2020

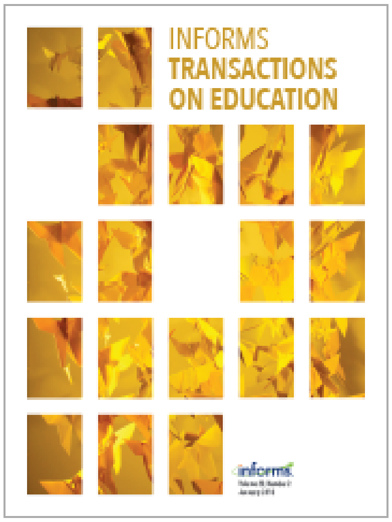
The Dark Side of Narrow Gamification: Negative Impact of Assessment Gamification on Student Perceptions and Content Knowledge

Hee Yoon Kwon

Koray Özpolat
University of Rhode Island, koray@uri.edu

Citation/Publisher Attribution

Hee Yoon Kwon, Koray Özpolat (2020) The Dark Side of Narrow Gamification: Negative Impact of Assessment Gamification on Student Perceptions and Content Knowledge. *INFORMS Transactions on Education*. https://doi.org/10.1287/ited.2019.0227



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Dark Side of Narrow Gamification: Negative Impact of Assessment Gamification on Student Perceptions and Content Knowledge

Hee Yoon Kwon, Koray Özpolat

informs
http://pubsonline.informs.org/journal/ited
INFORMS TRANSACTIONS ON EDUCATION
Articles in Advance, pp. 1–15
ISSN 1532-0545 (online)

The Dark Side of Narrow Gamification: Negative Impact of Assessment Gamification on Student Perceptions and Content Knowledge

Hee Yoon Kwon,[a] Koray Özpolat[b]

[a] Tommy and Victoria Baker School of Business, The Citadel, Charleston, South Carolina 29409; [b] College of Business, The University of Rhode Island, Kingston, Rhode Island 02881

Contact: hkwon@citadel.edu, https://orcid.org/0000-0001-9662-3859 (HYK); koray@uri.edu, https://orcid.org/0000-0001-9791-8966 (KÖ)

Received: October 16, 2019
Revised: March 22, 2020
Accepted: March 25, 2020
Published Online in Articles in Advance: September 1, 2020

https://doi.org/10.1287/ited.2019.0227

Abstract. We explored the effects of assessment gamification on students' content knowledge and perceptions of satisfaction, course experience, learning, and impact of teaching techniques. The course preparation, attendance, quizzes, classroom activities, and team projects of an undergraduate operations and supply chain management course had game elements that accumulate to team advantages in the collaborative midterm and final exam. Interestingly, we found that gamifying assessment activities resulted in significantly lower content knowledge, satisfaction, and course experience. Difference in perceived learning was not significant. Also, team exam scores were significantly lower in the gamified group, whereas individual exam scores were not significantly different. This study contributes to the literature by providing empirical evidence that gamification in classroom may produce unintended consequences and implementing gamification restrictively to assessment is ineffective at best. Directions for further research are discussed.

Keywords: pedagogical research • teaching supply chain management • assessment • gamification

"Games are showing us exactly what we want out of life: more satisfying work, better hope of success, stronger social connectivity, and the chance to be a part of something bigger than ourselves." —Jane McGonigal, *Reality Is Broken* (2011, p. 114)

1. Introduction

In recent years, gamification has become more popular and pervasive in various contexts as a means of embracing the benefits of positive human emotions invoked by games (Seaborn and Fels 2015, Dias 2017, Subhash and Cudney 2018). Gamification is noticed by academics, educators, and practitioners from a variety of domains, following a trend within the business and marketing sectors (Seaborn and Fels 2015). Specifically, gamification has been used and found to enhance motivation and improve user experience in brand loyalty (Zichermann and Cunningham 2011), healthcare and health awareness (Hamari and Koivisto 2015, González et al. 2016), management and training (Saunders 2017), and education and learning (Buckley and Doyle 2014, Stansbury and Earnest 2016). Gamification by itself has become an emerging segment in the industry, expected to grow to more than US$ 22.9 billion by 2022 (P&S Market Research 2016).

With a goal to bring some of these benefits to our classrooms, we gamified the assessment activities of an undergraduate level operations and supply chain management (OSCM) course. The course is mandatory to all students majoring in business and contains various concepts, constructs, and analytical content that requires high retention of student engagement for success. Specifically, reading the textbook to understand basic concepts and theories before lectures is crucial for student success, as well as to the efficient progression of the course. To this end, we implemented game mechanics and elements selectively to assessment activities of two sections of the OSCM course and measured the effects of gamification on student perceptions of satisfaction, course experience, learning, impact of teaching method, and student's content knowledge. Contrary to our expectations, gamification of assessment activities hurt students' content knowledge, satisfaction, and course experience. This study contributes to the literature by demonstrating a dark side of gamification in a classroom setting and by exploring the effects of selectively gamifying a course design element (assessment) with no online platform involved.

The rest of the paper is structured as follows. First, an overview of gamification and its application in higher education is provided, followed by the theoretical background for gamification. The basis for selecting assessment activities as the focal area for gamification is clarified, accompanied by the experimental design and details on gamified assessment in the course. The method section outlines the data collection protocol and survey items. The findings are reported in the results section, followed by detailed analysis in the discussion section. Finally, major findings and implications are outlined in the conclusion, followed by limitations and directions for future research.

2. Literature Review

2.1. Gamification in Higher Education

Gamification is emerging as a research topic in higher education among various disciplines (Seaborn and Fels 2015). We adopt the definition of gamification as the "process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation" (Huotari and Hamari 2012, p. 19). Gamification differs from game-based learning (GBL) in that gamification uses elements and mechanics of videogames to improve user experience and engagement in nongame contexts (Aldemir et al. 2018), whereas GBL uses full-fledged games or videogames, often labeled *serious games*, incorporated to the curriculum activities, as described by Kong (2019).

Educators from various disciplines have demonstrated benefits of gamification in higher education with evidence of improved attitude, engagement, enjoyment, motivation, (perceived) learning, participation, practical skills, retention, satisfaction, and student performance (grades) (Aldemir et al. 2018, Subhash and Cudney 2018). Subhash and Cudney, through a systematic review of the literature focusing on gamified learning in higher education, reveal a growing number of research in recent years, as well as business-related research being the second most published subject area after computing (Subhash and Cudney 2018).

A large body of gamification research has focused heavily on technology to create gamified experiences (Stansbury and Earnest 2016). In the education literature, the majority of gamification research in higher education entail the use and/or development of a dedicated online platform or a gamified learning management system (G-LMS) (Villagras et al. 2014, de-Marcos et al. 2017, Dias 2017). This focus and reliance on technology may be hindering the wider adoption of gamification in higher education, because of the difficulty of designing and managing a complex information system (Sobocinski 2017). There is a paucity of research in how gamification can be implemented without using or developing a complex technological system. Also, research on gamification of specific course design elements is desired but remains scarce (Nacke and Deterding 2017) because most are focused on gamifying the full-scale course design and information systems development, which can be an overwhelming task and burden for any single instructor exploring the benefits of gamification without organizational support (Sobocinski 2017).

Two recent exceptions stand out from these trends of technology dependence and full-scale points, badges, and leaderboards (PBL) approach in gamification research. Song (2017) designed a smaller-scale implementation of gamification focusing on the element of asking questions, examining the engaging effects of gamification within the individual and social interactions of the classroom (Song 2017). Morillas Barrio et al. (2016) found positive effects of gamified student response system (SRS) on student motivation, attention, and learning performance. Their work also presented a novel path of gamification in higher education by implementing gamification around another innovative technique to enhance its benefits.

2.2. Theoretical Background

The majority of gamification research in the educational environment identify student engagement and motivation as core behavioral benefits of using gamification (Hew et al. 2016, Kuo and Chuang 2016, Subhash and Cudney 2018, van Roy and Zaman 2018). Popularly cited in this stream of research is self-determination theory (SDT), which argues that "an understanding of human motivation requires a consideration of innate psychological needs for competence, autonomy, and relatedness" (Deci and Ryan 2000, p. 227). Ryan and Deci (2000) identify the needs for competence, relatedness, and autonomy as essential for facilitating optimal functioning of the natural propensities for growth and integration, as well as for constructive social development and personal well-being. Deci and Ryan (2000) present a self-determination continuum, projecting type of motivation, type of regulation, and perceived locus of causality affecting varying degrees of self-determination. Intrinsic motivation is projected as most desirable type of motivation, in the extreme of self-determined behavior. Deci and Ryan (2000) also present intrinsic motivation, autonomous regulation of extrinsic motivation, and intrinsic aspirations as antecedents of high-quality performance, healthy behavior, and positive experiences.

Gamification research grounded in SDT can be categorized into structural and content gamification, according to the type of motivation (Kapp et al. 2014, Hudiburg 2016). Structural gamification relies on external motivation by adding game elements (e.g. points, badges, leaderboards) without altering content, whereas content gamification uses intrinsic motivation by applying game elements, mechanics, and thinking to the content, making an activity more game-like (Kapp et al. 2014, Hudiburg 2016).

In the educational gamification research, intrinsic motivation is widely cited from SDT as desired outcome (Hew et al. 2016, Kuo and Chuang 2016, Subhash and Cudney 2018, van Roy and Zaman 2018). Intrinsically motivated students are more engaged, retain information better, and are generally happier (Hanus and Fox 2015). Intrinsic motivation is desired because the desire to learn comes from within the student, whereas extrinsic motivation is less desired, where the motivation is because of some outside force (Deci and Ryan 2000, Hanus and Fox 2015). In an empirical study of graduate students, Hew et al. (2016) find that students in the gamified course chose more difficult assignments and produced higher-quality artifacts than those in the nongamified course. We design assessment activities of the coursework with game mechanics following the suggestion of SDT on competence, relatedness, and autonomy.

3. Research Design

3.1. Design Elements and the Course Choice

Considering the advice of many scholars on the need for careful design and alignment (de-Marcos et al. 2016, Fitz-Walter et al. 2017, Sobocinski 2017, Aldemir et al. 2018), the ideal scenario for a course gamification design would be where the instructor can survey the students in advance to identify their needs and goals, thus customizing the gamified course design accordingly. In reality, however, most university students enroll in courses via an online system, and the course membership tends to change until the semester begins, as well as until weeks into the semester with drops and switches. Thus, the instructor rarely has a chance to meet or communicate with all of the enrolled students effectively until the semester begins. The alternative, then, is to identify the needs and goals that are vastly common among university students.

In higher education, students are exposed to high levels of stress and anxiety revolving around their performance in assessments, such as quizzes, reports, and exams (Kapitanoff 2009, Dahlström 2012, Cantwell et al. 2017, Johanns et al. 2017, Levine et al. 2017, Khansari and Coyne 2018). In this research, we apply gamification on the course element of assessment activities with a reward structure that connects them toward midterm and final exams to explore its effects on students' content knowledge and perceptions on satisfaction, experience, learning, and impact of teaching methods. The focus on student satisfaction, explored by Reinig et al. (2011), addresses the remark of Jassawalla et al. (2009, p. 43): "What is curiously missing in the rich body of research is the perspective of the student?"

Finally, we decided to gamify the introductory OSCM course, which is one of the most challenging courses to teach business students. Similar to the motivation of Kong (2019) to choose a modeling and simulation course for game-based teaching, our choice of gamifying OSCM reflects our observations of business students being challenged by the analytical content (e.g., statistical process control, safety stocks, forecasting) in this course and the increased need to better engage students.

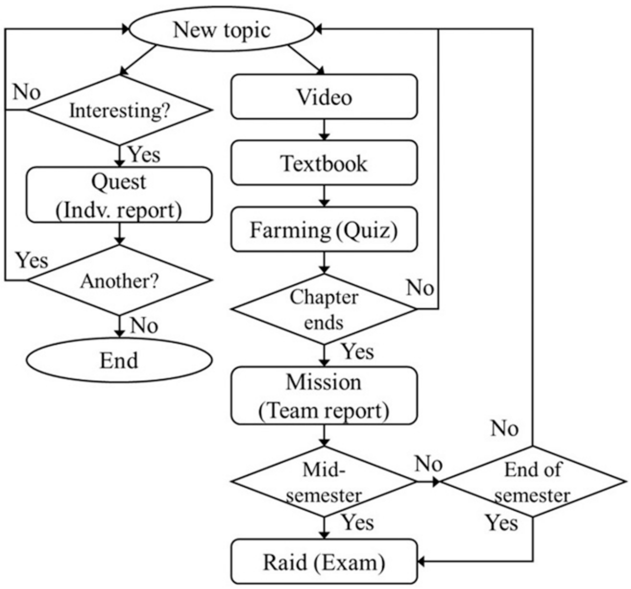
3.2. Experimental Design

For this study, two OSCM courses were designed. Both Section 1 (10 a.m.) and Section 2 (1 p.m.) were taught three times a week (Monday, Wednesday, and Friday) for 50 minutes each. Section 1 was designated as the experimental (gamified) group, whereas Section 2 was designated as the control (nongamified) group. The two sections had the same instructor, course structure, learning objectives, assignments, quizzes, term projects, and exams. Section 1's curriculum had (i) a gamified structure of assessment activities and (ii) an overlay of videogame nomenclature integrated to the assessment activities, following the application by Lieberoth (2015) and Stansbury and Earnest (2016).

Both sections were administered following a flipped classroom model (Herreid and Schiller 2013, Asef-Vaziri 2015), where brief lectures were delivered to students via prerecorded videos, uploaded to YouTube. Students from both sections had access to the same videos at the same time. The length and content of videos were limited so that they function as a study guide rather than a full lecture. Students were required to complete class preparation by reading the textbook and watching the study guide video before the class meeting.

Figure 1 depicts the progression of assessment activities in a flowchart. In the first weeks of the semester, students were assigned to teams by the instructor to ensure membership diversity in terms of major, sex, age, and nationality; also, team membership was maintained throughout the semester (Koppenhaver and Shrader 2003). In both sections, a short quiz of five questions was administered at the beginning of each class session as formative assessment of content knowledge and attendance taking. The quiz questions tested knowledge of key terminology and concepts. After finishing each chapter, student teams were required to produce chapter reports in one of the following formats: writing, presentation slides, infographic, or video. Teams were required to identify key concepts and topics from the chapter and apply the knowledge to a real-world context.

Figure 1. Flow of Assessment Activities



In both sections, one individual research report per semester was mandatory. Students were rewarded extra credit for each optional extra research report on some emerging supply chain topics.

The exams were designed as two-stage collaborative tests in both sections. Students were required to first individually answer questions of the individual exam and then converge as a team to answer the questions of the team exam. The midterm exam consisted of 15 individual questions and 15 team questions. The final exam consisted of 20 individual questions and 20 team questions.

3.3. Gamification in Course Assessment Activities

Gamification was applied to this course by structuring assessment activities to build up toward the midterm and final exams with team rewards and by adding a layer of videogame-inspired nomenclature. Figure 2 depicts the gamified structure of assessment activities.

The quizzes were labeled *Farming* and counted toward individual grades. Also, each team earned a ticket as a reward if all members of the team were present and scored higher than 60%, and the team average score is higher than 80%. The individual research report was labeled *Quest*. Additional to individual grades and extra credit, a team ticket was rewarded for every additional individual report. The team chapter reports were labeled *Missions*. On-time, high-quality output rewarded team tickets additional to team grade credits.

The midterm exam and the final exam were labeled *Season 1 Boss Raid* and *Season 2 Boss Raid*, respectively. The individual exam was labeled *singleplayer*, and the team exam was labeled *multiplayer*. For the multiplayer, teams had conditional access to unique advantages, labeled *power-ups*, such as one textbook, one calculator, and one page of handwritten notes. These power-ups were available for teams to purchase as a result of the team's performance and member contributions in Farming, Missions, and extra Quests, thereby connecting the basic activities of watching lecture videos and reading the textbook with the exams. Before each Raid (exam), teams were given ample opportunity to win more tickets than needed to purchase all three power-ups. In the control group, in contrast, one textbook, one calculator, and one page of handwritten notes were readily accessible to teams during the team exam. Table 1 provides a comparison of how gamification was applied to assessment activities in the course, as well as the theoretical focus within SDT in each element.

Farming (quizzes) promotes competence through positive feedback. The quiz items ask basic conceptual questions, which is easy to answer if the student watched the video and read the textbook. The team reward system from Farming promotes relatedness. Missions (team chapter reports) promote all three focal areas of SDT: competence through adjustment of difficulty adequate to chapter progression (start with summaries and move on to case studies), relatedness through teamwork and team reward, and autonomy through meaningful choices for report format and case topic. Quest (individual research report) promotes autonomy through meaningful choice in topic

Figure 2. Assessment Gamification Structure

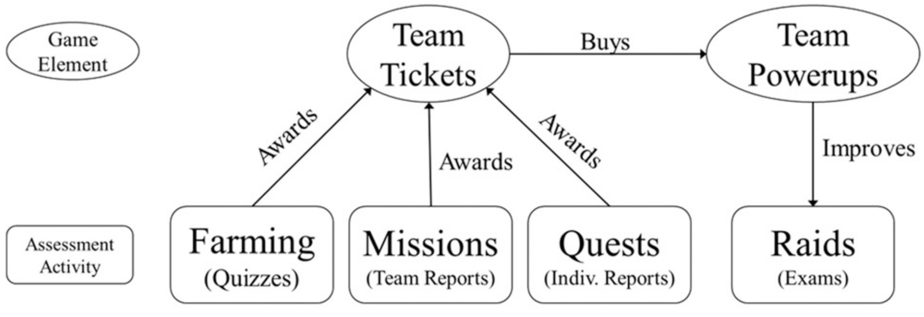


Table 1. Gamification Elements Applied to Course

Course component/ gamified term	Nongamified course function	Gamified course function	SDT focus
Quiz/Farming	Proof of work, attendance	Proof of work, attendance, team tickets for power-ups	Competence, relatedness
Team chapter report/ Mission	Team grade	Team grade and team tickets for power-ups	Competence, relatedness, autonomy
Individual research report/Quest	Individual assignment, extra credit	Individual assignment, extra credit, and team tickets for power-ups	Autonomy, relatedness
Midterm and final exams/ Season 1 and 2 Boss Raids	No structure in question order	Progressive difficulty	Competence
	1 textbook, 1 calculator, 1 page notes available for teams	Power-ups earned with team tickets from quizzes, quests, raids	Autonomy, relatedness

selection, number of additional reports, and timing of submission, as well as relatedness through contribution to team reward for additional reports. Boss Raids (exams) promote competence through the progressive difficulty of questions, relatedness through collaborative exam, and autonomy through meaningful choice regarding which Power-up to purchase with the team reward tickets.

3.4. Participants

The study took place at a university in the northeast of the United States in the course of one semester in the fall of 2018. Sixty-two undergraduate students participated, divided into two sections of 33 (Experimental, gamified) and 29 (Control, nongamified) by enrollment. In order to control for selection bias, the online enrollment system displayed the same generic course description, standard to the College of Business, for the two sections. This way, students were not exposed to the pertinent experimental conditions and manipulations before enrolling. Information on the course design was only given after the semester started. OSCM is an introductory course that is required for all business major students of the university and is also open to students of other majors that satisfy the prerequisites: 19.35% of the students in the current study were female; 6.45% were freshmen; 29.03% were sophomores; 53.23% were juniors; and 6.45% seniors. Specialized major areas of the students included supply chain management (16.13%), finance (22.58%), marketing (16.13%), accounting (6.45%), management (8.06%), chemical engineering, and communications.

3.5. Measures and Procedures

We tested the effect of assessment gamification on student's content knowledge, satisfaction with learning method, course experience, perception of learning, and impact of teaching techniques, following the survey methods of Stansbury and Earnest (2016) and Reinig et al. (2011), as detailed later. Content knowledge, course experience, perception of learning, and impact of teaching techniques were replicated or modified from Stansbury and Earnest (2016). Satisfaction with learning method was modified from Reinig et al. (2011).

3.5.1. Content Knowledge Measures. The content knowledge of students was calculated as the mean score of each student's total quiz and exam mean scores. At the beginning of every class session, a five-item quiz was given to students. Throughout the semester, a total of 24 quizzes were administered. The quiz items were either short-answer or multiple-choice questions on basic concepts of the session topic, which students were required to have studied in advance with the textbook and lecture video. Items on the quizzes were adapted from the test bank provided by the authors of the course textbook and were identical for both groups throughout the semester. Student performance on the quizzes was measured as a percentage of right answers. Because the quiz had a double purpose of attendance taking, the quiz grade penalizes absenteeism, resulting in an inaccurate measurement of student's content knowledge. We resolved this issue by only accounting for quizzes taken at the time of class, excluding absentees from the average.

The midterm and final exam were administered as a two-stage collaborative test with different sets of questions for each stage. Individual student performance in exams was measured as a percentage of right answers. The mean exam score was averaged with the mean quiz score to create a single variable *Knowledge*.

3.5.2. Formative Perception Measures. In the latter half of the semester, we conducted formative surveys in eight different time points assessing student's perceived satisfaction with learning method, perceived experience, perceived learning, and perceived impact of teaching techniques used in the course. For the formative survey instrument, a four-item perceived satisfaction survey from Reinig et al. (2011) was used. Also, a four-item perceived experience survey and a seven-item perceived learning survey were adopted from Stansbury and Earnest (2016).

A seven-item perceived impact of teaching method survey was adapted from Stansbury and Earnest (2016), with modifications to fit the context of course. Formative survey time points occurred over a 6-week period between the 10th and 15th weeks of classes. The analysis used 176 accumulated surveys from the experimental group and 167 accumulated surveys from the control group ($n = 343$). Table 2 provides a full list of survey items, constructs, and descriptive statistics.

Items for perceived satisfaction with the learning method were adapted from the survey design of Reinig et al. (2011) and stated, "I feel satisfied with the learning method used in this class," "I liked the learning method used in this class," "I would like to use this learning method in other classes," and "I was happy with the learning method used in this class." Responses were measured on a five-point Likert scale with end points: 1 = strongly disagree to 5 = strongly agree. All four items loaded onto a single construct. The items on this measure were highly reliable ($\alpha = 0.955$). The mean score for the four satisfaction items was calculated as perceived satisfaction with learning method variable *Satis*.

Items for perceived course experience were modified from the survey design of Stansbury and Earnest (2016) and asked students' honest feelings on their course experience in six areas: motivating, engaging, fun, boring (inverse measure for fun), challenging, and relevant. Responses were measured on a five-point Likert scale with end points: 1 = strongly disagree to 5 = strongly agree. The items on this measure were highly reliable ($\alpha = 0.813$). The mean score for the six experience items was calculated as perceived course experience variable *Exper*.

Items for perceived learning were adapted from the survey design of Stansbury and Earnest (2016) and asked, "When comparing this course to my traditional courses, I would rate this course as being…"

Table 2. Formative Survey Descriptive Statistics by Group

		Experimental (gamified) group ($N = 176$)		Control (nongamified) group ($N = 167$)	
Dependent measures		Mean	SD	Mean	SD
Perceived satisfaction with the learning method[a,*]		4.04	0.79	4.22	0.68
CS1	I feel satisfied with the learning method used in this class.	4.04	0.062	4.29	0.057
CS2	I liked the learning method used in this class.	4.06	0.064	4.25	0.057
CS3	I would like to use this learning method in other classes.	3.97	0.069	4.08	0.066
CS4	I was happy with the learning method used in this class.	4.09	0.063	4.26	0.055
Perceived course experience[b,*]		3.81	0.70	3.96	0.72
CE1	Motivating	3.79	0.062	3.98	0.063
CE2	Engaging	4.02	0.066	4.07	0.068
CE3	Fun	3.93	0.061	3.86	0.07
CE4	Enjoyable[c]	3.49	0.073	3.93	0.072
Perceived learning[a]		3.96	0.63	4.08	0.60
PL1	Increasing understanding of course content	4.12	0.055	4.32	0.056
PL2	Increasing my confidence	3.93	0.061	3.95	0.057
PL3	Keeping me involved in the classroom	3.94	0.064	3.94	0.065
PL4	Reinforcing key concepts	4.08	0.052	4.2	0.053
PL5	Motivating my learning	3.85	0.066	3.96	0.063
PL6	Developing my ability to reason	3.76	0.064	3.91	0.065
PL7	Increasing application of course content	4.04	0.059	4.26	0.058
Perceived impact of teaching technique[b]					
IT1	Group activities	5.39	1.12	5.51	1.15
IT2	Team projects	5.09	1.29	5.01	1.30
IT3	Quizzes	5.43	1.12	5.37	1.26
IT4	Professor*	5.52	1.26	5.81	1.28
IT5	Exams	4.59	1.46	4.80	1.45
IT6	Lecture videos***	4.18	1.63	4.72	1.42
IT7	Textbook***	5.59	1.30	4.90	1.77

[a]Five-point Likert scales: 1 = worse to 5 = better.
[b]Seven-point Likert scales: 1 = not at all to 7 = extremely.
[c]Recoded from inverse measure: Boring.
*$p < 0.05$; ***$p < 0.001$.

followed by seven items including "motivating my learning" and "reinforcing key concepts." Responses were measured on a five-point Likert scale with end points: 1 = strongly disagree to 5 = strongly agree. The items on the perceived learning measure were highly reliable (α = 0.934). The mean score for the seven perceived learning items was calculated as perceived learning variable *Learn*.

Items for perceived impact of teaching technique measures were adapted from the survey design of Stansbury and Earnest (2016) and asked students how much their learning in the course was aided by seven teaching techniques: classroom discussions, team chapter reports, quizzes, professor, exams, lecture videos, and textbook. Responses were measured on a seven-point Likert scale with end points: 1 = not at all to 5 = extremely. The items on the perceived impact of teaching technique measure were treated as individual variables, following the suggestion of Stansbury and Earnest (2016).

3.5.3. Summative Perception Measures. We also conducted a 27-item summative follow-up survey for the experimental section at the end of the semester after the final exam (n = 23). The summative perception survey contains 19 items on their overall perceived experience with the gamified course design (e.g., I enjoyed the gamified design of the class, the gamified design kept me engaged), six items on perceived efficacy of individual game elements (e.g., alteration of terms, reward system, and meaningful choices), and items on perception of videogames and learning (e.g., collaboration and teamwork, creativity and problem-solving) and learning style preference (competitive, cooperative, and individual). In the overall perceived experience measure and perceived efficacy of individual game elements measure, responses were assessed using a five-point Likert scale with end points: 1 = strongly disagree to 5 = strongly agree. The perception of videogames and learning measure provided a list from which to select all that applies. The learning style preference measure provided a list from which to choose only one that applies. The items on the summative follow-up overall experience measure were highly reliable (α = 0.95) and were explored to provide descriptive insights. Table 3 contains a full list of items and descriptive statistics.

4. Results

4.1. Content Knowledge

4.1.1. Quiz Scores. Results of a two-independent-samples t-test (Table 4) shows that the mean score for quizzes taken differs between the experimental (gamified) group (mean = 81.86, standard deviation (SD) = 7.74, n = 33) and the control (nongamified) group (mean = 86.39, SD = 7.67, n = 29) at the 0.05 level of significance (t = −2.31, df = 60, p = 0.024, 95% confidence interval (CI): −0.085, −0.006). On average, students in the nongamified control group scored higher in the quizzes than those in the gamified experimental group as seen in Figure 3.

4.1.2. Exam Scores. Scores from the midterm and final exams were divided into individual and team score from the two stages. A two-independent-samples t-test shows that the total mean score for both exams differs between the experimental group (mean = 70.71, SD = 6.59, n = 33) and the control group (mean = 76.41, SD = 7.46, n = 29) at the 0.05 level of significance (t = −3.20, df = 60, p = 0.002, 95% CI: −9.269, −2.133). On average, students in the control group scored higher in the exams than those in the experimental group. Item-level analysis of the t-test on exam scores (Table 4) shows that students in the experimental group scored significantly less in the team exams in both midterm and final exams, and their individual score mean is not significantly different in either exam. The experimental section's total mean scores of midterm (mean = 74.29, SD = 6.07, n = 33) and final (mean = 67.12, SD = 9.77, n = 33) exams were both significantly lower than the control group's midterm (mean = 78.51, SD = 7.94, n = 29) and final (mean = 74.31, SD = 9.28, n = 29) total mean scores (Table 4).

The content knowledge variable calculated from quiz and exam mean scores shows significant difference between experimental group (mean = 76.28, SD = 6.47, n = 33) and control group (mean = 81.40, SD = 6.43, n = 29) at a 0.05 level of significance (t = −3.12, df = 60, p = 0.003, 95% CI: −8.40, −1.83; Table 5).

4.2. Formative Perceptions

4.2.1. Perceived Satisfaction with Learning Method. Results of the two-independent-samples t-test (Table 5) show that the mean perceived satisfaction differs

Figure 3. Quiz Scores by Group and Date

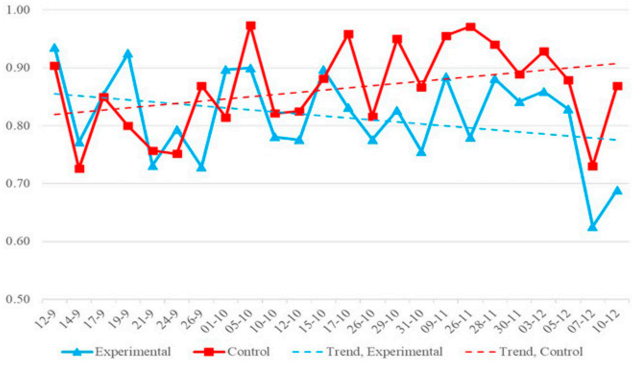


Table 3. Summative Survey Descriptive Statistics

Items		*N* = 23	
		Mean	SD
Perceived overall experience with gamified course design[a]			
GS1	I enjoyed the gamified design of this class.	3.96	0.98
GS2	The gamified design kept me engaged.	3.78	0.95
GS3	I would recommend gamified design in future courses to other instructors.	3.83	0.94
GS4	Being part of a team motivated me to study.	3.96	0.77
GS5	Gamified design in class provided excitement to the course.	3.83	0.89
GS6	Gamified design in class provided me with a challenge.	3.70	0.97
GS7	Gamified design allowed me to compete with others in the class.	3.39	1.08
GS8	Gamified design stimulated my curiosity regarding the course material.	3.78	0.80
GS9	Gamified design in class did not motivate me to study.[b]	3.00	1.00
GS10	Gamified design stimulated me emotionally.	3.43	0.90
GS11	Gameful design helped to stop me from being bored.	3.74	0.92
GS12	I enjoyed being part of a team.	4.09	0.90
GS13	I believe the gameful design fit well with each chapter's text material.	3.87	0.92
GS14	Gamified design allowed me to collaborate with others in the class.	3.96	0.93
GS15	I did not learn anything about the intended topic through gameful design in the class.[b]	3.17	1.15
GS16	I believe gameful design increased my content knowledge of operations and supply chain management.	3.74	0.92
GS17	If I had the choice, I would choose to enroll in courses where gameful design is used.	3.57	0.99
GS18	If I had to vote, I would vote against using gameful design in the operations and supply chain management classroom.[b]	2.74	1.01
GS19	I am enthusiastic about instructors using gameful design in the classroom to teach operations & supply chain management.	3.70	0.82
Perceived efficacy of individual game elements[c]		*M*	*SD*
GE1	Alteration of terms (Farming, Mission, Quest, Raid, etc.)	3.54	1.23
GE2	Resemblance of chapter progression to level progression (clearing a chapter with team Mission report)	3.61	1.03
GE3	Reward system (tickets for power-ups)	3.68	1.06
GE4	Teamwork and team dynamics in tasks and exams	3.71	1.05
GE5	Meaningful choices (report format, number of reports, etc.)	3.43	1.14
GE6	Difficulty progression of exam questions	3.21	1.07
Perception of skills learned via videogames		Percentage selected	
SL1	Collaboration and teamwork	65%	
SL2	Creativity and problem-solving	65%	
SL3	Critical thinking and leading/motivating	52%	
SL4	Analyzing/classifying	52%	
Learning style [reference][d]		Selection	Percentage
LS1	Working against other students	4	17%
LS2	Working with other students	15	65%
LS3	Working alone	8	35%

[a]Five-point Likert scales: 1 = strongly disagree to 5 = strongly agree.
[b]Inversely measured.
[c]Five-point Likert scales: 1 = extremely negatively to 5 = extremely positively.
[d]Double selections counted.

between the experimental (mean = 4.04, SD = 0.79, n = 176) and the control (mean = 4.22, SD = 0.68, n = 167) group at the 0.05 level of significance ($t = -2.30$, $df = 341$, $p = 0.022$, 95% CI for mean difference: −0.340, −0.026). On average, students in the experimental group were less satisfied with the learning method compared with those in the control section. The fixed-effect analysis of variance (ANOVA) results (Table 6) show a significant between-group difference in satisfaction ($F = 5.776$, $p = 0.017$). There was no significant difference associated with survey sessions (SRV, proxy for time, $p = 0.806$) or the interaction of group (Sec) and survey session ($p = 0.504$).

4.2.2. Perceived Course Experience. Results of the two-independent-samples t-test (Table 5) show that the mean perceived course experience differs between the experimental (mean = 3.81, SD = 0.70, n = 176) and the control (mean = 3.96, SD = 0.72, n = 167) group at the 0.05 level of significance ($t = -1.97$, $df = 341$, p = 0.05, 95% CI for mean difference: −0.302, 0.000). On average, students in the experimental group had more negative experiences compared with those in the control section. The fixed-effect ANOVA results (Table 7) show a significant between-group difference in experience ($F = 4.323$, $p = 0.038$). There was no significant difference associated with survey sessions

Table 4. Results of *t*-Tests and Descriptive Statistics of Content Knowledge Measurement Items by Group

	Experimental group (N = 33)		Control group (N = 29)		95% CI for mean difference				
Measures (percentage score)	Mean	SD	Mean	SD	Lower	Upper	t	df	Significance (two-tailed)
Quizzes total	81.86	7.74	86.39	7.67	–0.085	–0.006	–2.31*	60	0.024
Exams total	70.71	6.59	76.41	7.46	–9.269	–2.133	–3.20**	60	0.002
Midterm exam									
Individual	74.34	12.92	75.86	13.62	–0.083	0.052	–0.45	60	0.654
Team	74.24	6.25	81.15	11.31	–0.117	–0.021	–2.92**	42.37	0.006
Total	74.29	6.07	78.51	7.94	–0.078	–0.006	–2.36*	60	0.021
Final exam									
Individual	71.21	11.63	70.69	17.56	–0.070	0.080	0.14	60	0.889
Team	63.03	14.3	77.93	5.26	–0.203	–0.095	–5.57***	41.5	0.000
Total	67.12	9.77	74.31	9.28	–0.120	–0.023	–2.96**	60	0.004

*p < 0.05; **p < 0.01; ***p < 0.001.

(SRV, proxy for time, p = 0.441) or the interaction of group (Sec) and survey session (p = 0.947).

4.2.3. Perceived Learning. Results of the two-independent-samples t-test (Table 5) show that the mean difference of perceived learning between the experimental (mean = 3.96, SD = 0.63, n = 176) and the control (mean = 4.08, SD = 0.60, n = 167) group is not significant at the 0.05 level of significance (t = –1.73, df = 341, p = 0.085, 95% CI for mean difference: –0.246, 0.016). The fixed-effect ANOVA results (Table 8) show a nonsignificant between-group difference in perceived learning (F = 2.421, p = 0.121).

4.2.4. Perceived Impact of Teaching Techniques. Table 9 lists the t-test results for impact of teaching technique measurements. Students in the experimental group perceived the textbook (mean = 5.59, SD = 1.3, n = 176) to aid their learning significantly more than students in the control group (mean = 4.9, SD = 0.1.8, n = 167; t = 4.04, df = 303.8, p = 0.000, 95% CI for mean difference: 0.349, 1.013). Students in the control group perceived the professor (mean = 5.8, SD = 1.3, n = 167) and lecture videos (mean = 4.7, SD = 1.4, n = 167) to aid learning significantly more than in the experimental group (professor: mean = 5.52, SD = 1.3, n = 176; lecture videos: mean = 4.18, SD = 1.6, n = 176). Group differences in other variables were not significant. In the experimental group, students reported highest impact of textbook, followed by professor, quizzes, and classroom discussions. Students in the control group report highest impact of professor, followed by classroom discussions and quizzes.

A significantly higher mean score for impact of textbook supports the course design of using gamification to motivate students to read the textbook before coming to class.

4.3. Summative Perceptions

The summative survey analysis with 19 items on gamification perceptions (α = 0.945) reveals that students in the experimental section had favorable perceptions toward most individual items (Table 3). It is notable that the highest mean score is reported on item GS12, "I enjoyed being a part of a team" (mean = 4.09), followed by two other items regarding social dynamics: items GS4, "Being part of a team motivated me to study" (mean = 3.96) and GS14, "Gamified design allowed me to collaborate with others in the

Table 5. Results of *t*-Test and Descriptive Statistics of Dependent Variables by Group

	Group									
	Experimental			Control						
	Mean	SD	n	Mean	SD	n	95% CI for mean difference	t	df	Significance (two-tailed)
Knowledge	76.28	6.47	33	81.40	6.43	29	–8.40, –1.83	–3.12**	60	0.003
Satisfaction	4.04	0.79	176	4.22	0.68	167	–0.340, –0.026	–2.30*	341	0.022
Experience	3.81	0.70	176	3.96	0.72	167	–0.302, 0.000	–1.97*	341	0.050
Learning	3.96	0.63	176	4.08	0.60	167	–0.246, 0.016	–1.73	341	0.085

*p < 0.05; **p < 0.01.

Table 6. Fixed-Effect ANOVA for Satisfaction (Dependent Variable: *Satis*)

Source	Type III *SS*	*df*	*MS*	*F*	Significance	Partial η^2	Noncent. parameter	Observed power[a]
Corrected model	8.396[b]	15	0.560	1.017	0.437	0.045	15.256	0.665
Intercept	5,646.715	1	5,646.715	10,260.424	0.000	0.969	10,260.424	1.000
Sec	3.178	1	3.178	5.776	0.017	0.017	5.776	0.669
SRV	2.069	7	0.296	0.537	0.806	0.011	3.760	0.233
Sec × *SRV*	3.478	7	0.497	0.903	0.504	0.019	6.319	0.390
Error	179.961	327	0.550					
Total	6,031.938	343						
Corrected total	188.357	342						

[a]Computed using $\alpha = 0.05$.
[b]$R^2 = 0.045$ (adjusted $R^2 = 0.001$).

class" (mean = 3.96), because the design of gamified assessment in the course had a heavy focus on social interactions and teamwork.

Students in the experimental group also reported overall positive perceptions toward individual game elements used in the course (Table 3). Again, students reported highest mean scores on item GE4, "teamwork and team dynamics in tasks and exams" (mean = 3.71), followed by item GE3, "reward system (tickets for power-ups, mean = 3.68), which is also a social gamification element and the core driver for the gamified structure.

4.4. Supplemental Analyses

4.4.1. Academic Standing. To check for a potential bias from student's previous academic standing on content knowledge, additional analyses were conducted to explore the possible effect of gamification on content knowledge, after taking student's cumulative grade point average (GPA) into account. A scatterplot indicated a low degree of positive correlation between student's GPA and content knowledge. However, there was no significant difference between the two group's average cumulative GPA when the semester started. Thus, we can rule out the possibility that the difference in content knowledge and some perceptions between the experimental and control group were driven by previous academic standing.

4.4.2. Sex. Although the basic premise of gamification and SDT is that demographic differences do not matter, we ran another *t*-test for quiz scores, controlling for sex with the two groups combined. Female students ($n = 13$, mean = 84.07) show a significantly higher mean score compared with male students ($n = 49$, mean = 74.20; $t = 2.543$, $df = 60$, $p = 0.014$). In exam scores, female and male students show no significantly different scores.

A bootstrap *t*-test on mean quiz scores of female students shows no significant difference between the experimental ($n = 6$, mean = 83.84) and control ($n = 7$, mean = 89.56) group ($SE = 4.221$, $p = 0.206$, 95% *CI* for mean difference: −15.016, 3.566). A bootstrap *t*-test on mean quiz scores of male students of the experimental ($n = 27$, mean = 81.42) and control ($n = 22$, mean = 85.38) group resulted in no significant difference ($SE = 2.117$, $p = 0.070$, 95% CI for mean difference: −7.994, 0.313). These analyses allow us to suggest that assessment gamification had no interaction with sex.

5. Discussion and Self-Reflection

The results reported in Section 4 were surprising to us as we expected gamification of the OSCM course assessment element to have a positive impact on outcome variables. Although there have been a few critiques of gamification in the literature, they mostly focused on the use PBL as a stock approach and its negative effects on students' intrinsic motivation

Table 7. Fixed-Effect ANOVA for Experience (Dependent Variable: *Exper*)

Source	Type III *SS*	*df*	*MS*	*F*	Significance	Partial η^2	Noncent. parameter	Observed power[a]
Corrected model	6.626[b]	15	0.442	0.862	0.608	0.038	12.927	0.572
Intercept	4,979.050	1	4,979.050	9,714.576	0.000	0.967	9,714.576	1.000
Sec	2.216	1	2.216	4.323	0.038	0.013	4.323	0.545
SRV	3.539	7	0.506	0.986	0.441	0.021	6.905	0.426
Sec × *SRV*	1.132	7	0.162	0.316	0.947	0.007	2.210	0.146
Error	167.599	327	0.513					
Total	5,339.125	343						
Corrected total	174.224	342						

[a]Computed using $\alpha = 0.05$.
[b]$R^2 = 0.045$ (adjusted $R^2 = -0.006$).

Table 8. Fixed-Effect ANOVA for Perceived Learning (Dependent Variable: *Learn*)

Source	Type III *SS*	*df*	*MS*	*F*	Significance	Partial η^2	Noncent. parameter	Observed power[a]
Corrected model	4.732[b]	15	0.315	0.816	0.659	0.036	12.244	0.543
Intercept	5,375.284	1	5,375.284	13,908.026	0.000	0.977	13,908.026	1.000
Sec	0.936	1	0.936	2.421	0.121	0.007	2.421	0.342
SRV	1.722	7	0.246	0.636	0.726	0.013	4.455	0.274
Sec × *SRV*	1.875	7	0.268	0.693	0.678	0.015	4.850	0.299
Error	126.382	327	0.386					
Total	5,663.776	343						
Corrected total	131.114	342						

[a]Computed using $\alpha = 0.05$.
[b]$R^2 = 0.045$ (adjusted $R^2 = -0.008$).

(Deci et al. 1999, Aldemir et al. 2018). Considering the advice of these critical studies, we avoided the classical PBL approach, gamified only one course element, and used no online platform to design a generalizable course gamification experience. The surprising findings of our study make significant contributions to the literature by demonstrating a dark side of gamification in a classroom setting and by exploring the effects of selectively gamifying a course design element (assessment) with no online platform involved. We will interpret our findings and reflect on our experiences to assist other professors design their own course gamification projects.

First, we would like to share a few positive observations of gamification that are more aligned with the literature. Students in the experimental group were vocal about their excitement and motivation toward the course. In more than one instance, a group of students in the experimental group would approach the instructor and express their excitement for the course, contrasting their frustration with another course that was not gamified. In an anonymous written survey, administered in the ninth week, students responded with some positive comments such as "I really enjoy the videogame teaching style," "I like the videogame format," "the farming forces you to study almost daily," "I like that I have to rely on myself to teach myself," "I believe group work is very effective," and "I like how you offer the various ways we can do the mission reports because it helps me learn."

In addition, an analysis of perceived impact of teaching technique reveals that students in the gamified experimental group perceived the textbook as the most impactful resource, whereas those in the control group perceived the professor to be the most impactful element in their learning (Table 9). Student response to the instructor's proprietary midterm and final survey question on student perception of the textbook was overwhelmingly positive in the gamified group, whereas several in the control group expressed indifference or negative perceptions. This suggests that our iteration of assessment gamification motivated students to read the textbook before class to be ready and resulted in them being more self-regulating, which is associated with higher intrinsic motivation in SDT (Deci and Ryan 2000). Controlling for the effect of the flipped classroom design, which

Table 9. *t*-Test Results and Descriptive Statistics of Impact of Teaching Technique Variables

	Group							
	Experimental (*N* = 176)		Control (*N* = 167)					
Variables	Mean	SD	Mean	SD	95% CI for mean difference	*t*	*df*	Significance (two-tailed)
Classroom Discussions	5.39	1.1	5.5	1.2	−0.358, 0.124	−0.955	341	0.34
Team Chapter Reports	5.09	1.3	5	1.3	−0.196, 0.354	0.565	341	0.573
Quizzes	5.43	1.1	5.4	1.3	−0.192, 0.313	0.474	341	0.636
Professor	5.52	1.3	5.8	1.3	−0.555, −0.016	−2.08*	341	0.038
Exams	4.59	1.5	4.8	1.5	−0.521, 0.098	−1.34	341	0.18
Lecture Videos	4.18	1.6	4.7	1.4	−0.874, −0.223	−3.32***	341	0.001
Textbook	5.59	1.3	4.9	1.8	0.349, 1.013	4.04***	303.8	0.000

$^*p < 0.05$; $^{***}p < 0.001$.

was identical across sections, gamification of assessment activities seems to have had positive effect on making students read the textbook.

Our statistical results, interestingly, suggest that gamification of assessment significantly decreases student's content knowledge. The declining linear trend in the quiz scores from the experimental group is consistent with the observations made by Koivisto and Hamari (2014), where perceived usefulness, enjoyment, and playfulness are found to diminish with time interacting with gamified system. It is possible that gamifying assessment had some novelty value in the beginning of the semester resulting in higher motivation and engagement, but as students became accustomed to the course design and perceived the tasks as equally rudimentary as those in any other course, the novelty value may have faded and even harmed student performance in quizzes and exams. Lieberoth (2015) finds that adding a playful frame to tasks (*shallow gamification*) takes away the grit and output orientation of more goal-oriented work, which may further explain the declining trend in the content knowledge of the experimental group.

Another reflection comes from the reward system. We used common rolled-up drink tickets as reward/currency in the gamified system. Although students appreciated being awarded tickets for their quiz and team assignment performances, the items themselves held little sentimental value. Regarding the effect of rewards on intrinsic motivation, the literature provides controversial viewpoints. Deci and Ryan (2000), based on findings of a meta-analytic review of studies on the effects of extrinsic rewards on intrinsic motivation, argue that tangible rewards that are expected and task-contingent have negative impact on intrinsic motivation (Deci et al. 1999). In fact, Deci et al. (1999) find that all contingent tangible rewards, including monetary, significantly undermined intrinsic motivation, categorically refuting the argument of Eisenberger and Cameron (1996) that the detrimental effects of rewards is mostly a myth (Deci et al. 1999). However, these undermining effects of intrinsic motivation are only present in interesting tasks and not dull or boring ones (Deci et al. 1999). On the other hand, Aldemir et al. (2018), while extending the findings of Deci and Ryan (2000) with application on higher education, stress the need for a continuous and systematical reward distribution, with tangible items that are inexpensive but hold high sentimental value to avoid students suddenly losing their motivation to continue. Enhanced aesthetic and sentimental value of the reward objects may enhance the prolonged engagement with the gamified system. The value of the reward was also found to be a major modulating factor in enhancing episodic memory (Mason et al. 2017), making the association of high-value reward with an immersive experience an effective mechanism for gamification (Mullins and Sabherwal 2018).

A clearer disparity regarding the reward system exists between the unit of assessment and the unit of rewards. Although the quizzes were given as individual assessments, the tickets were awarded for the team performance. Typically, teams assigned one member to collect and keep the awarded tickets (typically in a zip-lock plastic bag) while the rest of the team had no access to them. It appears that setting the team as the unit for rewards for individual assessments may have hindered the *slicing up* of motivating effects to the individual level. We suggest that rewards should, first and foremost, benefit the individual before the team in order to retain prolonged motivation and engagement of all team members. Considering the focus of SDT on the relationship between competence and intrinsic motivation (Deci and Ryan 2000), proper reward, adequately valuable to the perception of students, should be given to individuals in addition to the teams.

Another meaningful observation comes from the individual and team score differences between the two groups. A closer look at the content knowledge measurements reveals that the gamified group scored significantly lower than the control group in quizzes and team exams but not in individual exam scores. These results suggest that gamification adversely affected student's team performance, although it did not affect their individual performance in the summative knowledge assessments. Hanus and Fox (2015) find that encouraging competition and social comparison harms intrinsic motivation, which in turn results in lower exam scores. Although their study environment uses an online leaderboard and badges, resulting in some degree of difference from the present study, the settings in present study may still invoke similar social comparison and competition. Specifically, after completing the quiz feedback, the instructor collected the quiz sheets by team and immediately awarded tickets to teams that qualified. Also, the instructor periodically asked teams to report the current quantity of tickets accumulated by each team.

The insights gained from the quizzes and exams show that the present gamification design involving the assessment of the course is, at best, insufficient to deliver a positive impact on student's content knowledge. This may be because of the long duration of the gamified experience, the simplistic reward system that rewards the team for individual work, and latent social competition that led to lower intrinsic motivation. It is apparent that the design needs to be improved to encompass more elements than assessment in order to retain student engagement and motivation, which will lead to enhanced content knowledge. Also, In the university's official course evaluation survey, one student from the experimental section commented: "Harder than most courses because students were expected to teach themselves. I wish the professor taught before we were quizzed." This suggests that the course's overall structure and the sequence of activities may interact with gamification.

5.1. Student Perceptions

The results of the formative perceptions survey suggest that assessment gamification had a significantly adverse effect on student's perception of satisfaction and course experience. However, perceived learning did not show a significant between-group difference. This suggests that, although our version of assessment gamification may have had a negative effect on student's satisfaction and course experience, they still felt like they were learning from the course. Also, it is important to note that, despite the significantly negative results in satisfaction and course experience in comparison with the control group, the survey results show that the gamified group is still highly satisfied (mean = 4.04/5) and having an overall positive course experience (mean = 3.81/5; Table 5).

These results and observations, combined together, suggest that students in the experimental section, as the initial hype and novelty value faded, perceived the gamified activities for coursework as what they actually are: coursework. Thus, they may have perceived slightly less satisfaction and poorer course experience overall because of a certain degree of disillusionment and the negative effects of the reward system, but it did not result in a detrimental decline in satisfaction or experience. Also, they developed a sense of self-regulation and autonomy in reading the textbook before class. The cooperative social emphasis of gamification that builds every assessment activity up toward the exams also created a sense of accountability and team membership, apparent in the high mean scores in social aspect items of both formative perceptions survey (Table 2: IT1, IT2) and summative gamified course survey (Table 3: GS4, GS12, GS14, GE3, GE4), although they did not manifest as higher team exam scores. These results are consistent with the emphasis on the social aspect of gamification in the literature (Hamari and Koivisto 2015; de-Marcos et al. 2016, 2017).

5.2. Gamification of Assessment

In the establishing work for SDT, Deci and Ryan (2000) refer to studies of Heider (1958) and de Charms (1968) on internal perceived locus of causality (PLOC), as well as additional studies that showed intrinsic motivation is undermined by threats, surveillance, evaluation, and deadlines because they shift the internal PLOC (I-PLOC) to external PLOC (E-PLOC). These studies are grounds on which Deci and Ryan (2000) argue the importance of satisfying the need for autonomy and sense of self-initiation for increased intrinsic motivation that can lead to greater creativity and better problem-solving.

In the present study, although choices were given for some components, such as team report format, the quizzes and exams were mandatory, and student performance in these mandatory activities had direct and heavy consequences in their grades. This is an issue especially prominent in the context of higher education because it is nearly impossible to completely eliminate certain factors that may control or manipulate students in the current design for undergraduate-level coursework. In order to achieve the desired level of choices and autonomy, perhaps assessment should not be the only element of the course to be gamified.

6. Conclusions and Future Work

We are surprised to find that gamification of assessment activities of a course has a negative impact on content knowledge and student perceptions. Our design was intended to explore the benefits that gamification can deliver to students when applied to course assessment. This design primarily answers the growing call for empirical studies on specific application contexts, as well as for isolating individual design elements. We did observe a heightened initial enthusiasm from the students for the gamified design, as well as the effect of gamification on promoting textbook reading before class. However, those observations did not lead to significant improvements in content knowledge and student perceptions and even resulted in significantly inferior results. Our negative results suggest that gamification can hurt content knowledge and student perceptions when applied only to assessment of coursework. Instead of simplistic game mechanics applied to narrow areas of the coursework, the gamified system must encompass a multitude of mechanics and elements that can entertain various needs of students. This principle should apply across contexts: education, marketing, training, management, etc. This does not mean that gamification must be applied to all aspects of coursework to be successful. Finding the optimal mix of gamification in the coursework would be a valuable future study.

As gamification research matures, there has been a subtle but growing call for elements of games and videogames that were not emphasized in previous iterations of gamification in either business or academia. These are calls for elements that contribute to the ever-elusive aspect of fun; elements on which videogame developers and game makers spend enormous time and resources to develop but somehow got lost in the transplantation of games to corporate and educational contexts: narrative and play style are among these elements. Narrative not only creates a more coherent experience but also facilitates the need to be part of something bigger than oneself, as addressed by McGonigal (2011) in the opening quote of this study. An interdisciplinary team spanning from business, computer science, literature, theater, and so on, may work best to design, implement, and execute a gamified coursework with rich, epic narratives and carefully placed elements and mechanics to deliver a fun, engaging, and effective learning experience.

This study extends the growing literature on gamification in higher education by implementing gamification to a narrower area of learning experience, focused on the single course element, assessment, and used no online platform. Our analysis and observation lead to a conclusion that gamification is at best insignificant when applied only to assessment and at worst may significantly harm student perceptions and academic performance.

6.1. Limitations

This study is limited in several ways. First, because of an iterative process of administrative approval for human subject research, the formative surveys were delayed until the 10th week of the 16-week semester. As a result, although assessments in content knowledge span the entire semester, the data from the formative perceptions surveys are limited to the later third of the semester. Also, because of confidentiality and anonymity requirements, the surveys contained no identifiers of individual students. Therefore, we were unable to treat the survey as a repeated measure to explore the effect of assessment gamification on individual student's perceptions over time. Although we do acknowledge this to be a major limitation, we opted for anonymity to avoid the possibility of students giving positively biased answers in the surveys because of coercion.

The quiz/exam scores and survey results of this study provide an interesting interpretation that lead to more questions. The controversial effect of rewards in a gamified system and effect of different types and units of rewards need more empirical testing as an isolated element. The effect of social interactions and the optimal mix of cooperation and competition, regarding their effects on intrinsic motivation, need to be addressed specifically in higher education context.

References

Aldemir T, Celik B, Kaplan G (2018) A qualitative investigation of student perceptions of game elements in a gamified course. *Comput. Human Behav.* 78:235–254.

Asef-Vaziri A (2015) The flipped classroom of operations management: A not-for-cost-reduction platform. *Decision Sci. J. Innovative Ed.* 13(1).

Buckley P, Doyle E (2014) Gamification and student motivation. *Interaction Learning Environ.* 24(6):1162–1175.

Cantwell ER, Sousou J, Jadotte YT, Pierce J, Akioyamen LE (2017) PROTOCOL: Collaborative testing for improving student learning outcomes and test-taking performance in higher education: A systematic review. *Campbell Systematic Rev*. 13:1–18.

Dahlström Ö (2012) Learning during a collaborative final exam. *Ed. Res. Evaluation* 18(4):321–332.

de-Marcos L, Garcia-Lopez E, Garcia-Cabot A (2016) On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Comput. Ed.* 95:99–113.

de-Marcos L, García-Cabot A, García-López E (2017) Toward the social gamification of e-learning: A practical experiment. *Internat. J. Engrg. Ed.* 33:66–73.

de Charms R (1968) *Personal Causation* (Academic, New York).

Deci EL, Ryan RM (2000) The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psych. Inquiry* 11(4):227–268.

Deci EL, Koestner R, Ryan RM (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psych. Bull.* 125(6):627–668.

Dias J (2017) Teaching operations research to undergraduate management students: The role of gamification. *Internat. J. Management Ed.* 15(1):98–111.

Eisenberger R, Cameron J (1996) Detrimental effects of reward: Reality or myth? *Amer. Psych.* 51(11):1153–1166.

Fitz-Walter Z, Johnson D, Wyeth P, Tjondronegoro D, Scott-Parker B (2017) Driven to drive? Investigating the effect of gamification on learner driver behavior, perceived motivation and user experience. *Comput. Human Behav.* 71:586–595.

González CS, Gómez N, Navarro V, Cairós M, Quirce C, Toledo P, Marrero-Gordillo N (2016) Learning healthy lifestyles through active videogames, motor games and the gamification of educational activities. *Comput. Human Behav.* 55:529–551.

Hamari J, Koivisto J (2015) "Working out for likes": An empirical study on social influence in exercise gamification. *Comput. Human Behav.* 50:333–347.

Hanus MD, Fox J (2015) Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Comput. Ed.* 80:152–161.

Heider F (1958) *The Psychology of Interpersonal Relations* (Wiley, New York).

Herreid CF, Schiller NA (2013) Case studies and the flipped classroom. *J. College Sci. Teaching* 42(5):62–66.

Hew KF, Huang B, Chu KWS, Chiu DKW (2016) Engaging asian students through game mechanics: Findings from two experiment studies. *Comput. Ed.* 92-93:221–236.

Hudiburg ML (2016). Motivation and learning in an online collaborative project using gamification. PhD dissertation, Old Dominion University, Norfolk, Virginia.

Huotari K, Hamari J (2012) Defining gamification—A service marketing perspective. *Proc. 16th Internat. Academic MindTrek Conf.* (ACM, New York), 17–22.

Jassawalla A, Sashittal H, Malshe A (2009) Students' perceptions of social loafing: Its antecedents and consequences in undergraduate business classroom teams. *Acad. Management Learning Ed.* 8(1):42–54.

Johanns B, Dinkens A, Moore J (2017) A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills. *Nurse Ed. Practice* 27:89–94.

Kapitanoff SH (2009) Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning Higher Ed.* 10(1):56–70.

Kapp KM, Blair L, Mesch R (2014) *The Gamification of Learning and Instruction Fieldbook* (John Wiley & Sons, San Francisco).

Khansari PS, Coyne L (2018) An innovative addition to team-based-learning pedagogy to enhance teaching and learning: Students' perceptions of team exams. *Currents Pharmacy Teaching Learn.* 10(1):90–95.

Koivisto J, Hamari J (2014) Demographic differences in perceived benefits from gamification. *Comput. Human Behav.* 35:179–188.

Kong N (2019) Active game-based learning of dynamics modeling and simulation in biomedical systems engineering. *INFORMS Trans. Ed*. 20(1):16–25.

Koppenhaver GD, Shrader CB (2003) Structuring the classroom for performance: Cooperative learning with instructor-assigned teams. *Decisions Sci. J. Innovative Ed.* 1(1):1–21.

Kuo M-S, Chuang T-Y (2016) How gamification motivates visits and engagement for online academic dissemination—An empirical study. *Comput. Human Behav.* 55:16–27.

Levine RE, Borges NJ, Roman BJB, Carchedi LR, Townsend MH, Cluver JS, Thompson BM (2017) High-stakes collaborative testing: Why not? *Teaching Learn. Med.* 30(2):133–140.

Lieberoth A (2015) Shallow gamification: Testing psychological effects of framing an activity as a game. *Games Culture* 10(3):229–248.

Mason A, Farrell S, Howard-Jones P, Ludwig CJH (2017) The role of reward and reward uncertainty in episodic memory. *J. Memory Language* 96:62–77.

McGonigal J (2011) *Reality Is Broken: Why Games Make Us Better and How They Can Change the World* (Penguin Group, New York).

Morillas Barrio C, Munoz-Organero M, Sanchez Soriano J (2016) Can gamification improve the benefits of student response systems in learning? An experimental study. *IEEE Trans. Emerging Top Comput.* 4(3):429–438.

Mullins JK, Sabherwal R (2018) Gamification: A cognitive-emotional view. *J. Bus. Res.* 106:304–314.

Nacke LE, Deterding S (2017) The maturing of gamification research. *Comput. Human Behav.* 71:450–454.

P&S Market Research (2016) Gamification market—Market intelligence report 2022. Accessed October 10, 2019, https://www.psmarketresearch.com/market-analysis/gamification-market.

Reinig BA, Horowitz I, Whittenburg GE (2011) The effect of team-based learning on student attitudes and satisfaction. *Decisions Sci. J. Innovative Ed.* 9(1):27–47.

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Amer. Psych.* 55(1): 68–78.

Saunders ME (2017). Gamification in employee training and development: Millennial perspectives. PhD dissertation, University of the Rockies, Denver, CO.

Seaborn K, Fels DI (2015) Gamification in theory and action: A survey. *Internat. J. Human Comput. Stud.* 74:14–31.

Sobocinski M (2017) I gamified my courses and I hate that.... *World J. Sci. Tech. Sustainable Development* 14(2/3):135–142.

Song D (2017) Engaged cohorts: Can gamification engage all college students in class? *Eurasia J. Math. Sci. Tech. Ed.* 13(7):3723–2734.

Stansbury JA, Earnest DR (2016) Meaningful gamification in an industrial/organizational psychology course. *Teaching Psych.* 44(1):38–45.

Subhash S, Cudney EA (2018) Gamified learning in higher education: A systematic review of the literature. *Comput. Human Behav.* 87:192–206.

van Roy R, Zaman B (2018) Unravelling the ambivalent motivational power of gamification: A basic psychological needs perspective. *Int. J. Human-Comput. Stud.* 127:38–50.

Villagrasa S, Fonseca D, Redondo E, Duran J (2014) Teaching case of gamification and visual technologies for education. *J. Cases Inform. Tech.* 16(4):38–57.

Zichermann G, Cunningham C (2011) *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps* (O'Reilly Media, Sebastopol, CA).