

Level Up: Using Gamification to Improve Student Evaluation and Motivation

Jarrod T. Kelly
Assistant Professor of Political Science
North Carolina Wesleyan College
JKelly@ncwc.edu

Abstract

Research spanning decades has shown that grading reduces student motivation and performance. Additionally, grading often induces feelings of powerless and reduced autonomy, resulting in an almost single-minded focus on avoiding a poor grade rather than a desire to meet course learning objectives. In an effort to overcome the negative effects associated with grading, I have developed a novel method of student evaluation, which I call Reflected Progression Grading (RPG). The RPG approach incorporates motivational elements, inspired by popular games and gamification research, in an attempt to encourage autonomous motivation and performance-approach goals among students in a general education course. This paper describes the RPG method in detail and presents the results of two studies examining the efficacy of this approach. The initial results are encouraging: the course failure rate was greatly reduced and students reported feeling greater perceived competence. In addition, large majorities of students reported feeling more autonomous and motivated. Likewise, most students expressed a preference for this type of grading over more traditional methods. Overall, this gamified approach seems to overcome many of the negative effects associated with traditional evaluation.

Paper to be presented at the 2020 American Political Science Association Teaching and Learning Conference in Albuquerque, NM, February 7-9, 2020.

This is a very preliminary and unpolished draft. Please do not cite without permission.

General education courses present a consistent challenge in higher education. While the value of a broad education is obvious to faculty and administrators, students often fail to understand the importance of these courses (Vander Schee 2011; Thompson, Eodice, and Tran 2015). In fact, students rank outcomes such as, a “sense of values, principles, and ethics,” an “expanded cultural and global awareness and sensitivity,” and an “appreciation of your role as a citizen and an orientation toward public service,” as some of the least important goals in pursuing an undergraduate degree (Humphreys and Davenport 2005, 40). Alas, this is particularly disheartening within political science as these are precisely the type of outcomes that we hope to achieve among our students. Instead, many students view general education as an obstacle to overcome on the way toward fulfilling their career goals. Unsurprisingly, a recent study found that a majority of students would rather take additional courses in their major, while approximately one-half of students would likely not enroll in general education courses at all if they were not required (Thompson, Eodice, and Tran 2015).

Given that students often have negative feelings toward general education, new strategies are needed to better motivate and engage students enrolled these courses. The method that I will describe and test in this paper seeks to address this need through improving another unpopular and ubiquitous component of education: student evaluation, i.e., grading. Specifically, the method that I have developed, which I refer to as Reflected Progression Grading (RPG), is intended for courses which are primarily populated by students seeking to fulfill a general education requirement. This method aspires to overcome many of the problems associated with traditional methods of student evaluation, which often serve to demotivate already unmotivated students within these courses.

I will proceed by first discussing some of the problems associated with traditional methods of student evaluation. Next, I will discuss some recent trends in evaluation that seek to address many of the concerns associated with traditional evaluation. I will then introduce the RPG method of evaluation and provide two sets of empirical tests. First, I conducted a pilot implementation of the method during the 2018-19 academic year in my introductory American politics and government course. I am then able to compare student outcomes pre-/post-implementation as well as measure overall student satisfaction with the method. Second, I conducted a natural experiment during the Fall 2019 semester whereby one section of my introductory course utilized a traditional method of evaluation and the other section utilized the RPG method. Here, I am able to directly compare the efficacy of this new approach to evaluation with the more traditional means of evaluation that I relied upon previously.

The Pernicious Effects of Traditional Student Evaluation

Ubiquitous across educational institutions of all shapes and sizes is the letter grade system of student evaluation. Students are assigned a course grade, from A to F, that typically corresponds to a percentage score generated by either a weighted average of assignment scores or the sum total of points earned divided by the number of points available. Further, individual assignments also often receive letter grades, which are themselves frequently based on a percentage score. This system of grading emerged in higher education in the late 18th century and became commonly adopted over the next century and a half (Durm 1993). Now, letter grades are nearly universal in their use from primary to post-secondary education.

While the letter grade system has greatly standardized academic transcripts, scholars have noted a litany of negative effects stemming from their use. Among other findings, grades have been shown to: increase feelings of powerlessness (Elliot and McGregor 2001), reduce feelings

of autonomy (Ryan and Deci 2000), reduce intrinsic motivation (Butler and Nisan 1986), as well as reduce interest and performance (Butler 1988; Harter 1978). Put simply, grades are a detriment to the learning process.

Ideally, grades would help to foster autonomous motivation, composed of extrinsic motivation, or behavior driven by an individual's desire to achieve a high grade, and intrinsic motivation, or behavior driven by interest in the material, coupled with a feeling of autonomy to realize these goals. In fact, grades have been shown to have the opposite effect. Specifically, grades tend to increase performance avoidance goals, i.e., the desire to avoid a bad grade, rather than increasing performance-approach goals, i.e., the desire to obtain a good grade (Pulfrey, Buchs, and Butera 2011).

What is to be done? At most institutions, individual instructors are required to assign a letter grade upon completion of the course. They do not have the authority to deviate from this established system of evaluation. Part-time and non-tenured faculty in particular face great pressure to conform to this system and may even feel coerced to conform to a specific grade distribution. However, many instructors retain great autonomy to determine *how* these letter grades are ultimately determined and in doing may pursue strategies intended to reduce the negative effects stemming from this type of evaluation.

Recent Trends in Student Evaluation

In recent years, there have been a number of novel strategies introduced to combat the pernicious effects of grading. As the assignment of letter grades is, in almost all cases, inevitable, instructors must instead think critically about how to deemphasize evaluation while continuing to deliver meaningful assessments and feedback. One approach, popular in the humanities and writing-intensive courses, is the use of portfolios. Rather than students receiving an evaluation

on each individual writing assignment, they will instead receive formative feedback on assignments throughout the semester. Instructors typically provide actionable comments about how to improve their work. Students then have the opportunity to revise the work and include it within a representative portfolio for evaluation at the end of the semester. The portfolio may be evaluated in a developmental manner, focusing on improvement and integration of feedback, or as a showcase of the student's best work, focusing on the overall quality of the final product (Mueller n.d.). In this way, the emphasis is on the process of learning and the continual development of students' abilities. However, the implementation of portfolio grading is limited to certain types of courses and requires a considerable time investment by the instructor.

Contract grading is another option that seeks to deemphasize grades while increasing student agency. In this method, students will typically make an agreement with the instructor at the beginning of the term specifying which assignments they will complete and the course grade that they will receive for successful completion of the assignments. What "successful completion" looks like is ultimately left to the instructor, but the idea is that students who wish to earn a higher grade complete more work than students who aspire toward a lower grade. This method seeks to foster greater student agency through mutual determination of the course workload by the student and instructor, as well as to provide greater clarity regarding evaluation (Inoue 2019). However, the applications of contract grading are again somewhat limited, as it is primarily designed for courses featuring written assignments and projects.

One new method of evaluation that is applicable to a wide variety of courses is specifications grading (Nilson 2015). This method of evaluation seeks to clarify expectations and reduce the stigma associated with poor performance on coursework. Specifically, in specifications grading, the expectations (or "specs") for all assignments are provided to students

in a detailed checklist. Students whose work satisfies all of the specs are awarded full credit for the assignment. Students whose work does not satisfy one or more specs does not receive credit. However, students have the opportunity to revise and resubmit work that does not initially receive credit. If the revised work meets all specs, they then receive full credit for the assignment. Then, similar to contract grading, final course grades are assigned based on the quantity of assignments completed. Unlike contract grading, however, students may alter their targeted grade throughout the semester without needing to notify the instructor. Additionally, instructors may integrate exams into this method by requiring threshold scores for each grade, e.g., an 80% average for a grade of B or higher, a 60% average for a grade of D or higher. However, this would serve to replace some of the emphasis on letter grades and may increase performance-related pressure and anxiety. While the course could be structured to allow numerous attempts on exams, this could be quite onerous for instructors.

Overall, a system of evaluation is needed that 1) deemphasizes letter grades, and 2) is flexible enough to be widely implemented across a wide variety of courses. While portfolio grading and contract grading are successful at the former, their application is limited to certain types of courses. Likewise, while specifications grading is more flexible, and provides the ability for students to revise and resubmit work, its method of assigning grades when exams are a component of the course can undermine some of its benefits. The evaluation method I propose seeks to solve some of these deficiencies by combining the best aspects of specifications grading with a more flexible method of allocating final course grades. Further, it seeks to bolster student engagement and motivation through incorporation of game-like components, another efficacious educational strategy.

Gamification

A game has been aptly defined as a “voluntary attempt to overcome unnecessary obstacles” (Suits 1978, 54). In his description of lusory attitudes, or a playful attitude that emerges when one enters into a game, Bernard Suits (1978) writes, “To play a game is to attempt to achieve a specific state of affairs [prelusory goal], using only means permitted by rules [lusory means], where the rules prohibit use of more efficient in favour of less efficient means, and where the rules are accepted just because they make possible such activity [lusory attitude]” (54-55). At their core, games contain objectives, rules, and a lusory (playful) attitude.

With this in mind, the overlap between a game and an academic course is readily apparent. When taking a course, students submit to overcoming some obstacles that they would not normally otherwise experience in the pursuit of an objective (ideally, this would be mastery of the course material). Further, the rules established in a course are quite different from those outside of the course and only exist within the context of the course itself. However, because student performance within a course often determines, or is perceived to determine, their success outside of the course, the lusory attitude that emerges within a game is not often present in most courses. The process of gamification seeks to introduce a lusory attitude within education and in the process restore students’ motivation to learn.

Specifically, gamification refers to the process of integrating game design elements and principles into non-game settings (Robson et al. 2015). In this case, the non-game setting is a course, though gamification in recent years has been more widely applied in non-educational settings. For example, loyalty programs are offered by many businesses. These loyalty programs are increasingly gamified, often with the use of a smartphone app that allows customers to track their earned rewards. Consider “Starbucks Rewards,” a loyalty program offered by the coffee

chain Starbucks.¹ For each purchase made, customers can earn “stars,” which appear in their Starbucks smartphone application. These stars can be redeemed for free beverages and food items. In addition, if customers earn a large number of stars within a year, they can earn “Gold Status,” which entitles them to additional rewards and results in a gold Starbucks card being issued. Starbucks further incentivizes purchases through double-star days, featured items for purchase that award bonus stars, and even games within the application itself. While these gamified elements may seem somewhat silly on the surface, they do effectively alter consumer behavior. Starbucks reports that their loyalty program accounted for approximately one-third of their growth in 2019 (Maze 2019). The question is: can gamification do the same for education? Specifically, can gamification alter students’ attitudes and behavior in the classroom?

Early research on gamification is quite positive, though much more research is needed. Recent studies on gamification report that, compared to courses using traditional grading, gamified courses increase performance (De-Marcos, Garcia-Lopez, and Garcia-Cabot 2016), attendance (Topîrceanu 2017), participation (Iosup and Epema 2014), and final course grades (Iosup and Epema 2014; Topîrceanu 2017). Yet, a recent review of the literature highlights the difficulties in identifying the specific gamification elements that engender these gains in motivation and performance (Dicheva et al. 2015).

Indeed, there are many common elements included in gamification approaches. Most of these elements seek to mimic the experience of being in a game setting in order to foster lusory attitudes. The most frequently cited elements include points, leaderboards, badges, and rewards (Dicheva et al. 2015; Zichermann and Cunningham 2011). Points are typically earned as they are in most courses: through the completion of assignments and exams, class participation,

¹ Details on the rewards program can be found here: <https://www.starbucks.com/rewards/>

attendance, projects, etc. However, in a gamified course, the points are usually accrued from a starting point of zero and students' current scores are displayed as such.² Student points may often be displayed on a course leaderboard, either listing all students or just the current students with the most points.³ Badges are also a commonly used element and are given to indicate superb performance on specific tasks, such as a high score on an exam, frequent class participation, or an excellent paper. Finally, rewards are often utilized to further increase engagement and motivation. Many approaches, such as my own, feature a token economy which allows students to earn and redeem tokens for specific rewards, such as excusing an absence or skipping a quiz. All of these elements seek to engender a game-like feel within a course and to cultivate lusory attitudes.

The RPG Approach

Reflected Progression Grading (RPG) incorporates many of the gamification elements described previously, as well as other evaluation methods, specifications grading (Nilson 2015) in particular, in an attempt to deemphasize letter grades and to promote student engagement and motivation. As a starting point, the RPG method considers Huizinga's (1938) examination of play. According to Huizinga, a game contains choices, rules, pathways, community, and feedback, with the end result being a feeling of empowerment. Accordingly, the RPG method seeks to give students choices regarding assignments, clear and straightforward rules for progressing in the course, multiple pathways (through assignment choice) to achieving a high

² While, in reality, all courses utilizing a points-based system work this way, with the advent of online gradebooks students are often presented with the current score displayed as points possible out of the current number of possible points. Thus, students rarely seem to view their grade as starting at a zero, as a percentage score is consistently shown and updated.

³ Given FERPA restrictions, students' point totals on a leaderboard need to be made anonymous.

grade, community through a course leaderboard, and meaningful feedback on their progress. The foremost goal is to empower students to excel and master the course material.

Points. As in most evaluation systems, the RPG approach centers on earning points. All students begin at zero points and are only ever shown their total number of points, never a percentage score corresponding to a grade. Critically, students can only gain points and never lose points. There are many opportunities to earn points, and students attempting any opportunity never risk their current points. If students do not successfully complete a points opportunity, they do not earn points, but they also do not lose points. In a sense, this is similar to any grading system: students may earn 0 points or a 0% on an assignment, but they do not earn a negative score. The core difference lies in the fact that there are many more points available than are needed for even the highest course grade (i.e., an A). Thus, while skipping an assignment entirely in most courses will negatively impact a student's percentage score, as those points are no longer available to earn, students may choose to skip a certain number of assignments in this course without being penalized. In my courses, I typically offer about one-third more available points than are needed for the top course grade. Specifically, about 1,350 points are available, with an A grade requiring 1,000 earned points.

By expanding the number of possible points, one may argue that I am merely lowering the percentage score required for a certain grade. For example, earning 1,000/1,350 points would simply lower the percentage score required for an A from 93% (which is typical) to 74%. In one sense, yes, I am lowering the percentage score and framing points in a different way. However, to combat concerns of simply lowering standards, I only award credit to students who complete assignments at a high level. This should help to negate concerns that students may seek to complete many assignments at a low level and be rewarded with a high grade.

Specifically, I utilize specifications grading to hold students accountable when completing assignments and to ensure that only high-quality work receives credit. Each written assignment in the course is graded using a checklist. These checklists are provided to students on the first day of the course so that they know precisely how their work will be evaluated. When students submit an assignment, I complete the associated checklist, noting each specification (“spec”) that they successfully complete. If students meet all specs, they receive full credit. If they meet almost all specs, but fall short on one or two minor specs (such as those related to formatting), they receive half credit. If they do not meet several specs, or just one major spec (such as one related to substantive content), they receive no credit. However, students are able to revise and resubmit assignments up to two times, thereby giving students multiple attempts to earn credit on each assignment. An example specs sheet, for a blog post assignment, is shown in **Figure 1**.

Essentially, the RPG method affords students greater choice in determining which assignments to complete by providing a large pool of points. In order to maintain rigor, I utilize specifications grading when awarding points for completing assignments. Only high-quality work will receive points and students may revise assignments in order to gain points. This should encourage a progress-oriented, or growth, mindset in multiple ways. First, students begin with zero points and only gain points. Second, students are given multiple opportunities to succeed on most assignments. I also encourage a progress-oriented mindset in a third way: through the gaining of levels, i.e., “leveling up.”

Blog Post Assignment (30 Points)

Student Name:

Blog Post #

Specs List	✓ #1	✓ #2	✓ #3	Comments
The post uses Times New Roman 11 pt. or 12 pt. font	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The text is single-spaced	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
There is a one-line space between paragraphs (paragraphs should not start with an indent; you may need to insert two spaces between paragraphs in order to separate them)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The blog post describes a political event or story that took place during the course	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The post is written in an editorial style meaning that the author shares his/her views on the issue and defends them persuasively	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The post cites at least three news articles in APA style. There are in-text citations and a reference list at the end of the post	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The articles cited are from at least two different sources (websites)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The articles cited are from <i>reputable</i> media outlets or sources (see list on MyNCWC for examples of reputable versus non-reputable sources)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The post contains a title that describes the position taken on the issue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The post is 300-450 words in length	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
The post contains no more than a few minor spelling and/or grammatical errors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Paragraphs are used to effectively organize the essay	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
There are no contractions in the post (isn't, aren't, won't, didn't, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Submission 1 Score: ☐ Full Credit (30 points) ☐ Partial Credit (15 points) ☐ No Credit (0 points)

Submission 2 Score: ☐ Full Credit (30 points) ☐ Partial Credit (15 points) ☐ No Credit (0 points)

Submission 3 Score: ☐ Full Credit (30 points) ☐ Partial Credit (15 points) ☐ No Credit (0 points)

Additional Comments:

Figure 1 – Example Specs Sheet. Specs sheets, such as the one pictured above, are used to evaluate all written work in the course. As students have up to three attempts to achieve full credit, there are three checklists (one per attempt).

Levels and Rewards. As students hit certain points milestones, they gain levels or “level up.” Each time they level up, they also gain rewards. This mechanic serves two purposes. First, it affords students a sense of status. Students know that higher levels indicate greater progress made within the course. Their final grade in the course is assigned based on the level they attain at the end of the course, as shown in **Table 1**. For example, students who achieve level seventeen

by the end of the course will receive a grade of C- while students who achieve level twenty-three will receive a grade of B+. Students know that they can only gain levels, not lose levels, so they can be confident that they will receive at least a certain grade by the time they attempt the final exam.

Second, students earn rewards for each level gained. Specifically, they accumulate tokens that can be redeemed for a variety of purposes. For example, tokens may be used to resubmit work that does not receive full credit (this also helps to prevent initial submission of shoddy work). They may also be used to excuse tardiness, for a short extension on assignments, and to skip individual questions on exams, among other benefits. Typically, I find that students prefer to save their tokens and do not spend them frivolously. This is partially because unused tokens may be redeemed for points at the end of the course at a predetermined rate. In addition, at certain level milestones, such as level ten, students have the opportunity to roll a six-sided die to determine the number of tokens awarded. This is yet another means by which to encourage lusory attitudes among students.

Table 1 – Levels and Rewards

Level	Points Required	Tokens Earned	Final Course Grade
1	25	1	F
2	50	1	F
3	75	1	F
4	100	1	F
5	125	1	F
6	150	1	F
7	200	1	F
8	250	1	F
9	300	1	F
10	350	1-6 (roll)	F
11	400	2	F
12	450	2	F
13	500	2	F
14	550	2	F
15	600	2-6 (roll)	D
16	630	2	D
17	670	2	D+
18	700	2	C-
19	730	2	C
20	770	2-6 (roll)	C+
21	800	3	B-
22	830	3	B
23	870	3	B+
24	900	3-6 (roll)	A-
25	1000+	-----	A

The Course Leaderboard. Rather than utilize the default gradebook in the learning management system (LMS), students' points total, current level, and number of tokens are displayed on a course leaderboard, as shown in **Figure 2**. All students are assigned an anonymous username at the beginning of the course (or they may choose their own username, as long as it does not clearly identify them). Their username is known only to them. In this way, another game-like element is introduced while students' progress remains anonymous. The leaderboard is connected to a spreadsheet containing scores on all assignments (though they cannot see individual assignment scores). Thus, whenever I enter scores on any assignments into

the spreadsheet, students' point totals, levels, and tokens are automatically updated on the embedded public leaderboard. This method is rather seamless and, once set up, requires little additional instructor effort to maintain compared to any normal grade-tracking spreadsheet.

In addition to the full leaderboard, I also have a smaller leaderboard which lists the top three students in the course by number of points earned. Students who are in the top three at the end of the unit receive bonus tokens. This serves to create a friendly competition among students, with small but meaningful rewards available. Ideally, it should serve to further motivate high-achieving students, while not raising the stakes so high as to undermine comradery.

Username	Points	Level	Tokens	Progress to 1000 pts.	Progress to Level 15	Progress to Level 25
	13	0	5	<div></div>	2%	1%
	23	0	5	<div></div>	4%	2%
	31	1	7	<div></div>	5%	3%
	18	0	5	<div></div>	3%	2%
	62	2	7	<div></div>	10%	6%
	73	2	7	<div></div>	12%	7%
	72	2	7	<div></div>	12%	7%
	85	3	8	<div></div>	14%	9%
	70	2	8	<div></div>	12%	7%
	49	1	6	<div></div>	8%	5%
	7	0	7	<div></div>	1%	1%
	84	3	8	<div></div>	14%	8%
	44	1	5	<div></div>	7%	4%
	25	1	6	<div></div>	4%	3%
	80	3	8	<div></div>	13%	8%
	57	2	6	<div></div>	10%	6%
	70	2	8	<div></div>	12%	7%
	81	3	8	<div></div>	14%	8%
	24	0	5	<div></div>	4%	2%

Figure 2 – The Course Leaderboard. Students usernames are shown in the first column (not shown to protect anonymity), followed by their accumulated points, level, and tokens. There is then a progress bar showing progress toward 1,000 points (an A grade) as well as percentage progress indicators toward level 15 (the minimum needed to pass the course) and level 25 (an A grade). This snapshot of the leaderboard was taken during the third week of the course, prior to most assignments being completed.

Summary. The RPG method uses a gamification approach to encourage student motivation and engagement. Specifically, the points system is structured to encourage students to view their performance as unidirectional toward progress and a higher grade. All students begin at zero points and accumulate points throughout the course. Students never see a percentage score and are instead encouraged to focus on gaining levels at certain point milestones, which will eventually be converted to a letter grade at the end of the course. There are many more points available than are needed and specifications grading on written assignments encourages accountability and upholds high standards. The amount of points also allows students to choose to complete assignments that are of most interest to them. Students are rewarded for gaining levels through the accumulation of tokens, which can be used for a variety of benefits within the course. **Table 2** displays a summary of the gamification elements present, mapped to Huizinga's (1938) framework of play.

Table 2 – Summary of Game Elements with the RPG Method

Game Element (Huizinga 1938)	Course Element
Choice	Choice of assignment types; flexible due dates; tokens can be used for many different purposes
Rules	Progression clearly defined; assignments evaluated using specifications grading checklists
Pathways	Multiple ways to earn points and (eventual) grades
Community	Course leaderboard
Feedback	Clear expectations through specifications grading; Rewards for progression

Study 1: Student Outcomes Pre-/Post- RPG Implementation and Student Satisfaction

Method

I first conducted a pilot of the RPG method during the 2018-19 academic year. I implemented the method in two sections of my introductory American politics and government course during the fall and spring semesters. To examine the initial efficacy of the method, I administered surveys at three different time points each semester. The first survey was given to students around the midpoint of the semester, approximately six weeks after the course began. At this point, students were familiar with the grading method and were able to complete a variety of assignments, including one exam. The second survey was delivered near the end of the course at approximately the eleventh week. By this time, students had completed a majority of the course and were able to predict (approximately) how well they would ultimately do in the course. The third survey was administered immediately after students completed the final course exam. Students at this time point were aware of their total points earned and their final grade. The surveys were delivered online (outside of class) and students received extra credit points for participating in the study. A total of 107 students completed the course and a majority completed each survey. Specifically, 69 students completed the first survey (64% completion rate), 62 students completed the second survey (58% completion rate), and 63 students completed the third survey (59% completion rate).

The surveys included measures of satisfaction with the grading system, academic motivation, and other attitudes. I also include students' final course grades within the dataset, in order to compare the distribution of grades obtained with the RPG method to the previous academic year (prior to implementation of the method). While this data is limited in the sense

that the responses to the survey items cannot be compared to previous semesters, it is instructive in gauging students' response to this novel method of evaluation.

Results

Course Outcomes. As shown in **Figure 3**, the distribution of course grades differs quite dramatically, and significantly ($\chi^2=9.81, p=0.04$) before and after implementation of the RPG method. The most important difference is the number of failing grades. The course failure rate pre-RPG implementation was 18.9%. However, post-implementation the failure rate fell to 10.3%, nearly a 50% reduction. This is a substantively important result that speaks to the efficacy of this method of student evaluation. The result is underscored by the fact that there is no statistically significant difference in the mean course grade. On an eleven-point scale, from F (0) to A (10), the mean course grade pre-implementation was 5.7 (between a C+ and B-) and 4.7 (between a C and C+) post-implementation, a non-significant difference ($t(1, 158)=1.62, p=0.11$). Thus, while the course failure rate fell quite dramatically, it does not appear to be due to an overall increase in course grades. In essence, grade inflation did not occur following the implementation of the RPG method.

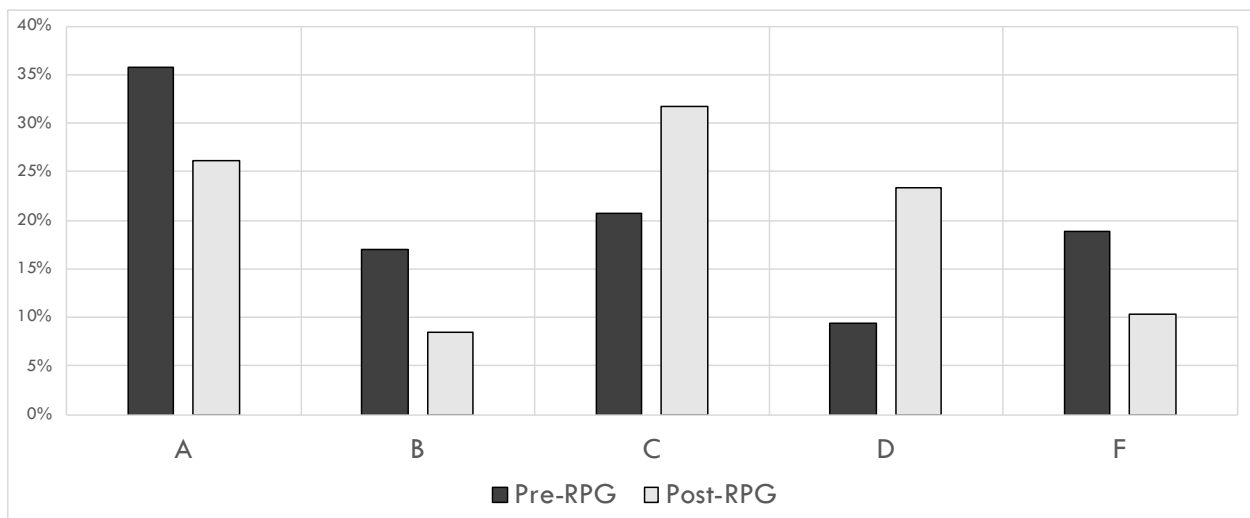


Figure 3 – Distribution of Course Grades Pre-/Post- RPG Implementation.

In fact, the post-implementation distribution is much more normal in shape than the pre-implementation distribution. The modal grade post-implementation is a C (on a five-point A-F scale) while the modal grade pre-implementation was an A. It would thus appear that the RPG method tends to draw course grades toward the center of the distribution, away from the poles (A and F). While I would, of course, prefer to assign higher grades, and I delight in awarding A grades to students who earn them, I believe that the number of A and B grades pre-implementation may have been the result of curving overall course grades in an effort to prevent an ever higher failure rate. I did not need to curve grades after implementing the RPG method as students were given greater agency to achieve a higher grade. That is, I felt confident that the grades students achieved at the end of the course were reflective of their progress and effort.

Student Satisfaction. Overall, students expressed high satisfaction toward the grading system. As shown in **Table 3**, a large majority of students agreed that the RPG method is easy to understand (74.6%) and understood how they would be evaluated (90.5%). This is encouraging given the novelty of many aspects of this evaluation method. Additionally, 73% of students preferred the method to a more traditional (points) method and 69.8% agreed that they would be more likely to sign up for a class that utilizes the RPG approach. Likewise, 73% agreed that the grading system motivated them to do their best and 76.2% agreed that the grading system made the course more enjoyable. Despite a large majority (79.4%) of students feeling in control of their grade, 65.1% still reported anxiety about their grade. While I hoped that this method would reduce anxiety regarding grades, it would appear that many students continue to feel anxious. However, given that I am unable to compare anxiety before and after implementation, it is possible that anxiety is lower than it would have been using a more traditional evaluation method.

Table 3 – Student Attitudes toward the RPG Method

Statement	Mean (7-point scale)	% Agreement
I find the grading system in this course easy to understand	5.14	74.6
I understand how my work will be evaluated in this course	5.63	90.5
I prefer the grading system in this course compared to the traditional points method	5.33	73.0
The grading system in this course motivates me to do my best	5.17	73.0
I would be more likely to sign up for a class if it used this type of grading system	5.22	69.8
The grading system makes this course more enjoyable	5.38	76.2
I feel in control of my grade in this course	5.46	79.4
I feel anxious about my grade in this course	4.85	65.1

Discussion

Overall, these preliminary results are quite encouraging. The course failure rate was reduced by nearly 50% following the implementation of the RPG method of evaluation. At the same time, the mean course grade did not differ pre-/post-implementation, which indicates that the observed reduction in the course failure rate was not a result of grade inflation. Further, students were highly satisfied with the new grading method, reporting that it was easy to understand, promoted autonomy regarding their grade, and was motivating. At the same time, student anxiety toward their grade was somewhat high, despite strategies to reduce a fixation on grades. This speaks to how grading is so ingrained in students' mindsets toward learning. Yet, because surveys were not distributed to students prior to implementation of the RPG method, it is difficult to know whether anxiety was reduced. Additional research and strategies are clearly needed to address student anxiety toward evaluation and the negative effects that result. The

following study will be able to make a direct comparison between students' attitudes in a course using the RPG method versus a more traditional points-based grading method.

Study 2: Natural Experiment

Method

During the Fall 2019 semester, I conducted a natural experiment to more directly evaluate the efficacy of the RPG method compared to a more traditional evaluation method. Specifically, I implemented the RPG method in only one section of my introductory American politics and government course (treatment). In the second section of the course, I utilized a more traditional, points-based method of evaluation (control). The sections of the course were offered on the same days of the week (Monday and Wednesday) and at similar times (11:30am-12:30pm [control] and 12:40-1:40pm [treatment]). The students in each section were similar across many factors, including year in school, age, gender, and race, as shown in **Table 4**.

Table 4 – Student Demographic Characteristics by Course Section

	Treatment Section	Control Section	Difference
Year in School	<ul style="list-style-type: none"> • 54% freshmen • 31% sophomores • 12% juniors • 4% seniors 	<ul style="list-style-type: none"> • 58% freshmen • 37% sophomores • 4% juniors • 0% seniors 	$\chi^2=1.98, p=0.58$
Age	18.61 years	19.09 years	$t(1, 44)=1.37, p=0.18$
Gender	<ul style="list-style-type: none"> • 58% male • 42% female 	<ul style="list-style-type: none"> • 67% male • 33% female 	$\chi^2=0.43, p=0.51$
Race	<ul style="list-style-type: none"> • 38% white • 62% non-white 	<ul style="list-style-type: none"> • 38% white • 62% non-white 	$\chi^2=0.01, p=0.94$

Every aspect of the course was kept similar across each section, with the exception of the grading system. In addition to similar class times and student characteristics, the course lectures

were all delivered electronically as audio files, thus ensuring identical content coverage. Specifically, the course utilized a flipped classroom approach whereby students listened to lectures prior to each class session and then came to class to participate in active learning activities, such as discussions, debates, and group activities.

The grading method in the control section of the course was designed to mimic the RPG method in many ways, without the gamification elements. Specifically, assignments were grouped by type (e.g., exams, writing assignments, debates) with students being required to complete a certain number of assignments in each category. For example, students were required to participate in five (of seven) total online debates on the course message board. Students could either complete all seven of these assignments, and have their lowest two scores dropped, or choose to complete only five (with the remaining “zero” scores dropped). Thus, similar to the RPG method, students still had some choice in determining which assignments to complete. Further, the assignments were the same across both sections and were evaluated in the same way (using a specs checklist). However, in the control section, each spec was assigned a certain point value and students would receive a percentage score on each assignment and were not allowed to resubmit. Overall, I believe that this approach affords a great deal of control over all aspects of the course, with the gamification elements being the only explicitly manipulated factor.

As in the previous study, students completed a series of three surveys throughout the semester. The surveys contained similar items measuring student attitudes toward the course and grading method, an intrinsic motivation inventory (Ryan 1982; Ryan, Mims, and Koestner 1983) as well as items measuring performance approach and avoidance (Elliot and McGregor 2001). The first survey was completed around the course midpoint (weeks 6-7), the second around two-thirds through the course (weeks 10-11), and the third after final course grades were determined

(week 15). In total, 50 students completed the course and 45 students (90%) completed the first survey, 33 students (66%) completed the second survey, and 38 students (76%) completed the third survey.

Results

Course Outcomes. The distributions of final course grades, by section of the course (treatment vs. control), are shown in **Figure 4**. These distributions are quite similar to those shown in the previous study, with a few notable differences. First, and most importantly, the failure rate in the treatment section was only about one-third of the failure rate in the control section. Specifically, 34.6% of students in the control section failed the course while only 12.5% of students in the treatment section failed. This is a substantively enormous difference and is much larger than the difference in the previous study. Second, the number of A grades was slightly higher in the treatment section (33.3%) than the control section (27%), though this difference is not particularly large. Like the previous study, grades in the treatment section tended to gravitate toward the center, with 33.3% of students in the treatment section earning a grade of C, compared to 11.5% in the control section. Indeed, the distributions are both bimodal, with the modal grades in the treatment section being an A and a C (33.3% each), and the modal grade in the control section is an F (35.6%), followed closely by an A grade (27%).

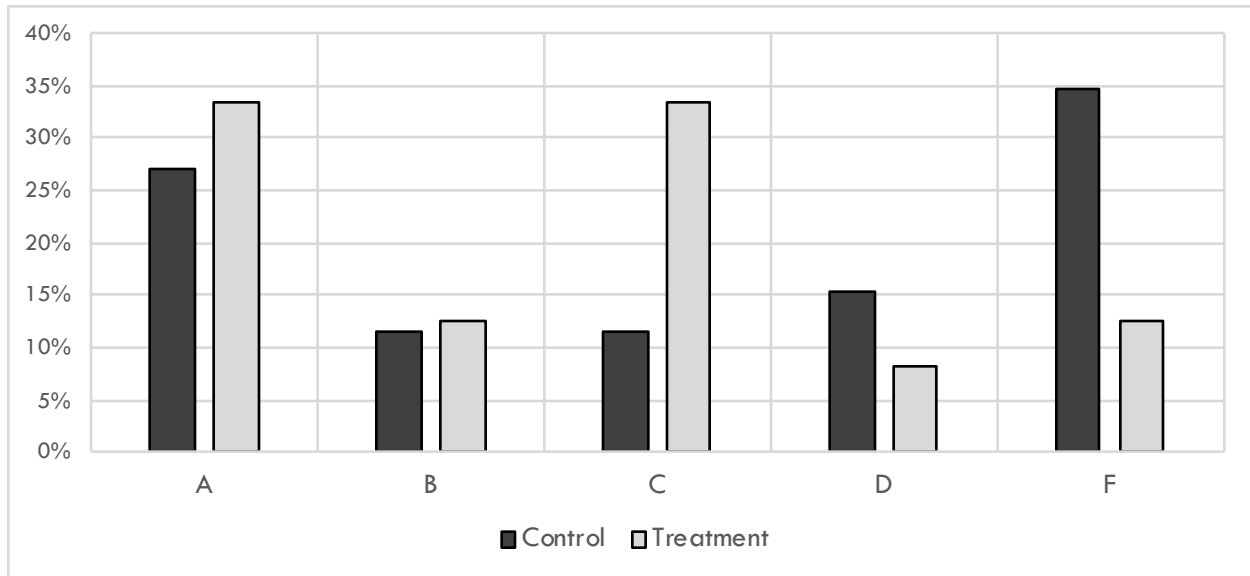


Figure 4 – Distribution of Course Grades by Section

To better understand how the RPG method resulted in changes to the final course grades achieved by students, I conducted a logit model predicting likelihood of course failure and an OLS regression model predicting course grades (on an eleven-point F (0) to A (10) scale). The primary covariate in each equation is section of the course (1=treatment section). I also included three controls: age, gender (1=male), and race (1=non-white). The coefficients for each model are plotted in **Figure 5**.

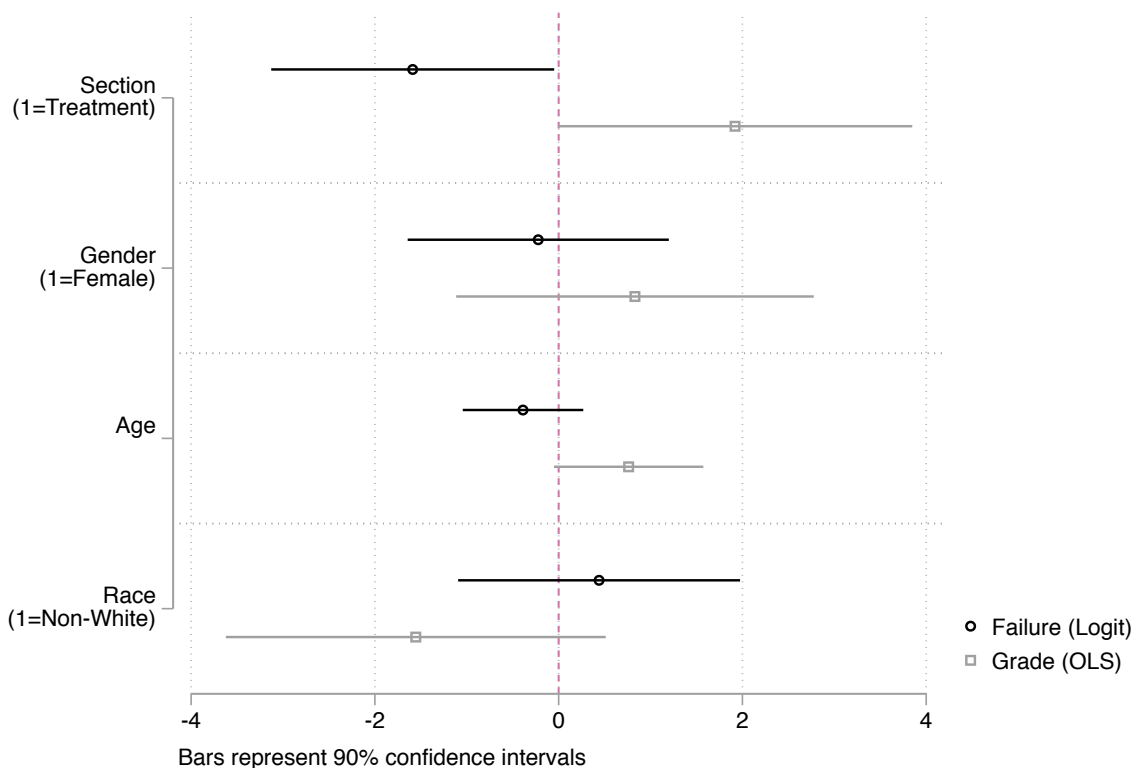


Figure 5 – Predicting Course Failure and Grade by Section and Demographic Characteristics

As can be seen, there is a significant difference in likelihood of failure between the two sections of the course ($p=0.09$). Students in the control section of the course had a 28.5% chance of failing the course while students in the treatment section had only a 7.9% chance of failing the course, a reduction in likelihood of 72.3%. In terms of overall grade, students in the treatment section were predicted to have a slightly higher course grade, though this effect did not quite reach significance at conventional levels ($p=0.102$). The mean grade (on the eleven-point scale) in the treatment section was 5.79 (between a C+ and B-) and 4.15 in the control section (between a C and C+). This is in contrast to the previous study, whereby the mean course grade in courses using the RPG method were slightly, though non-significantly, lower.

Regarding the differences in failure rate, I believe that the primary reason for such a high failure rate in the control section, compared to previous versions of the course, is the sheer number of potential assignments. One of the primary goals of the RPG method is to provide students with a great degree of choice regarding which assignments to complete. However, when framed in the context of a more traditional grading method, students in the control section appeared to be overwhelmed. While students were expected to only complete a certain number of assignments in each category, most students who failed the course did so simply because they did not complete many assignments at all.

Student Motivation. I examine student motivation in two ways. First, I included with the third survey an intrinsic motivation inventory (Ryan 1982; Ryan, Mims, and Koestner 1983). This measured student motivation across three areas. The first is perceived competence, which concerns students' perceived ability to perform well in the course. The second is effort, which measures the degree of effort that students exhibited throughout the course. Finally, the third is interest/enjoyment, or the degree to which students found the course to be enjoyable and of interest to them. I employed OLS regression to predict students' motivation in these three areas based on their section of the course (1=treatment) while again controlling for age, gender, and race.

The coefficients for the three models are plotted in **Figure 5**. As can be seen, students in the treatment section perceived themselves to be significantly more competent than those in the control section ($p=0.04$). While reported effort was greater in the treatment condition, it does not quite acquire statistical significance ($p=0.12$). Finally, there was also no statically significant difference in interest/enjoyment between the two sections ($p=0.41$).

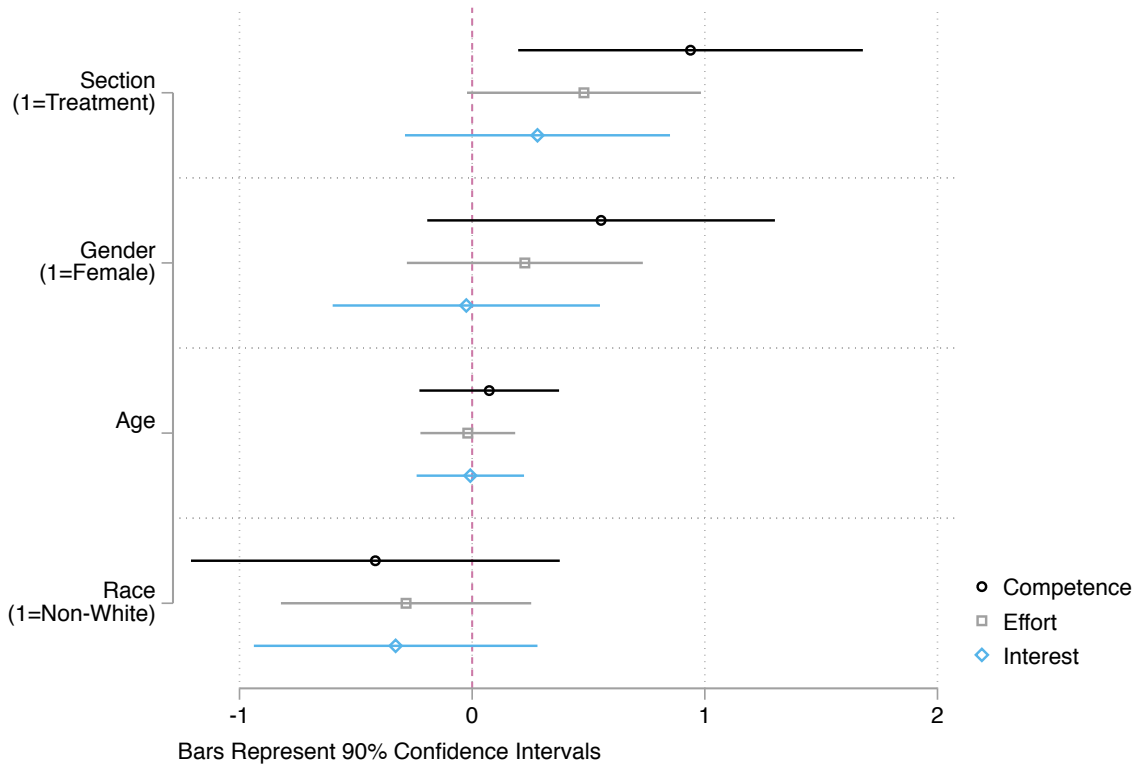


Figure 5 – Predicting Motivation by Section and Demographic Characteristics

I also included items measuring students' attitudes toward goal achievement using Elliot and McGregor's (2001) 2×2 framework. Specifically, these items measure whether students are motivated more by obtaining competence/success (approach) or avoiding incompetence/failure (avoidance). These are measured in two domains: performance, or normative competence defined externally (i.e., achievement of a grade), and mastery, or intrapersonal competence defined internally (i.e., feeling knowledgeable). I again conducted OLS regression models predicting motivation within these four domains based on section (1=treatment) and controlling for gender, age, and race.

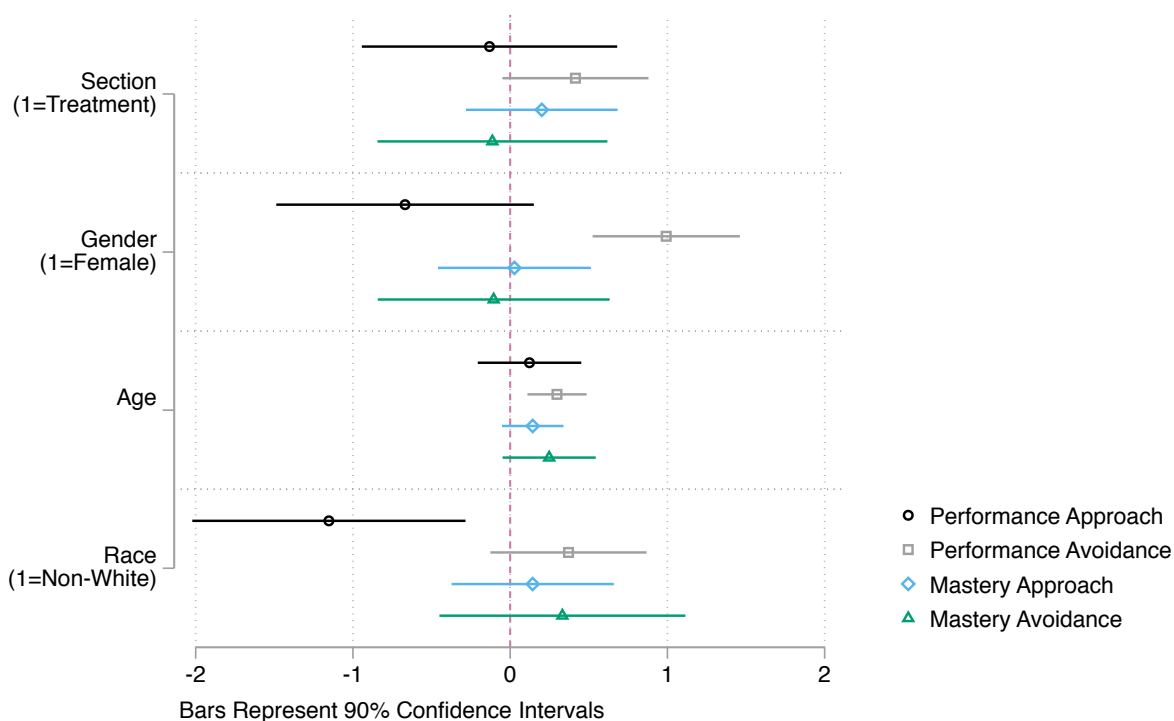


Figure 6 – Predicting Performance and Mastery Approach/Avoidance by Section and Demographic Characteristics

As shown in **Figure 6**, there are no statistically significant effects based on section, nor does there appear to be an intelligible pattern of results. Students reported slightly, though non-significantly, higher endorsement of performance avoidance and mastery approach goals in the treatment section, but also reported lower endorsement of performance approach and mastery avoidance goals. However, some interesting effects do emerge in regard to the control variables. Female students appear to be particularly motivated by performance avoidance ($p < 0.01$), as well as older students ($p = 0.01$). In contrast, non-white students show lower performance-approach

goals than white students ($p=0.03$). These differences speak to the particular motivational challenges faced by female and non-white students.

Discussion

The results of the natural experiment provide further evidence of the efficacy of the RPG approach to student evaluation. Students taking the course with the RPG method were far less likely to fail the course and walked away from the course with a greater sense of competence compared to those in the section with a more traditional method of evaluation. Specifically, students' likelihood of failing the course was reduced by over two-thirds in the RPG section of the course. Likewise, grades were not significantly higher in the RPG section, indicating again that this method can improve course outcomes without artificially inflating grades.

In addition to experiencing a much lower likelihood of course failure, students in the RPG section perceived themselves as more competent. Specifically, they viewed themselves as performing well, being skilled within the course, and generally feeling better able to perform course-related activities and assignments. This is another important result: one of the primary goals of the RPG method is to motivate students to do their best and to likewise reward them with a sense of accomplishment and empowerment as they progress within the course. While overall perceived competence was higher in the RPG section, it remains unclear how this competence was generated, as students in each section displayed no significant differences in their performance and mastery approach/avoidance goals. Clearly, additional research is needed in this domain to identify how students' goals drive motivation within the RPG framework.

Conclusion

At its core, Reflected Progression Grading seeks to better engage and motivate students within a general education setting. Students within general education courses are often

demotivated in two ways: 1) lack of interest in the course material (with a preference toward taking more courses related to their major), and 2) through the negative effects associated with letter grades. The RPG approach attempts to address both of these problems simultaneously through gamification. Gamified elements embedded within the course seek to draw students into the course material and to reduce the stigma arising from letter grades. The results presented here demonstrate the efficacy of this methodology, though certainly additional research is needed.

As with any evaluation method, there are advantages and disadvantages to the RPG approach. There are some downsides associated with this method, though I believe that the positive effects far outweigh them. First, setting up a grading system like this one requires a substantial initial time investment, as well as ongoing tweaks. The instructor must think like a game designer and carefully consider all aspects of the game elements embedded within the course. Point values for activities must be carefully determined, rewards must be meaningful and beneficial (though not too much so), and high academic standards must be maintained. Additionally, it is important to leverage technology to streamline the game processes, so as not to create an undue time burden on the instructor throughout the duration of the semester.

While my course includes many somewhat complex elements, I believe that many of the demonstrated benefits included here could be integrated in a much simpler design. For example, simply framing point accumulation as a journey of continuous accrual should be greatly beneficial to students. While the wide-scale implementation of online gradebooks allows students to carefully monitor their progress within their courses, students who see their grade fall over the course of the semester can easily become demotivated. I posit that students should instead only observe unidirectional, forward progress as they complete assignments. This could be implemented without necessarily including the level and rewards system that I use.

Overall, based on the results of this initial research, I strongly recommend that instructors think about how to restore the fun and joy that should result from learning. That is, I advocate for cultivating lusive attitudes within courses. While much of our course content is very serious (e.g., issues of social justice), and should be approached in a serious manner, this does not have to extend to the overarching course structure and evaluation method. We all begin our educational journey with a love of learning and, for most of us with advanced degrees, that love of learning continues. However, that enjoyment has been lost among many of our students. I propose that gamification provides one way to restore that enjoyment, and the approach presented in this paper provides a means to easily introduce gamification within many academic courses.

References

- Butler, Ruth. 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): 1–14.
- Butler, Ruth, and Mordecai Nisan. 1986. "Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance." *Journal of Educational Psychology* 78(3): 210–16.
- De-Marcos, Luis, Eva Garcia-Lopez, and Antonio Garcia-Cabot. 2016. "On the Effectiveness of Game-like and Social Approaches in Learning: Comparing Educational Gaming, Gamification & Social Networking." *Computers and Education* 95: 99–113.
- Dicheva, Darina, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. "Gamification in Education: A Systematic Mapping Study." *International Forum of Educational Technology & Society (IFETS)* 18(3): 1–14.
- Durm, Mark W. 1993. "An a Is Not an a Is Not an a: A History of Grading." *Educational Forum* 57(3): 294–97.
- Elliot, Andrew J, and Holly A McGregor. 2001. "A 2 x 2 Goal Achievement Framework." *Journal of Personality and Social Psychology* 80(3): 501–19.
- Harter, Susan. 1978. "Pleasure Derived from Challenge and the Effects of Receiving Grades on Children's Difficulty Level Choices." *Child Development* 49(3): 788–99.
- Huizinga, Johan. 1938. *Homo Ludens: A Study of the Play-Element in Culture*. Ketting, OH: Angelico Press.
- Humphreys, Debra, and Abigail Davenport. 2005. "What Really Matters in College: How Students View and Value Liberal Education. Liberal Education & America's Promise."

- Liberal Education* 91(3): 36–43.
- Inoue, Asao B. 2019. *Labor-Based Grading Contracts: Building Equity and Inclusion in the Compassionate Writing Classroom*. Fort Collins, CO: The WAC Clearinghouse.
- Iosup, Alexandru, and Dick Epema. 2014. “An Experience Report on Using Gamification in Technical Higher Education.” In *SIGCSE’14*, eds. J. Dougherty and K. Nagel. Atlanta, GA: ACM, 27–32.
- Maze, Jonathan. 2019. “Starbucks’ Rewards Program Pushes the Chain’s Growth.” *Restaurant Business Online*. <https://www.restaurantbusinessonline.com/marketing/starbucks-rewards-program-pushes-chains-growth> (January 26, 2020).
- Mueller, Jon. N.d. Retrieved from: <http://jfmueller.faculty.noctrl.edu/toolbox/portfolios.htm>
- Nilson, Linda. 2015. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Stylus Publishing, LLC.
- Pulfrey, Caroline, Céline Buchs, and Fabrizio Butera. 2011. “Why Grades Engender Performance-Avoidance Goals: The Mediating Role of Autonomous Motivation.” *Journal of Educational Psychology* 103(3): 683–700.
- Robson, Karen et al. 2015. “Is It All a Game? Understanding the Principles of Gamification.” *Business Horizons* 58(4): 411–20.
- Ryan, Richard M. 1982. “Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory.” *Journal of Personality and Social Psychology* 43(3): 450–61.
- Ryan, Richard M., and Edward L. Deci. 2000. “Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.” *American Psychologist* 55(1): 68–78.

- Ryan, Richard M., Valerie Mims, and Richard Koestner. 1983. "Relation of Reward Contingency and Interpersonal Context to Intrinsic Motivation: A Review and Test Using Cognitive Evaluation Theory." *Journal of Personality and Social Psychology* 45(4): 736–50.
- Vander Schee, Brian. 2011. "Changing General Education Perceptions Through." *International Journal of Teaching and Learning in Higher Education* 23(3): 382–87.
- Suits, Bernard. 1978. *The Grashopper: Games, Life, and Utopia*. Peterborough, ON: Broadview Press.
- Thompson, Clarissa A., Michele Eodice, and Phuoc Tran. 2015. "Student Perceptions of General Education Requirements at a Large Public University: No Surprises?" *Journal of General Education* 64(4): 278–93.
- Topîrceanu, Alexandru. 2017. "Gamified Learning: A Role-Playing Approach to Increase Student in-Class Motivation." *Procedia Computer Science* 112: 41–50.
- Zichermann, Gabe, and Christopher Cunningham. 2011. *Gamification by Design*. Sebastopol, CA: O'Reilly.