Class activities are fun to do.	○	○	○	○	○	○	○
While I do class activities, I think about how much I enjoy them.	○	○	○	○	○	○	○
I feel pressure while doing class activities.	○	○	○	○	○	○	○
It is important to me to do well in classes.	○	○	○	○	○	○	○
After working on class activities for a while, I feel pretty competent.	○	○	○	○	○	○	○
I feel very tense while doing class activities.	○	○	○	○	○	○	○
I don't try very hard to do well on class activities.	○	○	○	○	○	○	○
I think classes are boring.	○	○	○	○	○	○	○
Class activities were activities that I couldn't do very well.	○	○	○	○	○	○	○
I think I am pretty good at class activities.	○	○	○	○	○	○	○
I think I do pretty well at class activities, compared to other students.	○	○	○	○	○	○	○
I don't feel nervous at all while doing class activities.	○	○	○	○	○	○	○
I would describe classes as very interesting.	○	○	○	○	○	○	○
I am very relaxed when doing class activities.	○	○	○	○	○	○	○

Figure 8. Pre-Test Treatment Questionnaire.

Post-Test Control Questionnaire

Qualtrics link: https://coeuh.co1.qualtrics.com/jfe/form/SV_0emH6oIkrMoN0c5

Type your PSID number

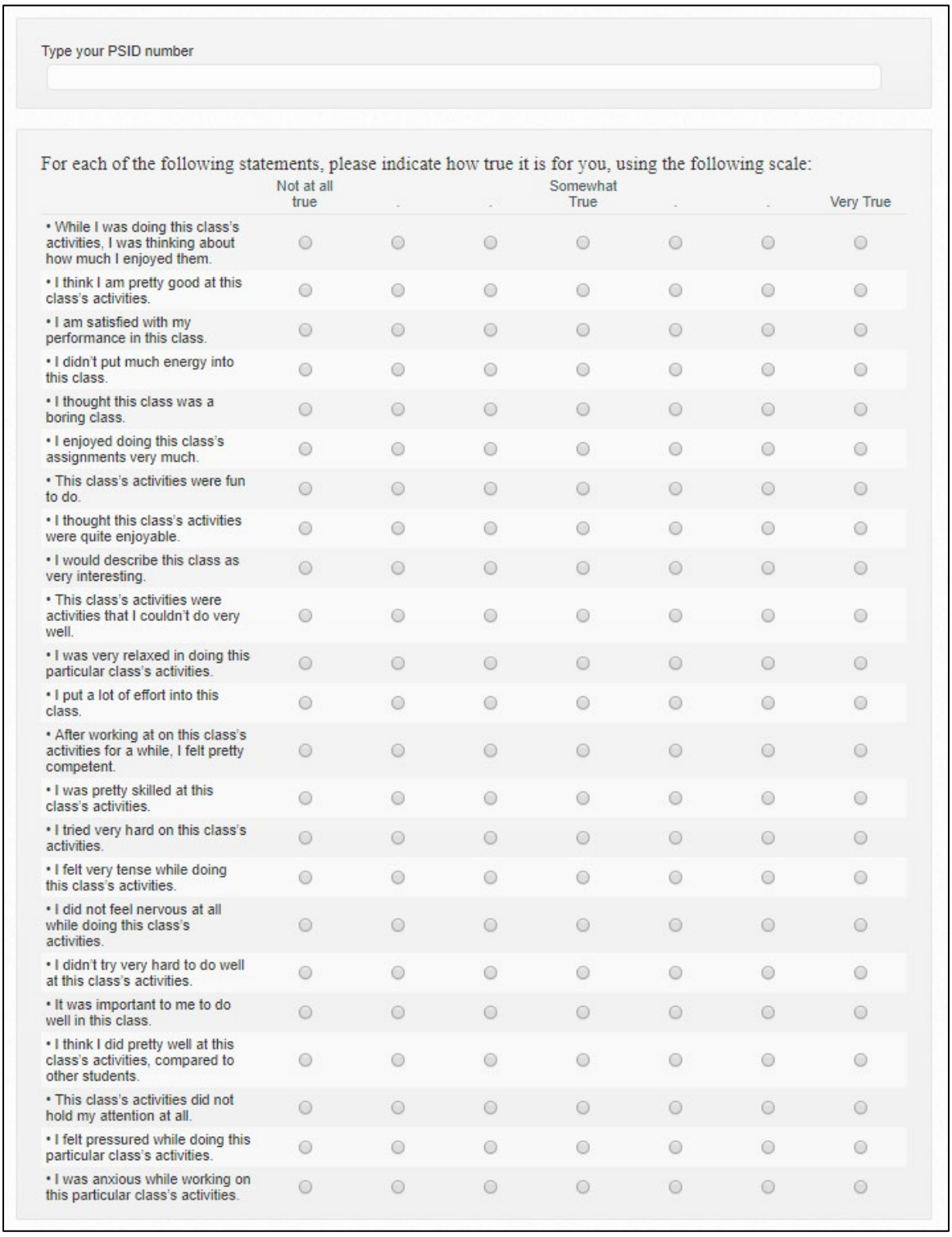
For each of the following statements, please indicate how true it is for you, using the following scale:

	Not at all true			Somewhat True			Very True
• I enjoyed doing this class's assignments very much.							
• This class's activities were fun to do.							
• I thought this class was a boring class.							
• This class's activities did not hold my attention at all.							
• I would describe this class as very interesting.							
• I thought this class's activities were quite enjoyable.							
• While I was doing this class's activities, I was thinking about how much I enjoyed them.							
• I think I am pretty good at this class's activities.							
• I think I did pretty well at this class's activities, compared to other students.							
• After working at on this class's activities for a while, I felt pretty competent.							
• I am satisfied with my performance in this class.							
• I was pretty skilled at this class's activities.							
• This class's activities were activities that I couldn't do very well.							
• I put a lot of effort into this class.							
• I didn't try very hard to do well at this class's activities.							
• I tried very hard on this class's activities.							
• It was important to me to do well in this class.							
• I didn't put much energy into this class.							
• I did not feel nervous at all while doing this class's activities.							
• I felt very tense while doing this class's activities.							
• I was very relaxed in doing this particular class's activities.							
• I was anxious while working on this particular class's activities.							
• I felt pressured while doing this particular class's activities.							

Figure 9. Post-Test Control Questionnaire.

Post-Test Treatment Group Questionnaire

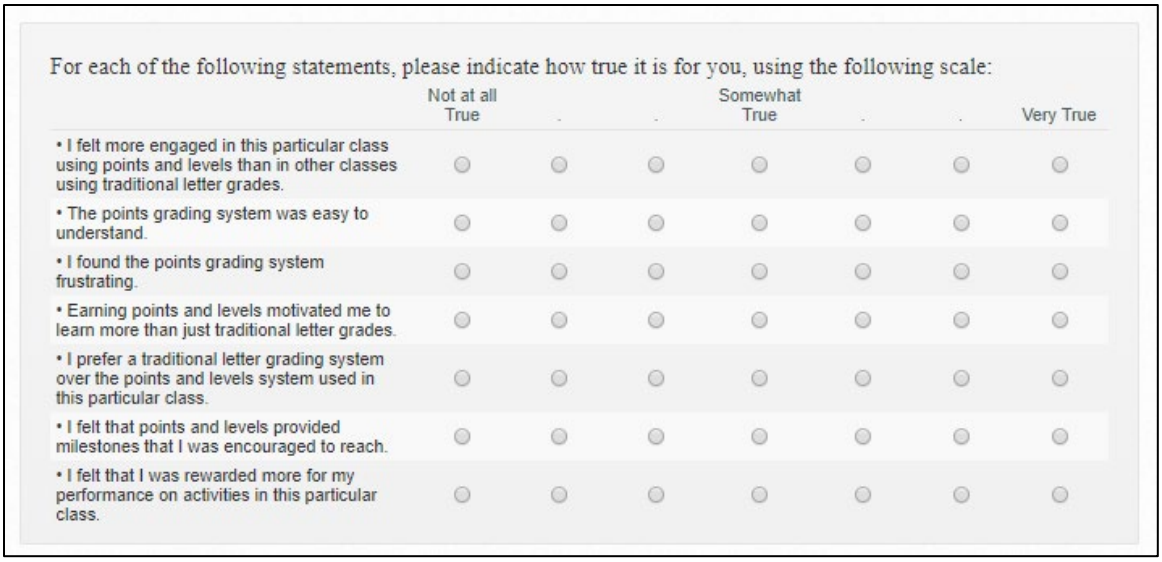
Qualtrics Link: https://coeuh.co1.qualtrics.com/jfe/form/SV_cZwHVvxe9S0ZkMt



Type your PSID number

For each of the following statements, please indicate how true it is for you, using the following scale:

	Not at all true	.	.	Somewhat True	.	.	Very True
• While I was doing this class's activities, I was thinking about how much I enjoyed them.	○	○	○	○	○	○	○
• I think I am pretty good at this class's activities.	○	○	○	○	○	○	○
• I am satisfied with my performance in this class.	○	○	○	○	○	○	○
• I didn't put much energy into this class.	○	○	○	○	○	○	○
• I thought this class was a boring class.	○	○	○	○	○	○	○
• I enjoyed doing this class's assignments very much.	○	○	○	○	○	○	○
• This class's activities were fun to do.	○	○	○	○	○	○	○
• I thought this class's activities were quite enjoyable.	○	○	○	○	○	○	○
• I would describe this class as very interesting.	○	○	○	○	○	○	○
• This class's activities were activities that I couldn't do very well.	○	○	○	○	○	○	○
• I was very relaxed in doing this particular class's activities.	○	○	○	○	○	○	○
• I put a lot of effort into this class.	○	○	○	○	○	○	○
• After working at on this class's activities for a while, I felt pretty competent.	○	○	○	○	○	○	○
• I was pretty skilled at this class's activities.	○	○	○	○	○	○	○
• I tried very hard on this class's activities.	○	○	○	○	○	○	○
• I felt very tense while doing this class's activities.	○	○	○	○	○	○	○
• I did not feel nervous at all while doing this class's activities.	○	○	○	○	○	○	○
• I didn't try very hard to do well at this class's activities.	○	○	○	○	○	○	○
• It was important to me to do well in this class.	○	○	○	○	○	○	○
• I think I did pretty well at this class's activities, compared to other students.	○	○	○	○	○	○	○
• This class's activities did not hold my attention at all.	○	○	○	○	○	○	○
• I felt pressured while doing this particular class's activities.	○	○	○	○	○	○	○
• I was anxious while working on this particular class's activities.	○	○	○	○	○	○	○



For each of the following statements, please indicate how true it is for you, using the following scale:

	Not at all True			Somewhat True			Very True
• I felt more engaged in this particular class using points and levels than in other classes using traditional letter grades.	○	○	○	○	○	○	○
• The points grading system was easy to understand.	○	○	○	○	○	○	○
• I found the points grading system frustrating.	○	○	○	○	○	○	○
• Earning points and levels motivated me to learn more than just traditional letter grades.	○	○	○	○	○	○	○
• I prefer a traditional letter grading system over the points and levels system used in this particular class.	○	○	○	○	○	○	○
• I felt that points and levels provided milestones that I was encouraged to reach.	○	○	○	○	○	○	○
• I felt that I was rewarded more for my performance on activities in this particular class.	○	○	○	○	○	○	○

Figure 10. Post-Test Treatment Group Questionnaire.

Appendix B

Consent to Take Part in a Human Research Study Materials

In this appendix you will find the consent form that were distributed to the participants at the beginning of the study. Furthermore, you will find the script that was used by the researcher's supervisor which was given prior to distributing the consent forms. The consent form was developed using the template provided by the university's institutional review board (form HRP-502a)

Consent Form:

Title of research study: Using a Gamified Points-Based Grading System in Technology Courses for Pre-Service Teachers

Investigator: ***--- under the supervision of ---***

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are an undergraduate student.

What should I know about a research study?

Someone will explain this research study to you.

Whether or not you take part is up to you.

You can choose not to take part.

You can agree to take part and later change your mind.

Your decision will not be held against you.

You can ask all the questions you want before you decide, and can ask questions at any time during the study.

Why is this research being done?

The traditional letter grading scale used in America, as well as many other countries in the world, is potentially harming student motivation. Through gamification, an alternate grading system is proposed that would remove this potential barrier to student achievement. Points-based grading is a shift in how students perceive their advancement in a course. This is intended to grant positive motivation. The purpose of this study is to measure the magnitude of these effects on students' motivation.

How long will the research last?

This research study will last for the remainder of the Semester.

How many people will be studied?

We expect to enroll about 80 people in this research study.

What happens if I say yes, I want to be in this research?

If you agree to participate in this research study you will be asked to complete a questionnaire that measures your motivation levels. During the last week of class, April, 30th to May 2nd, you will be asked to complete a second questionnaire to determine whether any changes may have occurred. Each questionnaire is expected to take no more than ten minutes to complete. Additionally, your name and grades will be collected for this study to examine whether the alternative grading system has any impact on scholastic performance. However, your identity will remain fully confidential and the data will not be able to link back to you.

All questionnaires will be distributed and collected in person by [redacted]. Your instructor will have no knowledge of whether you decided to participate or of your questionnaire answers during the semester.

What happens if I do not want to be in this research?

You can choose not to take part in the research and it will not be held against you. Choosing not to take part will involve no penalty or loss of benefit to which you are otherwise entitled. A decision to take part or not, or to withdraw from the research will have no effect on your grades or standing with the University of Houston.

What happens if I say yes, but I change my mind later?

You can leave the research at any time it will not be held against you. If you decide to leave the research, contact the investigator so that the investigator can exclude you from the treatment groups. When you decide to leave all data collected from you will be excluded from the final analysis and no further data will be collected.

Is there any way being in this study could be bad for me?

There are no foreseeable risks related to the procedures conducted as part of this study. If you choose to take part and undergo a negative event you feel is related to the study, please inform the study team.

Will I get anything for being in this study?

No.

Will being in this study help me in any way?

We cannot promise any benefits to you or others from your taking part in this research. However, points-based grading is believed to improve students' motivation. If the theories on gamification are correct, you should be more engaged in the course and therefore learn more from the class and obtain greater mastery of the content. However, the magnitude of the benefits is untested and unknown, which has led to this research project.

What happens to the information collected for the research?

Your taking part in this project is confidential. The information you provide cannot be linked to your identity. We may publish the results of this research. However we will keep your name and other identifying information confidential.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, you should talk to the research team at:

Supervisor: ---

Phone Number: ---

Email: ---

Principal Investigator: ---

Phone Number: ---

Email: ---

This research has been reviewed and approved by the University of Houston Institutional Review Board (IRB). You may also talk to them at (713) 743-9204 or cphs@central.uh.edu if:

Your questions, concerns, or complaints are not being answered by the research team.

You cannot reach the research team.

You want to talk to someone besides the research team.

You have questions about your rights as a research subject.

You want to get information or provide input about this research.

Signature Block for Capable Adult

Your signature documents your consent to take part in this research.

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	

Script for the Consent Process

This script is to be read by the research supervisor prior to distributing the consent forms.

Good afternoon,

My name is ----, I am here to invite you guys to participate in a research project that your instructor and I have been working on. We are studying the effects of various grading systems and how they potentially impact the intrinsic motivation of students. The letter grading system that we use here in America has been potentially harming student motivation and so in this study we will be looking at an alternative grading system that could lower this potential barrier to student achievement. It's called the points-based grading system, which is theorized to grant positive motivation to students. The purpose of this study is to measure the magnitude of these effects on students' motivation.

If you agree to participate in this study you will be given two short surveys to complete. One right now and one at the end of the semester. These questionnaires measure your intrinsic motivation and will be used to figure out whether any change has occurred. Additionally, we will also collect your name and grades for this study to examine whether the grading systems have any impact on scholastic performance. However, **your identity will remain fully confidential and the data will not be able to link back to you**.

This study is minimal risk, meaning that there will are no foreseeable risks related to participation. All questionnaires will be distributed and collected in person by me,

which means that all of the data will be kept secret from [redacted] until after the semester ends. **Your instructor will have no knowledge of whether you decided to participate or of your questionnaire answers**.

This study will last from today to the end of the semester.

You can find all of these details about the study in the consent form. If you find this study to be agreeable, please sign and date on the final page. If you do not wish to participate, that's fine too, you can leave your form blank. Also you can leave the study at any time, just send me an Email and I'll erase your data from the study. You can find my contact information on the consent form. So you can change your mind at any time and that's not a problem.

Let me know if you have any questions.

Appendix C

Data Analysis Assumptions Testing

This appendix displays the tests of assumptions for the multivariate analysis of variances procedures for research question (1) How do undergraduate technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of intrinsic motivation? Assumption testing is also shown for research question (2): How do undergraduate technology courses for preservice teachers taught using a points-based grading system compare to classes taught using a traditional letter grading system in terms of class performance?

Pre-Test Assumptions Analyses

In order to be able to directly compare the post-test data, tests first needed to be performed to determine whether the base IMI scores at the beginning of the treatment period were not statistically significantly different between the treatment and control groups. A MANOVA was first conducted on the pre-test questionnaire IMI scores (PreIMI) between the two groups (pre-test treatment, n = 20) (pre-test control, *n* = 29). Additionally, preliminary analyses were first conducted to test several assumptions: (1) Assumption of no univariate or multivariate outliers, (2) Assumption of multivariate normality, (3) Assumption of no multicollinearity, (4) Assumption of a linear relationship between the dependent variables for each group of the independent variable, (5) Assumption of no univariate or multivariate outliers, (6) Assumption of homogeneity of variance-covariance matrices, and (7) Assumption of homogeneity of variances. The results of the pre-test assumption analysis can be found in this appendix.

Outlier testing was assessed using box-plots (see Figure 11), and eight outliers were found across the four subscales and two groups. Six of the outliers were mild ($p < 0.01$), three per group and two outliers were extreme ($p > 0.01$). However, I believe that these perspectives are answered genuinely by the student participants and are realistic. By removing the outliers the results would be less reflective of the real world application of this research. As a result, the outlier data were kept for the analysis and no outliers were removed.

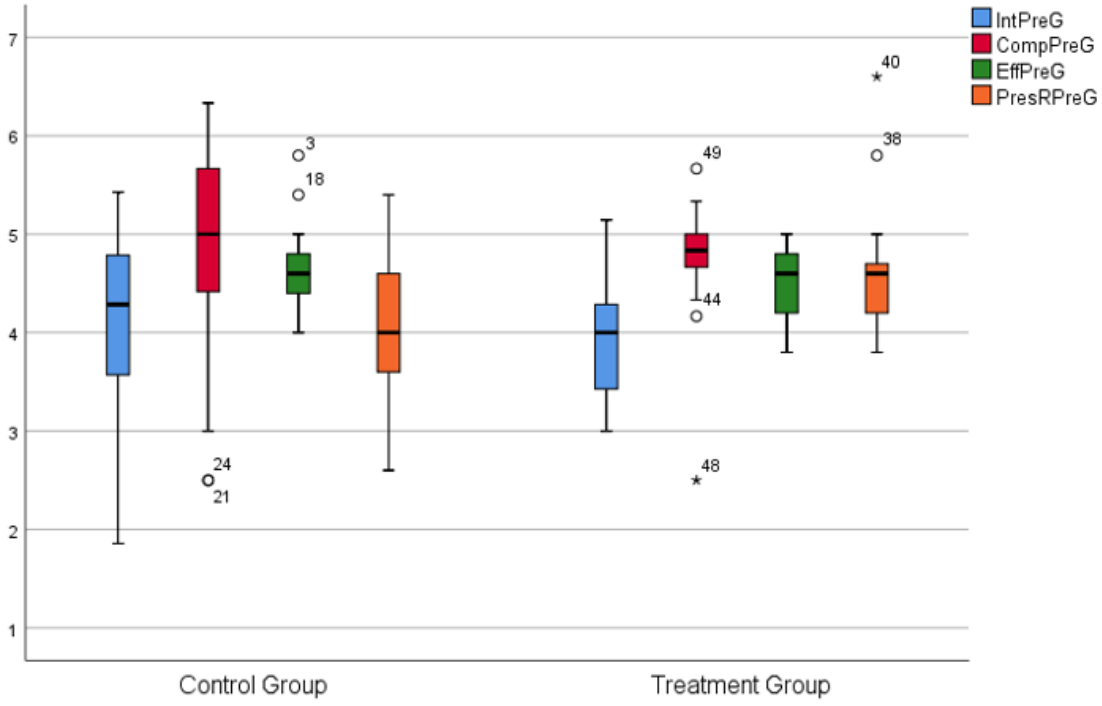
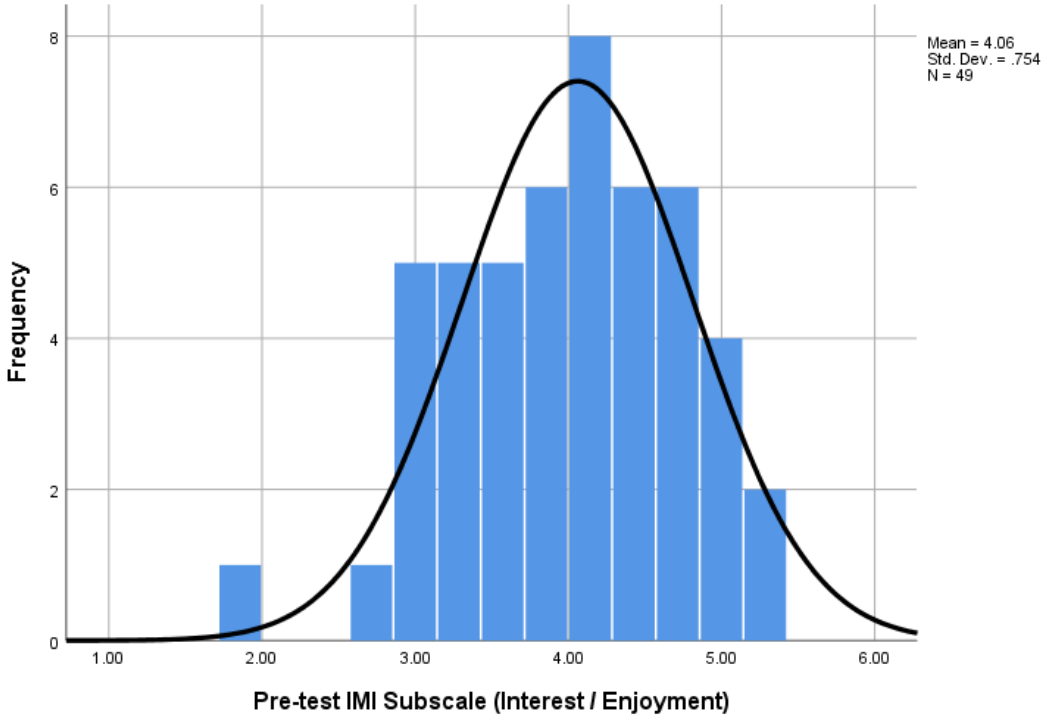
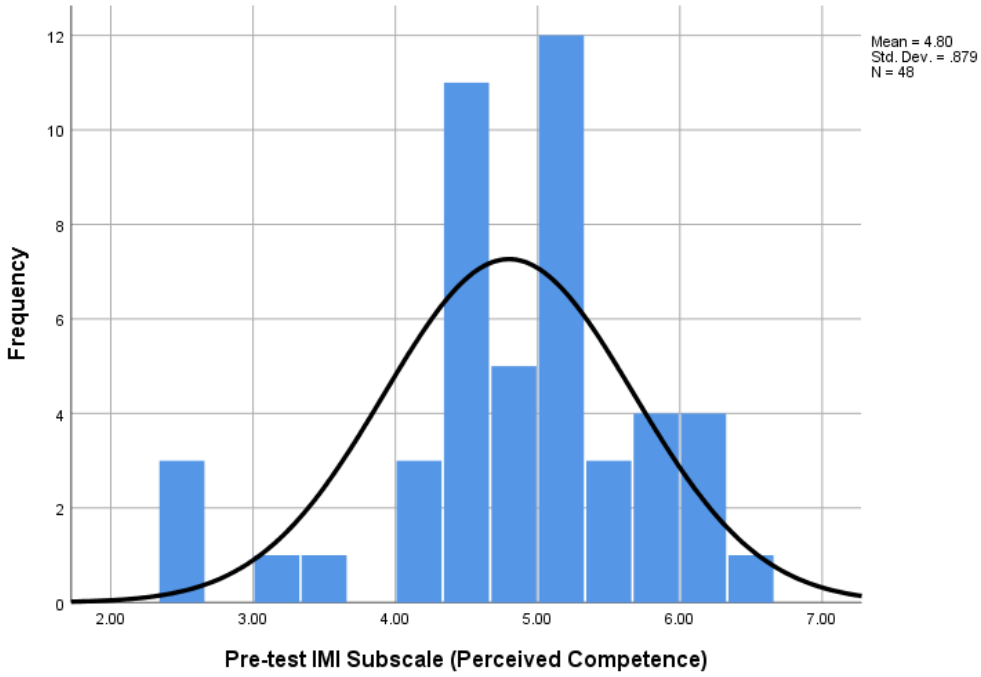


Figure 11. Boxplots of the pre-test IMI subscales: Interest / Enjoyment, Perceived Competence, Effort / Importance, and Pressure Tension. Created to determine and assess outliers.

In addition to the univariate outliers, tests were also performed to assess multivariate outliers by calculating each participant's Mahalanobis distance. There were no multivariate outliers in the data, as assessed by Mahalanobis distance ($p > .001$).

Preliminary assumption checking also revealed that data was normally distributed. Distribution was confirmed by histogram graphing through visual analysis.





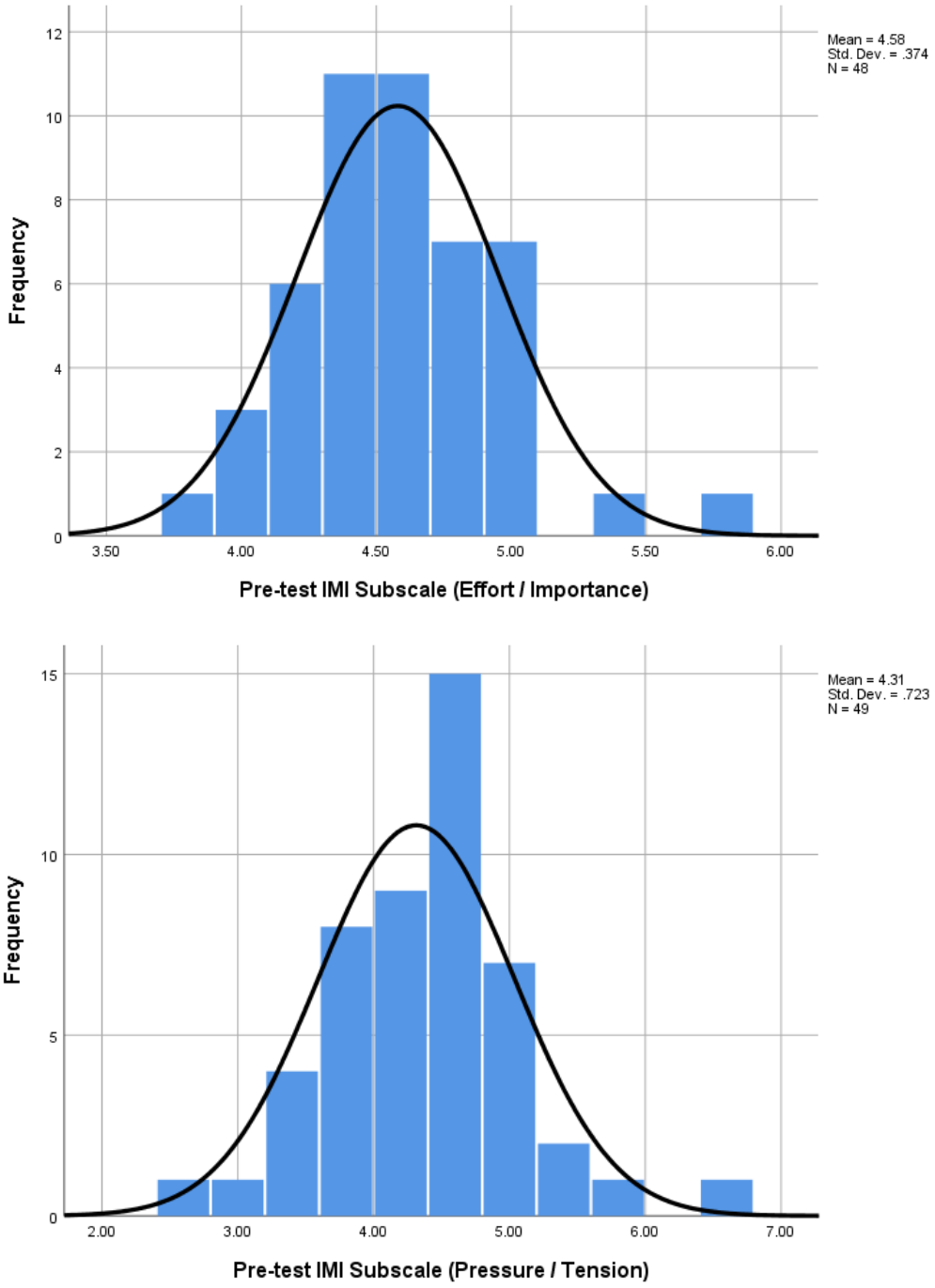


Figure 12. Histogram graphs of the Pre-Test IMI Subscales with a normal curve overlay. These graphs show that the pre-test subscales are normally distributed.

Tests for Multicollinearity were conducted next using Pearson correlation r.

There was no multicollinearity, as assessed by Pearson correlation (see Table 12). All

Pearson Correlation value are within the acceptable range of r < 0.8 (Field, 2009). These correlations indicate that there was no multicollinearity.

Table 12

Pearson Correlation table for the Pre-Test IMI Subscales.

		IntPreG	CompPreG	EffPreG	PresPreG
IntPreG	Pearson Correlation	1	.697**	.324*	-.081
	Sig. (2-tailed)		.000	.024	.581
	N	49	48	48	49
CompPreG	Pearson Correlation	.697**	1	.259	-.127
	Sig. (2-tailed)	.000		.079	.389
	N	48	48	47	48
EffPreG	Pearson Correlation	.324*	.259	1	.237
	Sig. (2-tailed)	.024	.079		.104
	N	48	47	48	48
PresRPreG	Pearson Correlation	.081	.127	-.237	1
	Sig. (2-tailed)	.581	.389	.104	
	N	49	48	48	49

The dependent variables (IntPreG, CompPreG, EffPreG, & PresRPreG were also tested for the assumption of linear relationship using scatterplot matrices (See Figure 11). A visual analysis of the scatterplot matrices showed that there was a linear relationship among the independent variables.

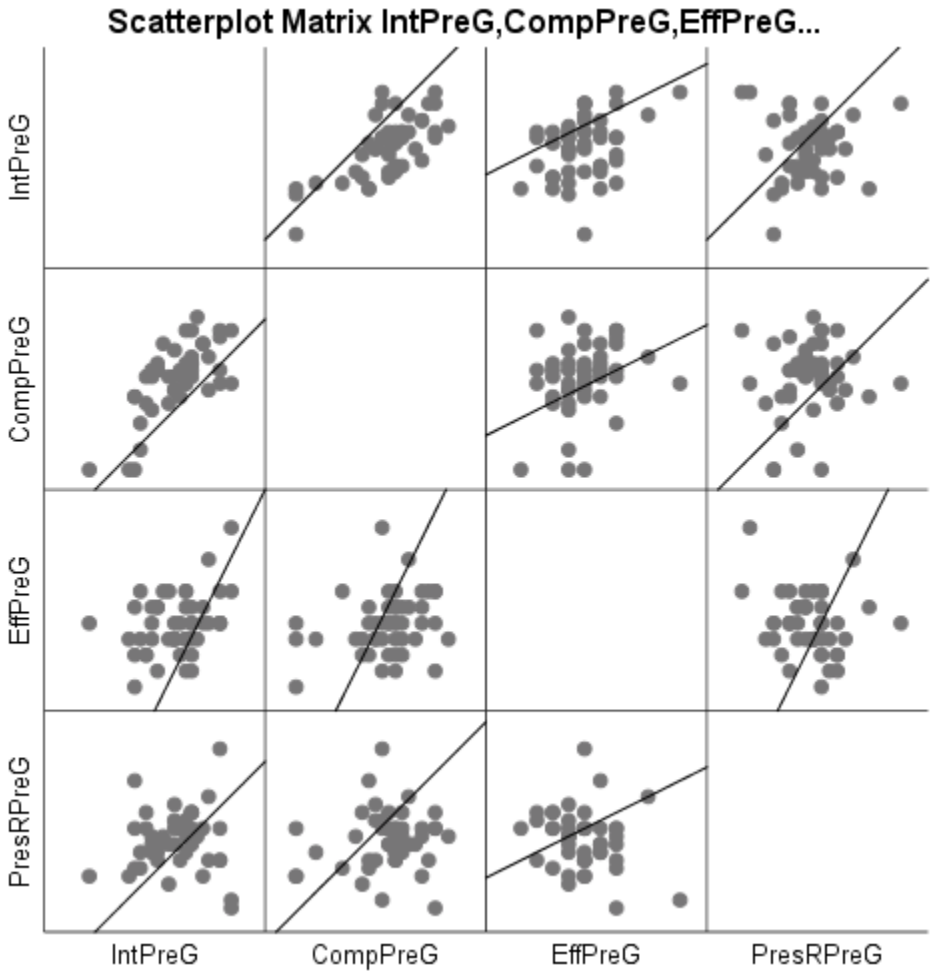


Figure 13. Scatterplot matrix of the pre-test IMI subscales: interest/enjoyment, perceived competence, effort/importance, and pressure/tension.

A Box's test performed on the pre-test data suggested that the dependent variables met the assumption of homogeneity of variance-covariances matrices ($p = .061$). Additionally, Levene's test showed that all dependent variables met the assumption of homogeneity of variances ($p > .05$) except for Perceived Competence (CompPreG, $p = .032$) (See Table 13).

Table 13

Pearson's Test for Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
IntPreG	Based on Mean	2.433	1	45	.126
	Based on Median	1.851	1	45	.180
	Based on Median and with adjusted df	1.851	1	37.787	.182
	Based on trimmed mean	2.199	1	45	.145
CompPreG	Based on Mean	4.890	1	45	.032
	Based on Median	4.795	1	45	.034
	Based on Median and with adjusted df	4.795	1	42.956	.034
	Based on trimmed mean	5.254	1	45	.027
EffPreG	Based on Mean	.042	1	45	.838
	Based on Median	.029	1	45	.865
	Based on Median and with adjusted df	.029	1	42.853	.865
	Based on trimmed mean	.082	1	45	.776
PresRPreG	Based on Mean	1.610	1	45	.211
	Based on Median	1.637	1	45	.207
	Based on Median and with adjusted df	1.637	1	43.334	.208
	Based on trimmed mean	1.844	1	45	.181

Post-Test Assumptions Analyses

To test change intrinsic motivation, a one-way multivariate analysis of variance test (MANOVA) was conducted on the end of course post-test questionnaire data to determine whether the two grading systems effected the intrinsic motivation levels of the

preservice teachers. The questionnaire questions we constructed using the Intrinsic Motivation Inventory (IMI), created by Deci & Ryan in 1982. As a result of the pre-test MANOVA conducted, it was determined that there was not a statistically significant difference in the pre-test IMI scores.

In order to conduct the MANOVA test on the post-test IMI subscales, preliminary analyses were first conducted to test several assumptions: (1) Assumption of no univariate or multivariate outliers, (2) Assumption of multivariate normality, (3) Assumption of no multicollinearity, (4) Assumption of a linear relationship between the dependent variables for each group of the independent variable, (5) Assumption of no univariate or multivariate outliers, (6) Assumption of homogeneity of variance-covariance matrices, and (7) Assumption of homogeneity of variances.

There were a few mild univariate outliers as assessed by boxplots (See Figure 14) and no multivariate outliers detected by Mahalanobis distance ($p < .001$). Unlike the pre-test boxplot outlier analysis, the post-test only had mild outliers. Since they were only mild outliers the data were kept for MANOVA analysis.

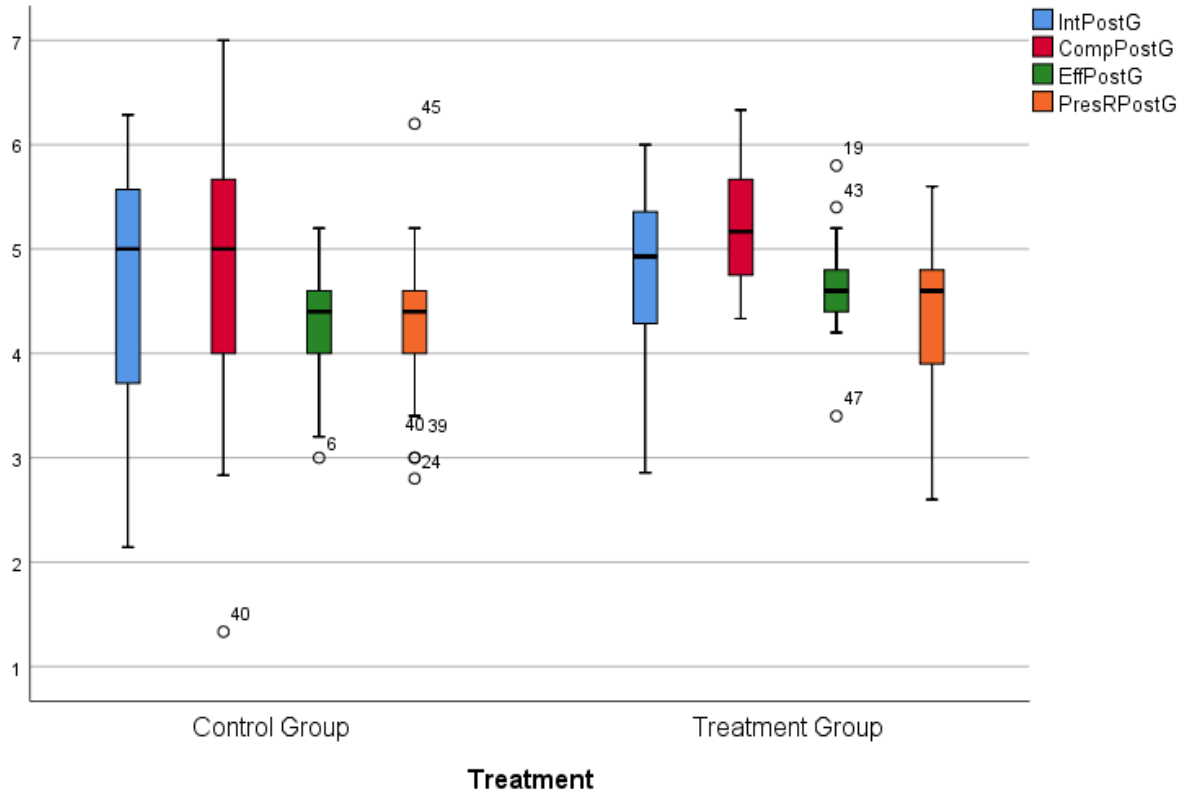
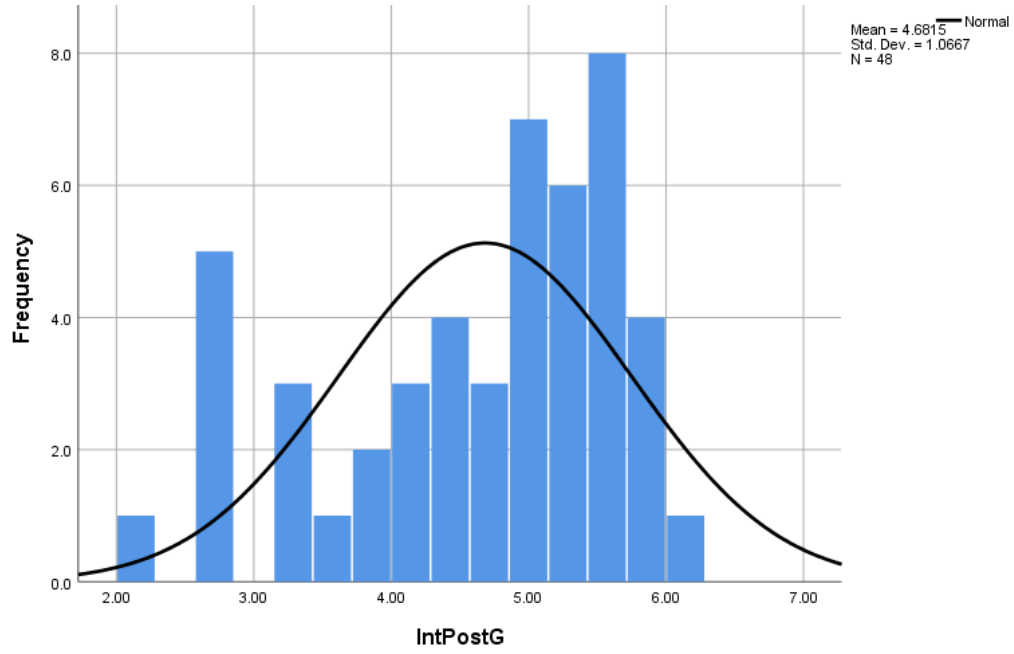
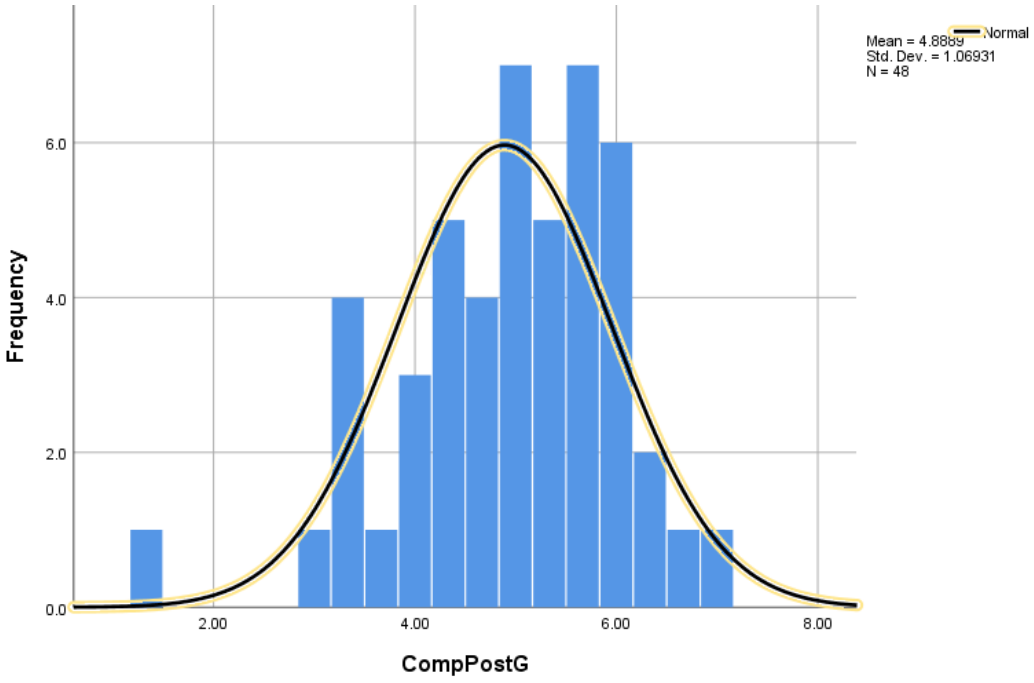


Figure 14. Boxplots for assessing outliers for all of the post-test IMI subscales.

Normality of distribution test was conducted using histograms of the post-test IMI subscales. Through a visual analysis, all four subscales are confirmed to be normally distributed (see Figure 15).


Mean = 4.6815
Std. Dev. = 1.0667
N = 48
Normal
Frequency
8.0
6.0
4.0
2.0
0.0
2.00
3.00
4.00
5.00
6.00
7.00
IntPostG


Mean = 4.8889
Std. Dev. = 1.06931
N = 48
Normal
Frequency
6.0
4.0
2.0
0.0
2.00
4.00
6.00
8.00
CompPostG

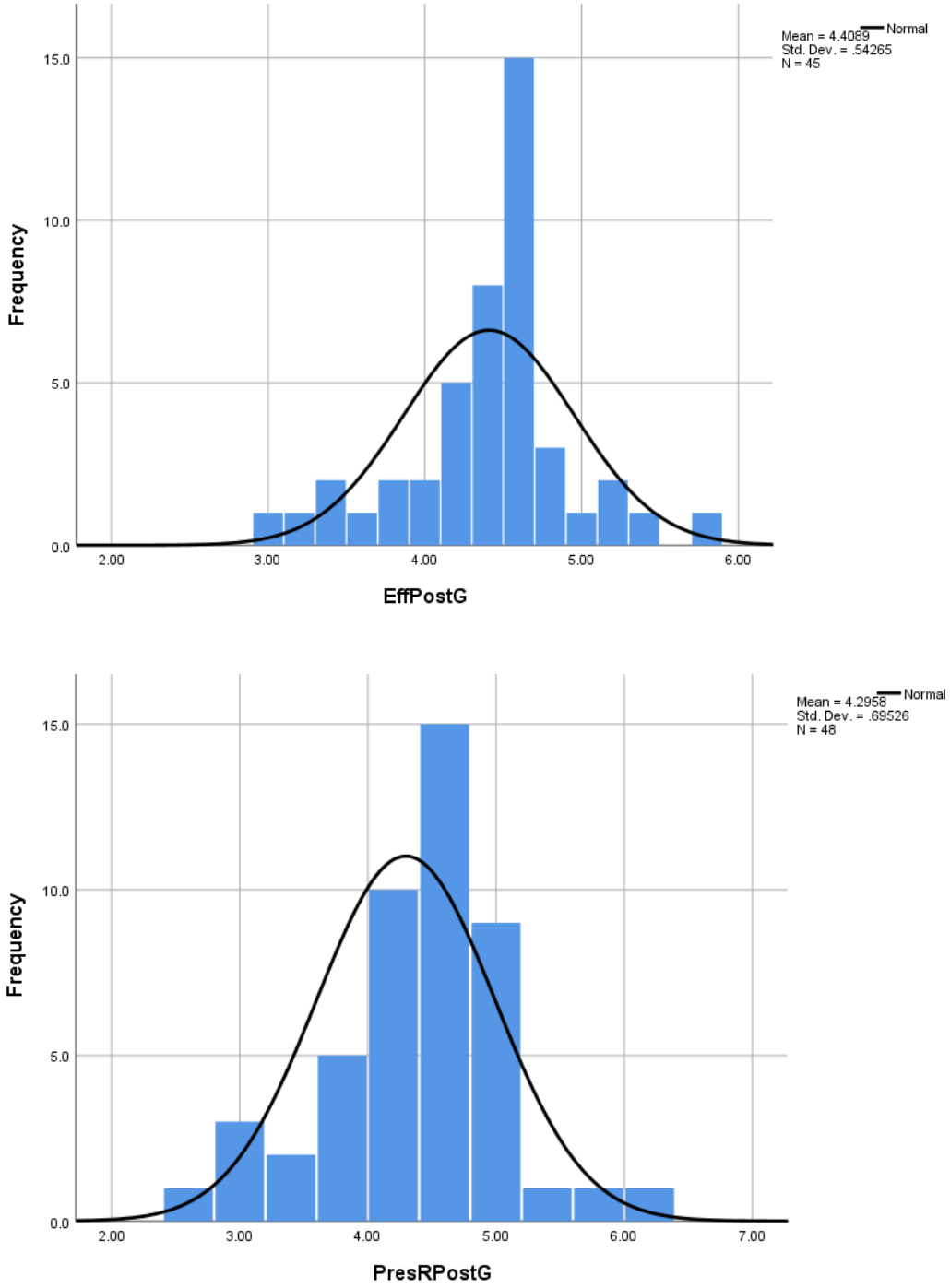


Figure 15. Distribution histograms for the IMI post-test subscales.

Pearson correlation r was run to detect multicollinearity. Table 14 shows the correlation matrix table of the dependent variables. The strongest correlation was between Interest and Competence subscales (r = .441, p = .002), while the weakest correlation was between Effort and Interest subscales (r = .036, p = .812). These correlations indicated that there was no multicollinearity.

Table 14

Pearson's r Correlation Matrix Table of Competence Subscales

		IntPostG	CompPostG	EffPostG	PresPostG
IntPostG	Pearson Correlation	1	.441**	.036	-.357*
	Sig. (2-tailed)		.002	.812	.013
	N	48	48	45	48
CompPostG	Pearson Correlation	.441**	1	.130	-.199
	Sig. (2-tailed)	.002		.393	.176
	N	48	48	45	48
EffPostG	Pearson Correlation	.036	.130	1	.246
	Sig. (2-tailed)	.812	.393		.104
	N	45	45	45	45
PresPostG	Pearson Correlation	-.357*	-.199	.246	1
	Sig. (2-tailed)	.013	.176	.104	
	N	48	48	45	48

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The dependent variables were also tested for the assumption of linear relationship using scatterplot matrices (see Figure 16). The scatterplot matrices showed that there was a linear relationship among the independent variables.

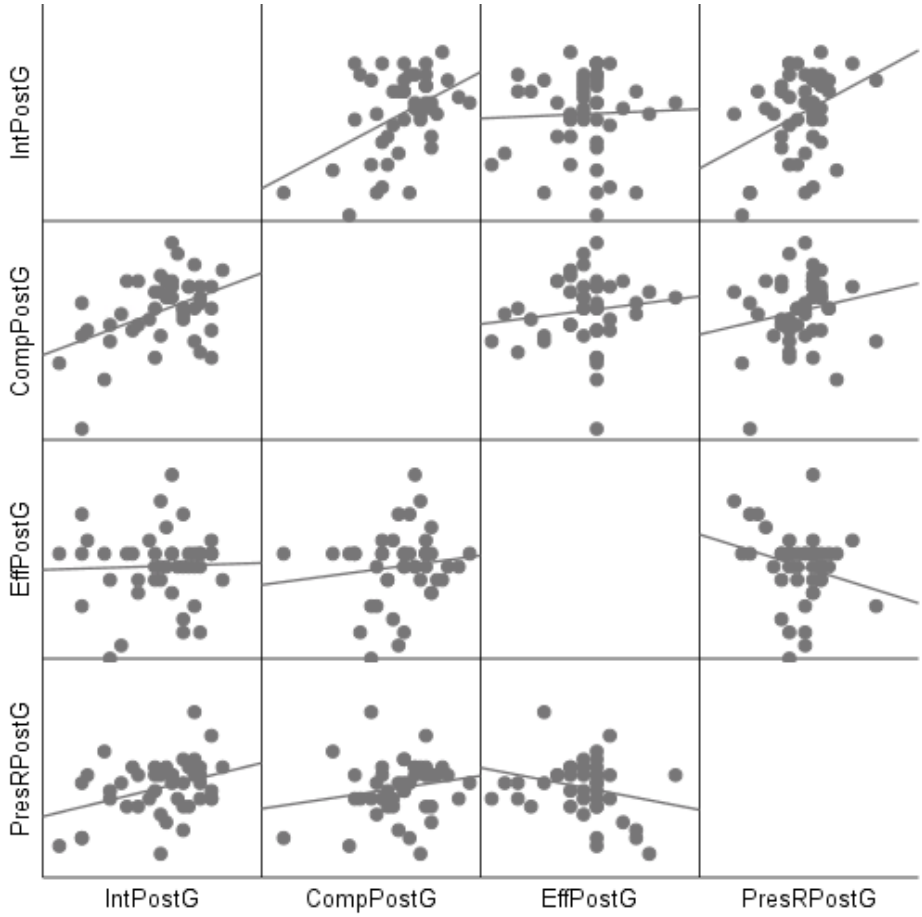


Figure 16. Scatterplot matrix of the Post-test IMI subscales: interest/enjoyment, perceived competence, effort/importance, and pressure/tension.

Box's test suggested that the dependent variables met the assumption of homogeneity of variance-covariances matrices ($p = .459$). Levene's test showed that all dependent variables met the assumption of homogeneity of variances ($p > .05$) except for competence ($p = .021$) (Table 15).

Table 15

Levene's Test of Equality of Error Variances for Post-Test IMI Subscales

		Levene Statistic	df1	df2	Sig.
IntPostG	Based on Mean	2.620	1	43	.113
	Based on Median	1.575	1	43	.216
	Based on Median and with adjusted df	1.575	1	38.706	.217
	Based on trimmed mean	2.446	1	43	.125
CompPostG	Based on Mean	5.953	1	43	.019
	Based on Median	4.957	1	43	.031
	Based on Median and with adjusted df	4.957	1	32.523	.033
	Based on trimmed mean	5.711	1	43	.021
EffPostG	Based on Mean	.029	1	43	.866
	Based on Median	.001	1	43	.982
	Based on Median and with adjusted df	.001	1	42.831	.982
	Based on trimmed mean	.013	1	43	.910
PresPostG	Based on Mean	.028	1	43	.868
	Based on Median	.005	1	43	.943
	Based on Median and with adjusted df	.005	1	42.003	.943
	Based on trimmed mean	.014	1	43	.908

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Treatment

Wilk's Lambda was selected for its robustness to of homogeneity of variances (Field, 2017). Wilk's Lambda showed that there was no statistically significant

difference between the treatment and control groups on the combined dependent variables, F (4, 40) = 1.465, Wilks' Λ = .231; partial η^2 = .128 (see Table 16).

Table 16

Homogeneity of Variances Testing for the Post-Test IMI Subscales

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.992	1186.667[b]	4.000	40.000	.000	.992
	Wilks' Lambda	.008	1186.667[b]	4.000	40.000	.000	.992
	Hotelling's Trace	118.667	1186.667[b]	4.000	40.000	.000	.992
	Roy's Largest Root	118.667	1186.667[b]	4.000	40.000	.000	.992
Treatment	Pillai's Trace	.128	1.465[b]	4.000	40.000	.231	.128
	Wilks' Lambda	.872	1.465[b]	4.000	40.000	.231	.128
	Hotelling's Trace	.146	1.465[b]	4.000	40.000	.231	.128
	Roy's Largest Root	.146	1.465[b]	4.000	40.000	.231	.128

a. Design: Intercept + Treatment

b. Exact statistic

Interestingly, tests of between-subjects effects showed that there was a statistical difference between the control and treatment groups specifically on the Effort/Importance IMI subscale (PostEffG), F(1) = 4.227, p = .046, partial η^2 = .089 (See appendix 3). However, since Wilk's Lambda takes precedent over the test of between-subjects effects, the ultimate result was that there was no statistically significant difference between the traditional grading group and the points-based grading group.

Table 17

Results of Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	IntPostG	.388[a]	1	.388	.346	.560	.008
	CompPostG	1.980[b]	1	1.980	1.801	.187	.040
	EffPostG	1.160[c]	1	1.160	4.227	.046	.089
	PresPostG	.082[d]	1	.082	.161	.690	.004
Intercept	IntPostG	927.564	1	927.564	827.078	.000	.951
	CompPostG	1026.671	1	1026.671	933.871	.000	.956
	EffPostG	819.435	1	819.435	2986.863	.000	.986
	PresPostG	566.809	1	566.809	1110.382	.000	.963
Treatment	IntPostG	.388	1	.388	.346	.560	.008
	CompPostG	1.980	1	1.980	1.801	.187	.040
	EffPostG	1.160	1	1.160	4.227	.046	.089
	PresPostG	.082	1	.082	.161	.690	.004
Error	IntPostG	48.224	43	1.121			
	CompPostG	47.273	43	1.099			
	EffPostG	11.797	43	.274			
	PresPostG	21.950	43	.510			
Total	IntPostG	1048.714	45				
	CompPostG	1141.167	45				
	EffPostG	887.680	45				
	PresPostG	644.760	45				
Corrected Total	IntPostG	48.612	44				
	CompPostG	49.253	44				
	EffPostG	12.956	44				
	PresPostG	22.032	44				

a. R Squared = .008 (Adjusted R Squared = -.015)

b. R Squared = .040 (Adjusted R Squared = .018)

c. R Squared = .089 (Adjusted R Squared = .068)

d. R Squared = .004 (Adjusted R Squared = -.019)

Class Performance Assumptions Analyses

This section of the appendix displays the tests of assumptions for the independent variables t-test procedures for research question (2) which compares the class performance variable (FinalGrades) between the treatment group and the control group. The following preliminary tests were performed to investigate the following assumptions: (1) Assumption of no outliers, (2) Assumption of normality, and (3) Assumption of homogeneity of variances.

Table 18

Descriptive Statistics for Class Performance

	Treatment	N	Mean	Std. Deviation	Std. Error Mean
FinalGrade	Control Group	29	92.10	5.205	.967
	Treatment Group	23	93.61	7.785	1.623

Outlier testing was assessed using a boxplot graph. One outlier existed, however it was still included in the analysis since it was just a mild outlier (See Figure 17).

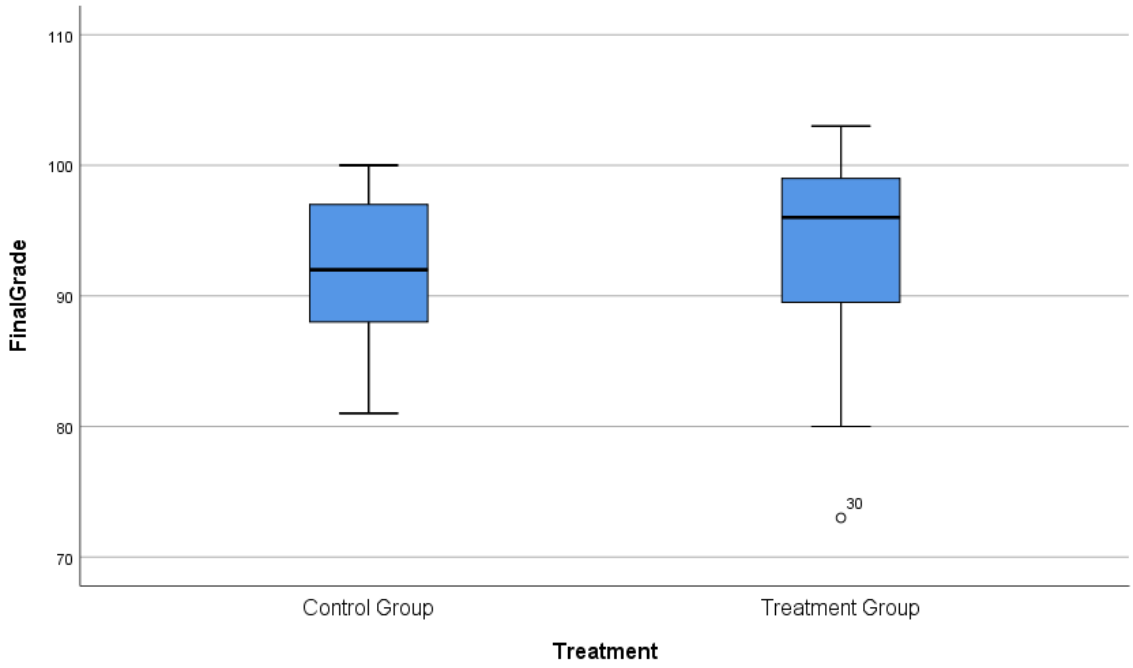


Figure 77. Boxplots to assess outliers for the final grade variable.

Normality distribution was tested using histograms. Both variables appear normally distributed for both groups (See Figure 16).

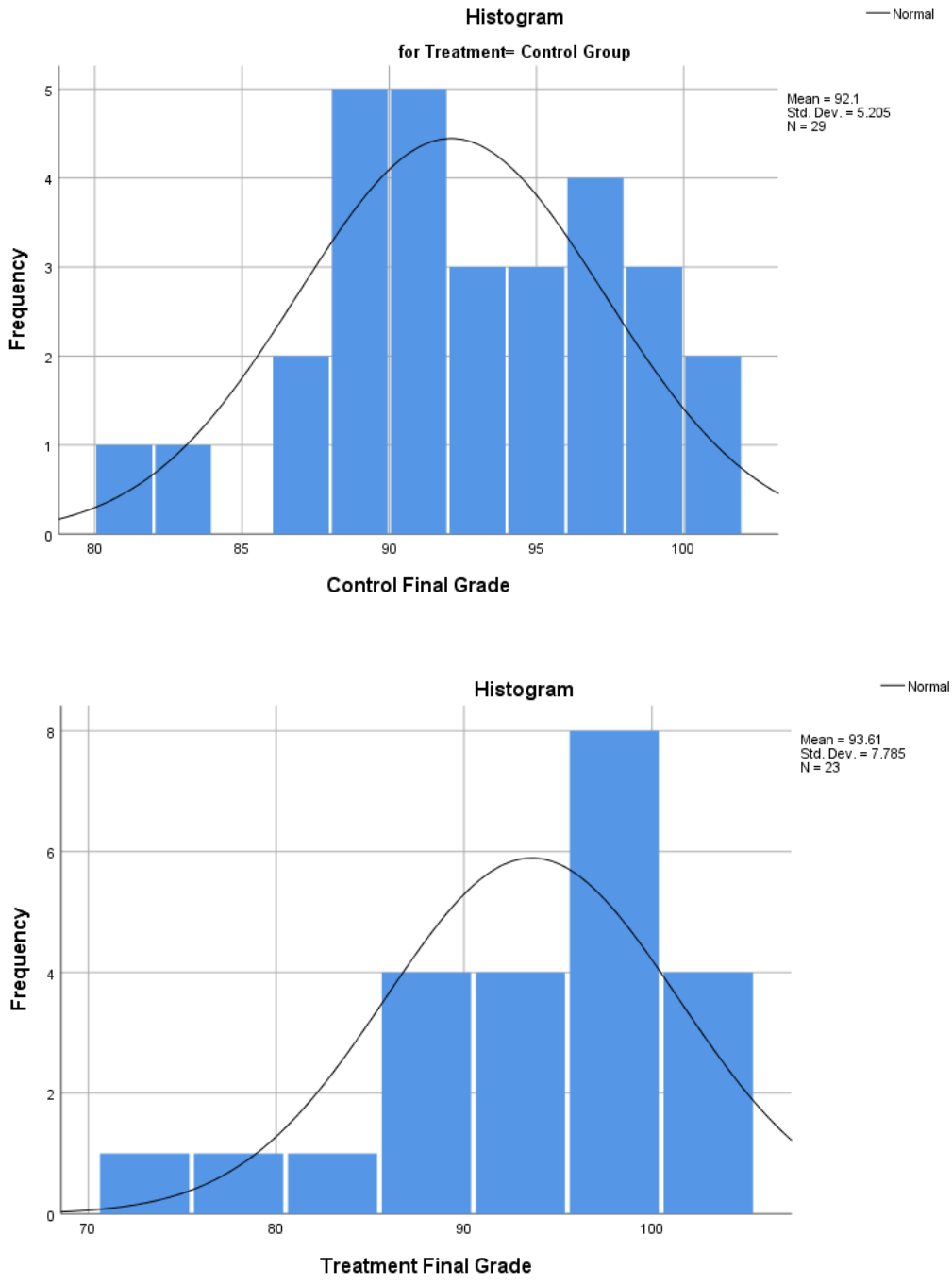


Figure 18. Histograms of the class performance variable to determine distribution normality.

There was homogeneity of variances for the class performances, as assessed by Levene's test for equality of variances (p = .109). Equal variances assumed.

Table 19

Levene's Test for Equality of Variances for the Class Performance Variable

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
FinalGrade	Equal variances assumed	2.670	.109	-.833	50
	Equal variances not assumed			-.797	36.737

Appendix D

Points-Based Grading Impressions Figures and Tables

Participants in the treatment group received additional questions on their impressions of the points-based grading system. These questions were only found on the post-test questionnaires.

The questions are as follows:

1. Earning points and levels motivated me to learn more than just traditional letter grades. (PointsQ1)
2. I prefer a traditional letter grading system over the points and levels system used in this particular class. (PointsQ2 Reversed)
3. I felt more engaged in this particular class using points and levels than in other classes using traditional letter grades. (PointsQ3)
4. I felt that I was rewarded more for my performance on activities in this particular class. (PointsQ4)
5. I felt that points and levels provided milestones that I was encouraged to reach. (PointsQ5)
6. I found the points grading system frustrating. (PointsQ6 Reversed)
7. The points grading system was easy to understand. (PointsQ7)

These questions were designed to be similar in structure to the IMI question items. Like the IMI, the questions participants were asked to evaluate how true to them each statement was. Also like the IMI, students responded using a 7-point Likert scale (1 = very untrue; 4 = somewhat true; 7 = very true). Finally, the questions were randomized when the students took the questionnaire, again matching the recommended IMI implementation. Twenty students completed the impressions section of the post-test questionnaire ($n = 20$).

Table 20

Descriptive statistics for the post-test points-based grading impressions questions

	PointsQ1	PointsQ2R	PointsQ3	PointsQ4	PointsQ5	PointsQ6R	PointsQ7
N	20	20	20	20	20	20	20
Mean	3.60	5.20	3.10	4.05	4.10	3.95	3.50
Median	3.50	6.00	3.00	4.00	4.00	4.50	3.50
Mode	1[a]	7	2[a]	5	4	6	4

Note. that the scores for question 2 and 6 have been reversed.

Tables and Figures for all Points Impressions Questions Combined:

Table 21

Descriptive Statistics for all Points Questions Combined

	PointsTotal
N	20
Mean	3.9286
Median	3.9286
Mode	3.71[a]

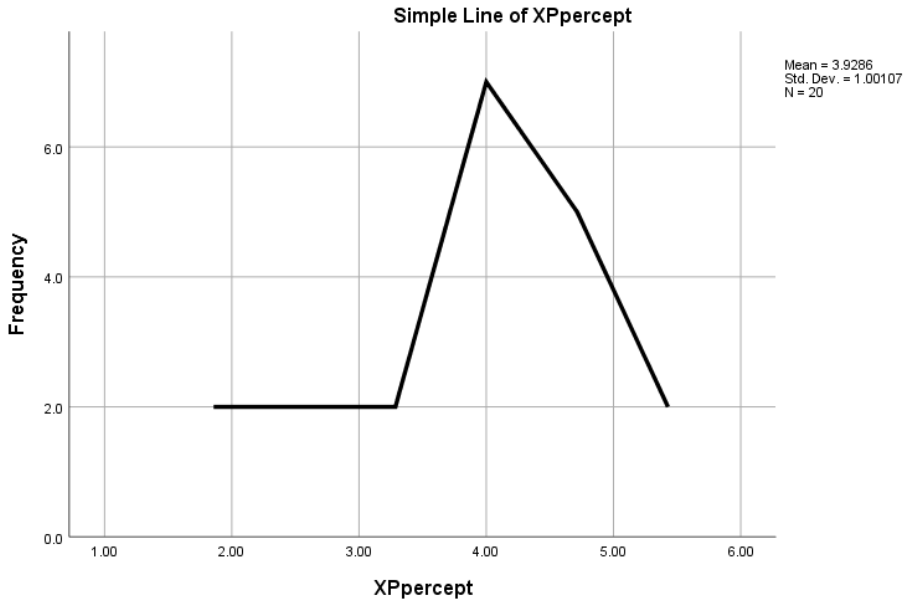


Figure 19. Line graph of the combined points question scores.

Points Impressions Question 1:

Question 1 asked participants to reflect on how much they agreed with the statement: *earning points and levels motivated me to learn more than just traditional letter grades.* With an overall average of 3.60, students were divided on their perceptions of the motivational benefits of the points-based grading system (see Table 21). However, more students disagreed (n = 10) with this statement than agreed (n = 8; see Table 22).

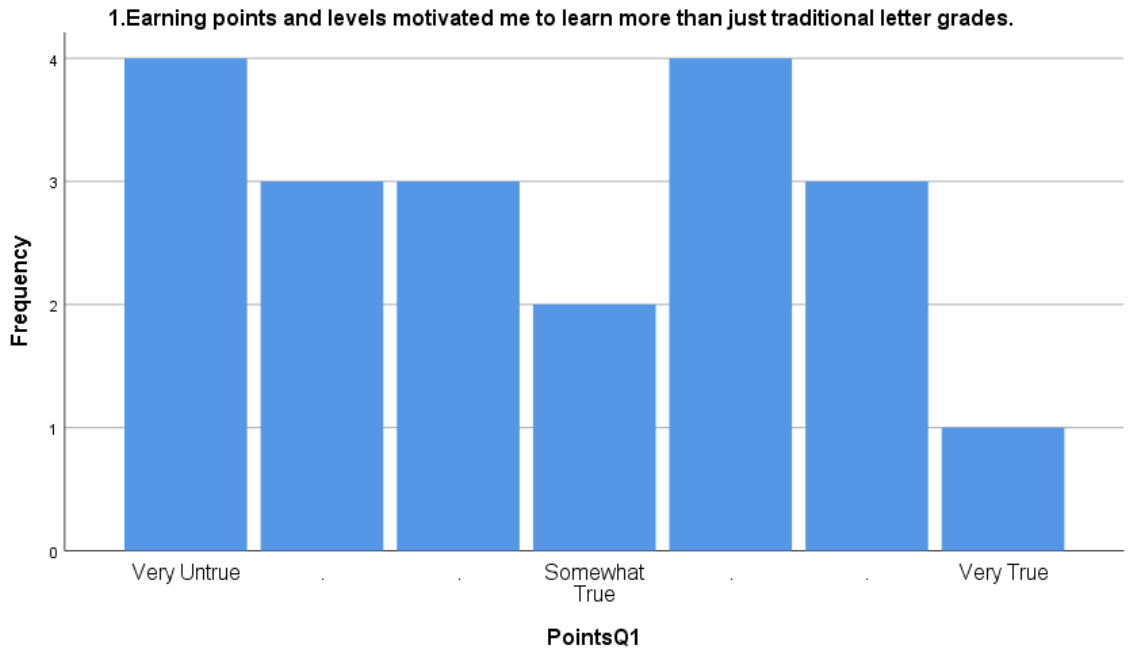


Figure 20. Bar graph of the results of points impressions question 1

Table 22

Earning Points and Levels Motivated Me to Learn More Than Just Traditional Letter Grades

	Frequency	Percent
1	4	20.0
2	3	15.0
3	3	15.0
4	2	10.0
5	4	20.0
6	3	15.0
7	1	5.0
Total	20	100.0

Note. 1 = Very Untrue, 4 = Somewhat true, 7 = Very True

Points Impressions Question 2:

Question 2 asked whether participants to reflect on how much they agreed with the statement: *I prefer a traditional letter grading system over the points and levels system used in this particular class.* Surprisingly, the great majority of the class disagreed with this sentiment (n = 14; m = 5.20). The most selected response to this question item was *very untrue* (n = 7) (see Table 24; see Figure 19). Even though there was no statistically significant difference on intrinsic motivation or class performance, the students still preferred the gamified approach over the traditional.

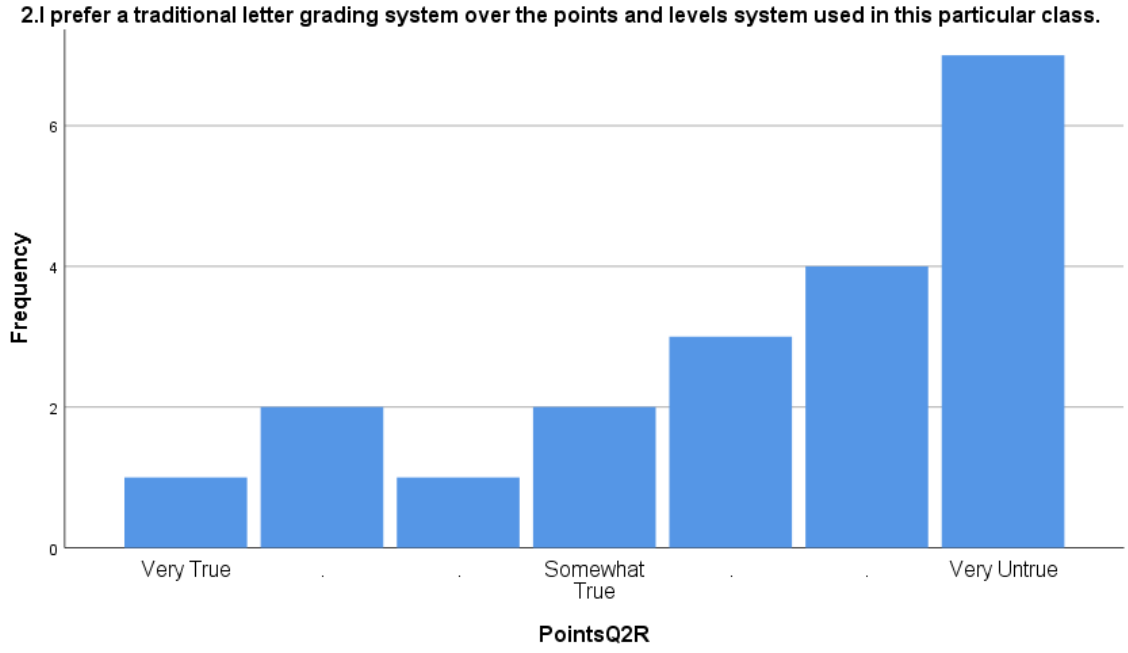


Figure 21. Bar graph for the results of points question 2.

Table 23

I Prefer a Traditional Letter Grading System Over the Points and Levels System Used in this Particular Class

	Frequency	Percent
1	1	5.0
2	2	10.0
3	1	5.0
4	2	10.0
5	3	15.0
6	4	20.0
7	7	35.0
Total	20	100.0

Note. The results of this question are reversed:
7 = Very Untrue, 4 = Somewhat true, 1 = Very True

Points Impressions Question 3:

Question 3 asked whether or not participants felt more engaged in the gamified class over a traditionally graded class. With a mean score of 3.10, a majority of students fell on the negative side of the scale. This implies that the students felt that the points-based grading system was not more effective on their engagement in the class over a traditional letter grading system.

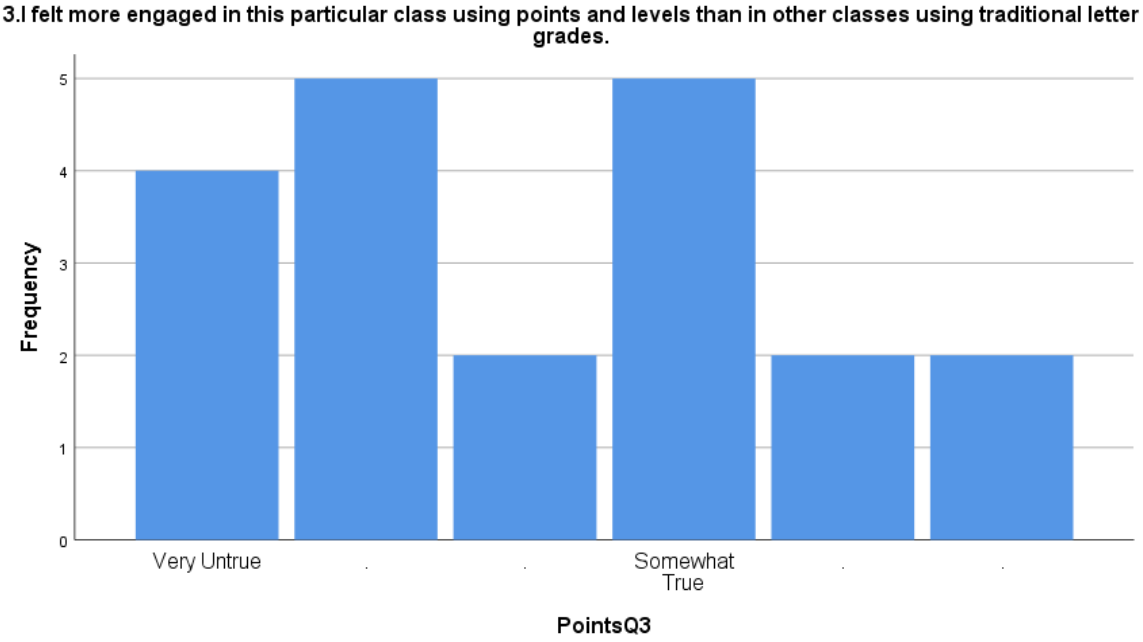


Figure 22. Bar graph of the results of points question 3.

Table 24

I felt more engaged in this particular class using points and levels than in other classes using traditional letter grades

	Frequency	Percent
1	4	20.0
2	5	25.0
3	2	10.0
4	5	25.0
5	2	10.0
6	2	10.0
Total	20	100.0

Note. 1 = Very Untrue, 4 = Somewhat true, 7 = Very True

Points Impressions Question 4:

Question 4 asked whether students felt more rewarded by the points system for their performance on class activities. While the results of this question leaned mostly positive. It was only by a differential of one student (disagree *n* = 8; agree *n* = 9; neutral n=3; See Table 25), and so no real conclusion can be made of on the effectiveness of points are in terms of rewarding student achievement.

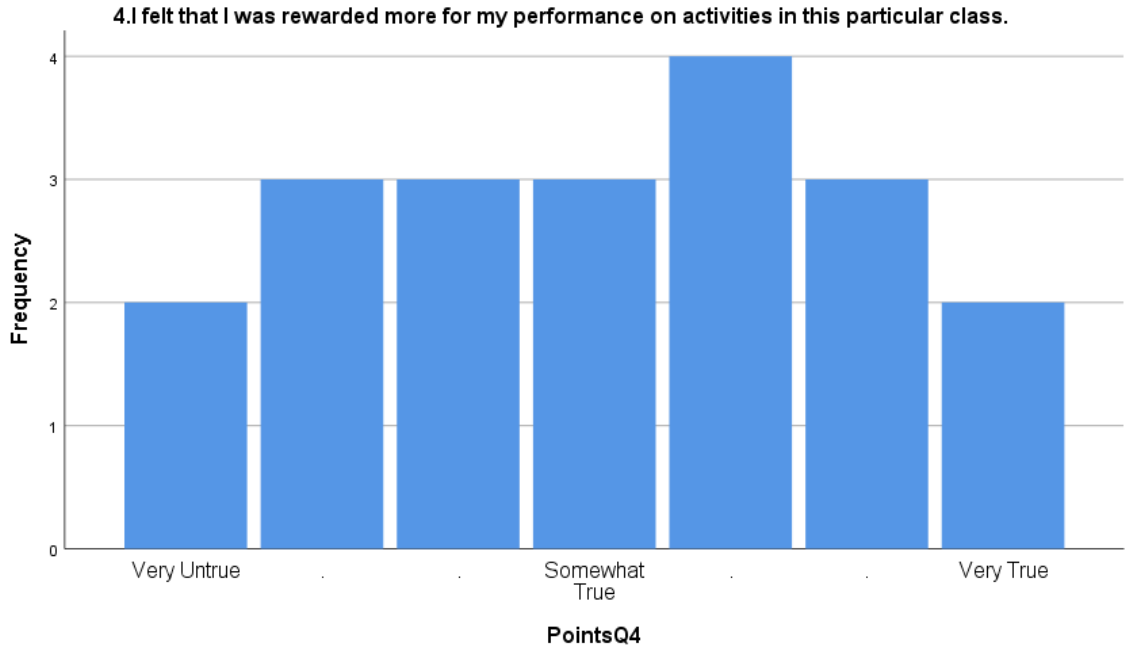


Figure 23. Bar graph of the results of points question 4.

Table 25

I felt that I was rewarded more for my performance on activities in this particular class.

	Frequency	Percent
1	2	10.0
2	3	15.0
3	3	15.0
4	3	15.0
5	4	20.0
6	3	15.0
7	2	10.0
Total	20	100.0

Note. 1 = Very Untrue, 4 = Somewhat true, 7 = Very True

Points Impressions Question 5:

Question 5 asked whether the level milestones encourages students towards reaching them. More students reported that they felt that the level milestones encouraged them rather than not (agree $n = 9$, disagree $n = 6$, neutral $n = 5$; see Table 26). However, the results were spread out across the scale overall.

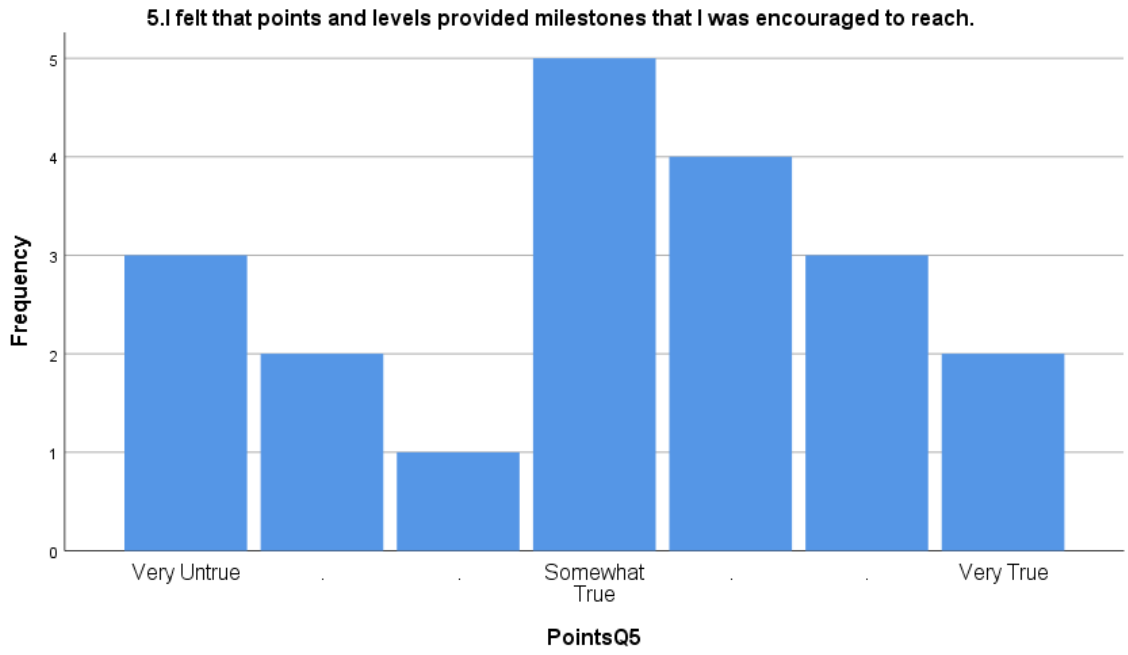


Figure 24. Bar graph of the results of points question 5.

Table 26

I felt that points and levels provided milestones that I was encouraged to reach

	Frequency	Percent
1	3	15.0
2	2	10.0
3	1	5.0
4	5	25.0
5	4	20.0
6	3	15.0
7	2	10.0
Total	20	100.0

Note. 1 = Very Untrue, 4 = Somewhat true, 7 = Very True

Points Impressions Question 6:

Question 6 asked if the students found the points-based grading system to be frustrating. Students overall did not find it frustrating and the majority of the participants declared that the points-based grading system did not frustrate them (n = 10). Unfortunately, there were 8 students that reported that they were frustrated (see Table 27).

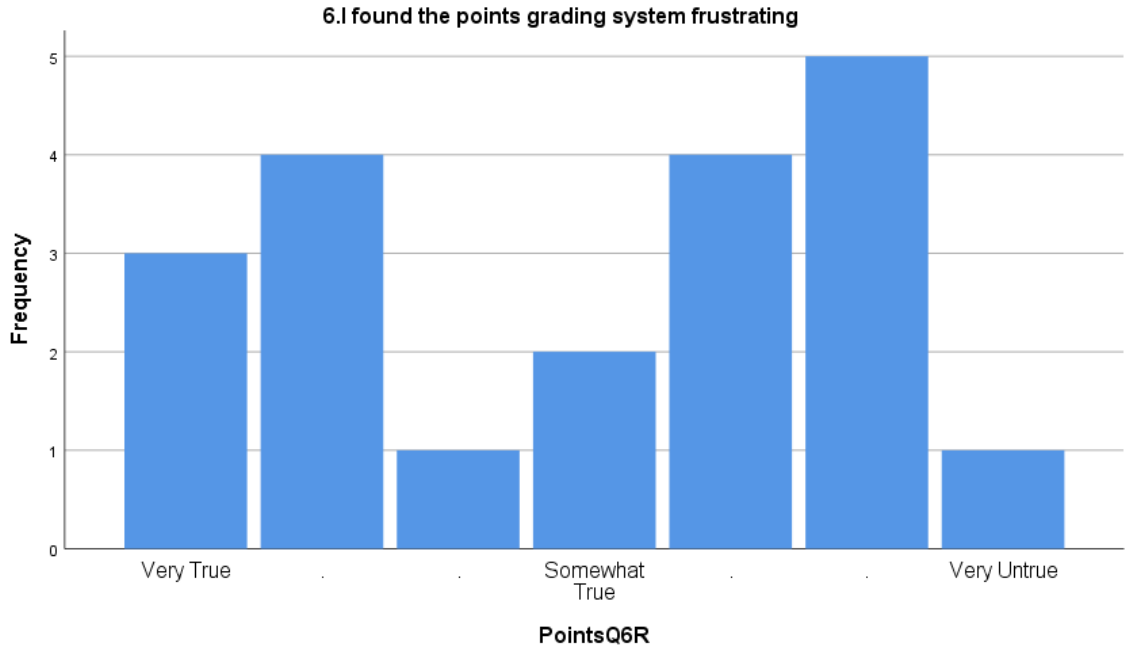


Figure 25. Bar graph of the results of points question 6.

Table 27

I Found the Points Grading System Frustrating

	Frequency	Percent
1	3	15.0
2	4	20.0
3	1	5.0
4	2	10.0
5	4	20.0
6	5	25.0
7	1	5.0
Total	20	100.0

Note. the results of this question are reversed:
7 = Very Untrue, 4 = Somewhat true, 1 = Very True

Points Impressions Question 7:

Question 7 asked if the grading system was easy to understand. Responses were varied. However, half of the students (n = 10; see Table 28) reported that the grading system was hard to understand. From these results I can infer that the points-based grading system was challenging for the treatment participants to understand.

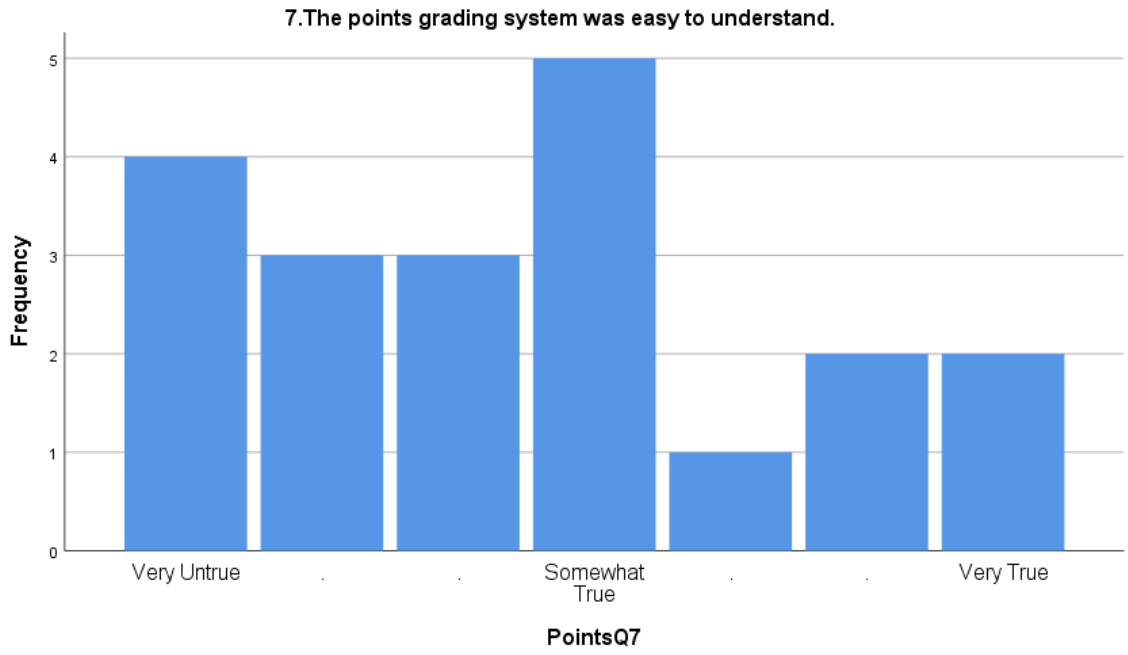


Figure 26. Bar graph of the results of points question 7.

Table 28

The Points Grading System was Easy to Understand

	Frequency	Percent
1	4	20.0
2	3	15.0
3	3	15.0
4	5	25.0
5	1	5.0
6	2	10.0
7	2	10.0
Total	20	100.0

Note. 1 = Very Untrue, 4 = Somewhat true, 7 = Very True

Appendix E

IRB Approval

UNIVERSITY of HOUSTON

DIVISION OF RESEARCH
Institutional Review Boards

APPROVAL OF SUBMISSION

March 2, 2018

George Zhao

gzzhao@uh.edu

Dear George Zhao:

On February 23, 2018, the IRB reviewed the following submission:

Type of Review:	Initial Study
Title of Study:	Using a Gamified Points-Based Grading System in Technology Courses for Pre-Service Teachers
Investigator:	George Zhao
IRB ID:	STUDY00000779
Funding/ Proposed Funding:	Name: Unfunded
Award ID:	
Award Title:	
IND, IDE, or HDE:	None
Documents Reviewed:	• XP Update Letter Sample.pdf, Category: Other; • Consent Script.pdf, Category: Recruitment Materials; • HRP-502a Consent Form 3 (Zhao).pdf, Category: Consent Form; • HRP-503 Protocol 2 (Zhao).pdf, Category: IRB Protocol; • Questionnaire Questions 2 (Zhao).pdf, Category: Study tools (ex: surveys, interview/focus group questions, data collection forms, etc.);
Review Category:	Expedited
Committee Name:	Not Applicable
IRB Coordinator:	Danielle Griffin

The IRB approved the study from February 23, 2018 to February 22, 2019, inclusive.

To ensure continuous approval for studies with a review category of "Committee Review" in the above table, you must submit a continuing review with required explanations by the deadline for the January 2019 meeting. These deadlines may be found on the compliance website (http://www.uh.edu/research/compliance/). You can

UNIVERSITY of HOUSTON

DIVISION OF RESEARCH
Institutional Review Boards

submit a continuing review by navigating to the active study and clicking "Create Modification/CR."

For expedited and exempt studies, a continuing review should be submitted no later than 30 days prior to study closure.

If continuing review approval is not granted on or before February 22, 2019, approval of this study expires and all research (including but not limited to recruitment, consent, study procedures, and analysis of identifiable data) must stop. If the study expires and you believe the welfare of the subjects to be at risk if research procedures are discontinued, please contact the IRB office immediately.

Unless a waiver has been granted by the IRB, use the stamped consent form approved by the IRB to document consent. The approved version may be downloaded from the documents tab.Attached are stamped approved consent documents. Use copies of these documents to document consent.

In conducting this study, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

Sincerely,

Research Integrity and Oversight (RIO) Office
University of Houston, Division of Research
713 743 9204
cphs@central.uh.edu
http://www.uh.edu/research/compliance/irb-cphs/