MAXIMIZING STUDENT LEARNING OUTCOMES WHEN USING BUSINESS STRATEGY SIMULATION TOOLS: THE ROLE OF PEDAGOGY, PARTICIPATION, AND PRACTICE (A MULTI-YEAR STUDY)

Kevin A. Fletcher

Kevin A. Fletcher, Assistant Professor, Russell Sage College, Email:
fletck3@sage.edu

Abstract

The use of simulations is prevalent across business schools in colleges and universities. Research has shown that they can be useful in bridging theory with practice and preparing students for “real world” employment. However, there is limited research on determining the best pedagogical approaches for using these tools in a classroom setting. In this study, two approaches to using the Capsim “Capstone” online business strategy software are explored and measured for overall effectiveness and meeting learning outcomes. Results show that the pedagogical approach matters less in determining learning success as compared to other variables including average team logins, individual login behavior, experience of the instructor in using the simulation, and the

simulation industry environment itself (i.e., number of student-run teams versus number of computer-run teams). The results have implications for the delivery of computer simulations in class as well as their use as part of a course grade.

Key Words: Business Strategy, Simulations, Pedagogy.

DOI: <https://dx.doi.org/10.15239/j.brcacadjb.2021.11.01.ja07>

INTRODUCTION

College and university management faculty are looking to produce graduates that are ready for “real world” management decision-making. One way that educators have been providing such “real world” experience in the classroom is through the use of business simulations. Specifically, online, or computer-based simulations allow for a dynamism and flexibility that benefits students and enhances the educational experience (Bell & Loon, 2015; Gove, 2012; Wolfe, 2016). As Brawer (1997) notes, “computer-based simulations provide environments wherein participants experience realities of the business world that are risk-free... simulations also offer sufficient insight into the actual operation of a business so that participants can later transfer the simulation model strategies into real-life situations.” This has been supported in more recent research as well (Lohmann et. al., 2019; Ledman, 2016).

Similar research has shown the value of simulations to students in making complex, functionally integrated decisions, specific to strategic management courses, in simulated real-life, competitive environments (Poisson-de-Haro & Turgut, 2012; Tiwari, Nafees & Krishnan, 2014). Likewise, course material has shown to be comprehended at a deeper, more cognizant level (Tan, Tse & Chung, 2010). These tools offer something traditional lecture and classroom teaching cannot offer. Yet, the effectiveness of business simulations to meet program learning objectives in both undergraduate and graduate programs has been questioned (Di Meglio, 2008; Umble, Umble & Artz, 2008).

The use of these types of classroom resources is pervasive. According to Ahn (2008), about 95% of The Association to Advance College Schools of Business (AACSB) accredited schools are using some sort of business simulation to further learning outcome goals. The potential for enhanced learning has been noted by researchers including increased retention of lecture materials as measured through exam results (Gremmen & Potters, 1997) and mastery of concepts (Klassen & Willoughby, 2003). Additional learning potential comes through navigating team dynamics, interpersonal communication, and decision-making with colleagues. This matches real-world employment demands. For instance, a Pew Center research study on student job preparation and employer needs identified the skills, capabilities, and attributes needed for success. These included emotional intelligence, curiosity, creativity, adaptability, resilience, and critical thinking. Also identified was the need for more practical experiential learning opportunities in higher education (Rainie & Anderson, 2017).

However, the experience with online classroom simulations is not all positive. Some research has shown the use of simulations in business coursework to be problematic in terms of delivering unsuccessful simulation results as a result of unproductive team dynamics (Anderson, 2005). Egereonu (2011) also found that students experienced stress in using simulations in class due to lack of time for preparation and implementation in a typical college classroom session. An additional issue identified by Neely & Tucker (2013) was the complicated nature of online simulations leading to ineffective use in classroom settings. Namely, significant time is spent learning how to use the simulation at the detriment of learning content based on the simulation. Others, such as Frezo, Behrens & Mislevy (2010), have highlighted the challenges instructors face in using these types of simulations effectively. Ledman (2016) reflects on the experience of using such simulations in business capstone classes, offering recommendations for addressing the complexity and challenge. These include patience in introducing the simulation to the students, opportunities to practice using it so the “mechanics” of the software is

understood and regular “check-ups” in reviewing decisions and results with the students. With these ongoing issues in mind, this research explores the use of a specific simulation, the Capsim Capstone, with two different pedagogical approaches to help further understand ways to use online simulations more effectively.

OVERVIEW OF CAPSIM CAPSTONE SIMULATION

The Capsim Capstone online business simulation tool (www.capsim.com) is a strategic management teaching aid created and delivered by Capsim Management Simulations, Inc. It is a popular tool in business strategy classrooms. Saulnier (2009) reports that over 600 colleges and universities around the world use the Capsim business simulation, and in 2008 alone, over 82,000 undergraduate and Masters of Business Administration (MBA) students were registered as Capsim users. As a teaching tool, Capsim provides competition among student teams while focusing on how sets of organizational decisions and their results can interact and are impacted by the decisions of other companies in an industry.

Through the Capsim simulation, individual students or student teams run a fictitious company (that makes various types of sensors) in an industry environment that includes five other student or computer-run companies. Decisions are made for each round (representing the fiscal year for the entire industry) in research and development (R&D), marketing, production, and finance, with additional decisions made in later rounds, at the instructor’s discretion, in total quality management (TQM), advanced marketing, human resources (HR), and labor negotiations. The five other teams competing in that industry also make decisions in each of these functional areas. Once decisions are made a round is “processed” representing a full year of all six of the company’s activity and results (i.e., producing product, making sales). In essence, numerical decisions from all companies (i.e., number of units of sensors to build in the coming year, costs for production, mix of debt or equity required to fund the decisions in the coming year,) are stored in an integrated spreadsheet which the

simulation uses to produce aggregate industry (and team specific) results representing a year of activity. At the end of each round, student teams receive detailed reports (the primary report is known as the “Capsim Courier”) on how well their products sold, how competitors performed, detailed profit and loss statements and balance sheets, etc. Likewise, a cumulative performance score, the Analyst Report score, provides a good summary of round by round company performance (*Capsim Team member Guide, 2018*).

To set up the scenario for the students, this fictitious industry provides a wide variety of sensors to a growing marketplace. The simulation assumes that a monopoly sensor company has just been broken up into six equal companies—each with the exact same cash, debt, inventory, market share, etc. Each company has one product in each of five distinct market segments—low, traditional, size, performance, high. Each market segment is defined by distinct customer wants and needs around product characteristics that include product size, speed, mean time before failure (MTBF), price, and perceived age (*Capsim Professor Guide, 2018*). (See Appendix #1 for a more complete description of the round-by-round decisions made in this simulation.)

For all decisions, the simulation provides round-by-round, detailed financial and market segment reports for students to analyze. Student teams have information on what customers want (i.e., sensor size, speed, price) for each market segment. They also have the ability to secure loans and issue stock to raise capital needed to fund their activities. Proformas within the simulation also provide additional predictive tools for higher quality decision-making (*Capsim Student Guide, 2018*). The biggest unknown variable in the simulation is how each of the other five companies (whether computer or student-run) will proceed (i.e., their strategic approach and tactical decisions).

FOCUS OF RESEARCH

As stated previously, there is some debate on the efficacy of these simulations as a result of usage and execution decisions by the course instructor (Cadotte, Riley, Bonney & MacGuire, 2013; Wolfe, 2016). For instance, Forsyth & Anastasia (2016) showed that the style of operationalizing the Capsim simulation can negatively affect the learning ability of the students as reflected in their company results. The authors identify several problems that give rise to the need to explore alternative pedagogical solutions in classroom delivery. These issues included the need for students to have time and to take seriously the practice tutorials and the need for students to have time to practice the simulation.

Capsim's Capstone business simulation offers two ways to present the game in classroom setting. A "Tournament" style allows individuals or teams to be placed in an industry to compete against each other. The rounds (i.e., years) are processed automatically (i.e., scheduled using a calendar function in the simulation) or manually by the instructor. The "Footrace" option places each individual or team against five computer-run companies and allows for students to process their own rounds and redo rounds (*Capsim Professor Guide, 2018*). This provides a flexible environment for students to experiment with the simulation, learn how it works, see how different decisions affect results, etc.

Perhaps the most interesting result from Forsyth & Anastasia's (2016) work relates to the use of the "Footrace" option with the Capsim simulation, as they state: "internally, there was evidence that using the Footraces early on in the course, contributes to better scores overall because there seems to be a strong correlation between more practice and better results." The notion that "practice makes perfect," (or in this case, better) for overall Capsim performance, seems to apply in this case. To re-examine this evidence within this research study, the following hypothesis is presented:

H1: The use of Footraces by individual students for practice rounds in Capsim leads to better results in team-based Tournament competition rounds (as measured by team cumulative Analyst Report scores).

Another question raised in the use of this particular simulation by Kilburn & Kilburn (2012) is whether or not individual login activity by teams (as measured by average logins per team member) or by individuals on the teams serve as a predictor of simulation success. A “login” is simply a measure of when a student opens a web-based simulation with their username and password. Their research showed that login rates by individuals in each team might be a better predictor of performance as compared to group-based measures. Likewise, the authors contend that, as one might imagine, certain students can create a competitive advantage for teams based on their interest and activity in the simulation. These “overachievers” effectively give their team a competitive advantage. To test these results within this research study, the following two hypothesis are presented:

H2: Higher average logins per team is a predictor of team-based Tournament competition round success (as measured by team cumulative Analyst Report scores).

H3: Individual login volume by overachievers on each team is a predictor of team-based Tournament competition round success (as measured by team cumulative Analyst Report scores).

Additionally, it might be expected that the Instructor’s ability to teach the simulation (i.e., how to use it, how to read and use the various reports it generates) should improve over time leading to improved Analyst Report scores with each subsequent class. In this study, the eleven classes take place over a four-year timeframe. As such, the following hypothesis is considered:

H4: The average Analyst Report Scores, by class, increases as the course is taught with increasing frequency by the same instructor.

Finally, the difference of industry competitiveness is likely impacted by not only the quality of the student teams for each class, but also by the quantitative mix of student-run and computer-run teams. Since class sizes vary semester-to-semester, the classes used in this study vary with three, four, and five student-run companies in the Tournament competition rounds. Subsequently, since the industry has six companies competing, the proportion of computer-run teams increases (one, two and three, respectively) as the number of student teams drop.

The computer-run companies have set and pre-determined business strategies (i.e., focused differentiation, broad low-cost) that they are programmed to follow. The decisions made by computer-run companies in the industry impact the ability for other teams to be successful. Success, in this case, is measured through the Analyst Report which is highly dependent on sales and profits. At the same time, student-run companies in the simulation make mistakes—under forecasting demand for their sensors and thus losing sales, over forecasting sales leading to high inventory and related inventory costs, not funding their operations adequately in advance through debt and equity and being forced to take on high interest emergency loans, etc. Computer-run companies do not make those mistakes. Thus, as expected, the better competitors, making good decision and leading to higher Analyst Report scores, are the computer-run companies. A higher number of computer-run companies in the industry leads to a more competitive industry environment. Since the computer-run companies are likely better competitors than student-run companies, the following is considered:

H5: The individual team Analyst Report scores for classes with Five Student Team members are higher than those Four Student Teams.

H6: The individual team Analyst Report scores for classes with Four Student Team members are higher than those Three Student Teams.

H7: The individual team Analyst Report scores for classes with Five Student Team members are higher than those Three Student Teams.

RESEARCH METHODS

Over a four year span (2013-2017), consisting of eight different fifteen-week, semester-long classes taught by the same instructor, the simulation was used as follows:

- Each student registered for the simulation (i.e., have an online account) and each student went through the set of online tutorials to learn how to use the simulation. These tutorials were completed in class allowing for the instructor to answer questions throughout the process.
- Students were then placed into three to six person teams by the instructor to run both a practice phase (usually lasting four rounds, or four fiscal years, of decisions) and then a full competition phase (lasting eight rounds of decisions).
- Student teams were created in the same way for every class (i.e., both pre- and post-pedagogical change). Namely, the instructor attempted to create balance by using a “snake draft” approach—ranking the students from strongest to weakest (based on previous course performance, consultation with department colleagues, and in-class observations) and then assigning students to teams based on that ranking. So, if there were four student teams (A, B, C, D), the numerically ranked student team assignments were as follows: A(1), B(2), C(3), D(4), D(5), C(6), B(7), A(8), etc. This approach, as opposed to letting students pick their teams, provides more control over student performance variables, creates the best chance for

fairness and evenness of teams' capabilities, and is a more realistic comparison to work situations (Adams, 2003).

- Each student-run company was automatically scored using the aforementioned rubric that is embedded in the simulation, known as the "Analyst Report." This scoring rubric evaluates performance in ten areas: Margin, Profits, Emergency Loans, Working Capital, Market Share, Forecasting, Customer Satisfaction, Productivity, Financial Structure, and Wealth Creation (Capsim Professor Guide, 2018).
- This score accounted for 20% of their final course grade—a typical value for a final grade found in business capstone classes using simulations (Karriker & Aaron, 2014).

The instructor set up the simulation in this "Tournament" style (i.e., team-based, competing against all the other student teams in the class and 1-3 computer-run teams), had access to each teams' scores, and could archive those past games' results. The simulation was also integrated into the course (not as a stand-alone experience outside of the classroom) which Snow, Gehlen, & Green (2002) found to be a more useful and effective for student learning. Specifically, teams met during class time, the instructor reviewed decisions and results with both the full class and individual teams (i.e., in a consultant role), and strategic management theory was referred to regularly during consultation sessions.

In order to explore the issues, and proposed solutions, raised by Forsyth & Anastasia (2016) and Kilburn & Kilburn (2012), the classroom delivery of the computer simulation to both undergraduate and MBA classes was changed in the following ways in Spring, 2018:

- Each student registered for Capsim and took the online tutorials as before.
- However, each student was assigned with running his/her own company against five computer companies for four practice rounds in the simulation's "Footrace" mode.
- Additionally, at the completion of rounds 2 and 4 in Footrace mode, the students were required to post their company results (i.e.,

Capsim Courier) on a course webpage allowing students and the instructor to post documents. Each student was grouped with 2-3 other students in an online discussion group, each of whom posted their results, and analyzed and commented on their peers' results. This activity was graded to help inspire engagement and effort.

- At the end of the fourth practice round, students were placed into teams and the "Tournament" competition rounds (consisting of 3-5 student teams) was launched over eight rounds.

These changes were applied to three different fifteen-week semester-long courses in 2018 (see Table 1 for a summary of each of the pre and post-change courses). Once again, each course had different students and 20% of the course grade was determined based on the teams' Analyst Report scores. Since Analyst Report is a "balanced scorecard" approach to scoring each rounds' result and cumulative results, it offers a rich and comprehensive way to assess performance by student company teams specific to decision-making (technical and procedural) and problem-solving learning outcomes.

It is worth noting that some researchers have questioned the ability to assess business skills through these types of simulations, specifically pointing to the lack of evaluative functionality around interpersonal, communication, leadership, and other soft skills (Frezzo, Behrens, & Mislevy, 2010; Neely & Tucker, 2012). However, for the purposes of this research, the focus was purely on analytical skills, decision-making, and knowledge and application of the wide range of content needed to succeed in the simulation. As a result, the Analyst Report serves as a functional assessment tool. For instance, the "margins" category is assessed by looking at an equal weight of Contribution Margin, Net Margin, and Return on Sales results. Likewise, each of the ten categories in the Analyst Report are scored through a combination of ratios and results, and students have full access to the scoring criteria and how points are allocated (Capsim Professor Guide, 2018). There is nothing "hidden" in this scoring system, and as a result, students possess all

the information needed—through the various end of round reports—to maximize those scores.

Over eight rounds, a maximum cumulative total of 8,000 points is possible through the Analyst Report scoring. However, a review of prior class results (of computer and student-run companies shown in Tables #1 and #2) and consultation with a representative at the company that created and supports the simulation reveals that a typical “best” score, even for a computer-run company, is in the 6,000-6,500 point range (B. Frank, Capsim Customer Service, personal communication, December 15, 2017).

The pre-pedagogical change data was collected from the eight classes (over the four year period), consisting of 33 student teams each with 3-6 individual members. The total number of students involved in these 33 teams was 226 (see Table #1). The post-pedagogical change data was collected from the three subsequent classes over a one-year period, consisting of 11 student teams and 39 total students (see Table #2). It is worth noting that the sample size for the scores collected after the change in pedagogy was unavoidable. The same “post-change” methodology could not be repeated beyond these three classes, since the use of the simulation in this course was completely changed in the subsequent year following a department-level curriculum and course content review.

Finally, it is worth acknowledging that Schmeller (2019) found that opening post-round reports (i.e., Analyst Report, Capsim Courier) was more predictive of simulation performance than login frequency. However, this study controlled for that by having every team member open and view reports during in-class team meeting times before and after each round of the simulation (both pre-pedagogical change and post-change). As discussed previously, the simulation was used as an integrated, in-class learning tool—not a stand-alone assignment.

RESULTS

Table #1 provides a summary of Analyst Report scores by teams from previous classes, along with both total team login information and the number of logins by the most “active” individual student on each team. There is no way to assess the amount of time spent by students on the simulation, only the number of times a student logged into the simulation. A summary of average score by class (i.e., semester) is also given—with Class 1 being the oldest class group, and so on.

Similar data collection was completed for the three class groups who participated in the “post-pedagogical change” class as noted in Table #2.

It’s important to note that over the eleven classes of team data there is a mix of “typical” daytime undergraduate student, evening class undergraduate student (i.e., typically a working adult going back to college), and MBA student (see Table #3).

When teams’ Analyst Report scores were ranked (1 = highest score; 44 = lowest score) Graduate student teams averaged a rank of 15.25, as compared to 19.8 (Mixed), 24.75 (Undergraduates-Day), and 24.9 (Undergraduate-Night). While not central to this study, one would expect higher performance from Graduate teams as compared to Undergraduate counterparts using Capsim in a similar simulation environment, and this appears to be supported by the data.

Another observation among class types is the scope of improvement pre and post-pedagogical change with Undergraduate Day student teams (see Table #3). The average Analyst Report scores before the change was 2785 (15 teams) while the one equivalent class in the post-change group showed an average of 4375 (5 teams).

The overall average Analyst Report scores for those post-change classes was 3792, compared to 3467 for the class scores prior to the change of teaching approach using the Footrace method. This indicates some initial support for *Hypothesis #1 – The use of Footraces by individual students for practice round in Capsim lead to better results in team-based competition*

rounds (as measured by team cumulative Analyst Report scores). A final note on total Analyst Report Scores: The computer score did drop as a result of the pedagogy change, but this is not a zero-sum scoring system. It is possible for all teams to score fairly well in any one simulation's industry environment.

Comparison of Mean Capsim Scores- Aggregate and By Class Type

The full set of data was analyzed using SPSS statistical software (version 24). First, to test if both the "non-footrace" (pre-test) and "footrace" (post-test) groups were normally distributed, the Shapiro-Wilk test was applied. The scores for both the non-footrace (.098) and footrace (.449) groups were found to be normally distributed. In addition, there were no outliers found in either independent group using a simple boxplot. Next, an Independent Samples Test ("t-test") was applied to the data. First, using Levene's Test for Equality of Variance, it was concluded that there was homogeneity of variance for the Capsim Scores ($p = .166$). Likewise, the "footrace" Capsim Score average was found to be 323 +/- 526 (mean +/- standard deviation) higher than the average "non-footrace" Capsim Analyst Report scores with a 95% degree of confidence. However, the t-test itself provided a p-value of .542, indicating no statistical difference between the two sets of mean Capsim Analyst Report scores. Thus, upon initial inspection there was no evidence to support Hypothesis #1 (H1).

The full data set, however, consists of both MBA and Undergraduate class scores. An inspection of Table #3 and the rank-order analysis shows higher, more consistent scores with the MBAs than the Undergraduate class groups. As such, a second t-test was completed separating the Undergraduate Capsim score from the MBA scores.

As was done with the aggregate t-test, the Shapiro-Wilk test was applied to the Undergraduate groups only to test for normal distribution. This yielded scores of (.177) and (.485), respectively from "non-footrace" and "footrace" groups, indicating a lack of normal distribution. Levene's

Test for Quality of Variance concluded that there was a lack homogeneity of variance for these Capsim Scores ($p = .044$). The “footrace” scores were found to be 345.37 ± 777 (Mean $\pm$ standard deviation) points higher than the average “non-footrace” scores, with a p-value of .667, assuming variances not equal, indicating that the difference in mean scores was not statistically significant for the Undergraduate class groups.

The same was completed for the MBA groups. Average Capsim score for the non-footrace teams in this group was 4133 ± 417 , as compared to 4504 ± 844 for the footrace groups. There were no outliers found using a simple box plot. The Shapiro-Wilk yielded scores of .193 and .097, for non-footrace and footrace groups, indicating normal distribution. Likewise, Levene’s Test for Quality of Variance showed homogeneity of variance (.973). However, the mean difference in scores (370.27) was not found to be statistically significant between non-footrace and footrace MBA teams $t(11) = .418$, $p = .684$. As such, there was no support for the following hypotheses:

H1: The use of Footraces by individual students for practice rounds in Capsim leads to better results in team-based Tournament competition rounds (as measured by team cumulative Analyst Report scores).

Relation of Team and Individual Logins to Performance

The simulation also allows the instructor to track and review login activity by individuals and teams. The data was analyzed to determine the level of correlation between Capsim Scores and both Average Team Logins (H2) and Top Individual Logins (H3) using the Pearson’s Correlation Coefficient. First, tests using Shapiro-Wilk indicated a lack of normality in distribution for Average Team Logins (.031) as well as for Top Individual Logins (.000). However, given the robustness of the Pearson’s correlation coefficient with sample groups that deviate from normality, the test was still used. As shown in Table #5, there was a moderate positive correlation between the Capsim Scores and Top Individual Logins ($r = .330$, $p = .029$).

Likewise, the correlation between Average Team Logins and Capsim Scores indicated a moderately positive relationship ($r = .424$, $p = .004$). Both were found to be statistically significant explaining 10.9% and 18% of the variation, respectively. As such, there was support for Hypothesis #2 and #3, respectively:

H2: Higher average logins per team is a predictor of group simulation success (as measured by team cumulative Analyst Report scores).

H3: Overachieving individual login volume in each team will be a predictor of team success (as measured by team cumulative Analyst Report scores).

Improvement Trend Over Time of When Class Was Taught

Next, the data was analyzed to see if Capsim Analyst Report scores had improved over time (i.e., with each subsequent attempt to use the simulation in class over the four year period). A scatterplot of Capsim Scores against Class (oldest to most recent) indicated a potential linear relationship between the variables (Graph #1).

After the Capsim Scores and Class (Time) data were found to have residuals that were independent (Darbin-Watson statistic of 2.012) and normally distributed, a regression analysis indicated that *when* the class was taught accounted for 4.2% of the variation in Analyst Report scores and was not statistically significant $\{F(1, 42) = 2.886, p = .097\}$. The same was true when looking at Average Capsim Score by Class $\{F(1, 9) = 1.937, p = .197\}$.

However, a closer look at the scatterplot shows more noise in the data in the post-pedagogy change classes. It is possible that the newness of this changed course design (different from the previous eight classes that had been taught the same way) was skewing the results. Therefore, a regression analysis was done on only the pre-pedagogical change classes (classes 1-8). There was an independence of residuals, as indicated with a Durbin-Watson statistic of 1.893. Residuals were shown to be normally

distributed indicating homoscedasticity. The linear regression established that the experience of teaching the simulation over time accounted for 10.8% of the predicted variability in Analyst Report scores. This result was found to be statistically significant, $F(1,31) = 4.87$, $p < .05$. As a result, there was mixed evidence in support of the following hypothesis:

H4: The average Analyst Report Scores, by class, increases as the course is taught with increasing frequency by the same instructor.

Relation to Number of Student versus Computer Teams

An analysis of the scores by number of student teams, shown in Table #5, indicated that fewer computer teams leads to higher scores by student teams.

To test if the differences between mean Analyst Report scores were statistically significant an Independent t-test was conducted between each of the three different student group populations listed in Table #5. First, to test if the groups were normally distributed, the Shapiro-Wilk test was once again applied. The Capsim scores for the Three Student Team (.119), Four Student Team (.324), and Five Student Team (.239) groups were found to be normally distributed. However, there were outliers found in the Four Student Team group using a simple boxplot. As such, the Independent t-test was used to compare the Three and Five Student Team groups, but for the comparison of Three to Four and Four to Five, the Mann-Whitney U Test was used.

The Mann-Whitney U test is a rank-based nonparametric test that can be used to determine if there are differences between two groups on a continuous or ordinal dependent variable. It is often used as a nonparametric alternative to the Independent-samples t-test when the data fails the assumptions of the independent-samples t-test—such as the presence of outliers found in the Four Student Team score in this case.

As such, the Mann-Whitney U test was run to determine if there were differences in engagement score between the Four and Five Student

Team groups first. Distributions of the engagement scores for the groups were not similar, as assessed by visual inspection. However, there was a statistically significant difference in engagement scores between Four and Five Student Team groups mean Analyst Report scores, $U = 171$, $z = 1.985$, $p = .048$, thus, showing support for Hypothesis #5:

H5: The individual team Analyst Report scores for classes with Five Student Team members are higher than those Four Student Teams.

Likewise, distributions of the engagement scores for the Three and Four Student Team groups were not similar, as assessed by visual inspection. However, there also was no statistically significant difference in engagement scores between Three and Four Student Team groups, $U = 93$, $z = 1.212$, $p = .242$. Thus, there was not support for Hypothesis #6:

H6: The individual team Analyst Report scores for classes with Four Student Team members are higher than those Three Student Teams.

Finally, for the Independent Samples Test ("t-test") to compare mean scores between Three and Five Student Team groups, first, using Levene's Test for Equality of Variance, it was concluded that there was homogeneity of variance among the Capsim Scores ($p = .697$). Likewise, the Five Student Team Capsim Score average was found to be 1469 +/- 554 (mean +/- standard deviation) higher than the average Three Student Team Capsim Analyst Report scores with a 95% degree of confidence. The t-test provided a p-value of .013, indicating a statistically significant difference between the two sets of mean Capsim Analyst Report scores. Thus, there was support for Hypothesis #7:

H7: The individual team Analyst Report scores for classes with Five Student Team members are higher than those Three Student Teams.

Relation to Number of Student versus Computer Teams (Undergraduates Only)

As shown in Table #5, Analyst Report scores for just the Undergraduate classes in this study seem to indicate a difference between the three, four, and five student team industry environments. As such, those groups were isolated. First, to test if the groups were normally distributed, the Shapiro-Wilk test was once again applied. The Capsim Analysis Report scores for the Four Student Team (.324) and Five Student Team (.557) groups were found to be normally distributed. The same was not found to be true for Three Student Team (.044). Given the similarity of sample sizes for each group, 9, 12, and 10 respectively, and with a visual inspection, the decision was made to assume normality and continue with the Independent t-test for this section (Laerd, 2015). However, there were outliers found in the Four Student Team group using a simple boxplot. With the robustness of the t-test in mind, the Independent t-test was used to compare the Three and Five Student Team groups, but for the comparison of Three to Four and Four to Five Student Team groups, the Mann-Whitney U Test was once again used.

First, the Mann-Whitney U test was run to determine if there were differences in engagement score between the Four and Five Student Team groups. Distributions of the engagement scores for the groups were not similar, as assessed by visual inspection. There was no statistically significant difference in engagement scores between Four and Five Student Team groups mean Analyst Report scores, $U = 88$, $z = 1.846$, $p = .069$. As such, there was no support for Hypothesis #5 (Undergraduates only).

The Mann-Whitney U test was used again, and the distributions of the engagement scores for the Three and Four Student Team groups were not similar, as assessed by visual inspection. However, there was a statistically significant difference between the Four Student Team engagement scores (mean rank = 13.58) and the Three Student Team groups (mean rank = 7.56), $U = 85$, $z = 2.203$, $p = .023$. Thus, there was support for Hypothesis #6 (Undergraduate only).

For the Independent Samples Test (“t-test”) to compare mean scores between Three and Five Student Team groups, first, using Levene’s Test for Equality of Variance, it was concluded that there was homogeneity of variance among the Capsim Scores ($p = .557$). Likewise, the Five Student Team Capsim Score average was found to be 2118.23 $\pm$ 672.18 (mean $\pm$ standard deviation) higher than the average Three Student Team Capsim Analyst Report scores with a 95% degree of confidence. The t-test provided a p-value of .006, indicating a statistically significant difference between the two sets of mean Capsim Analyst Report scores. Thus, there was evidence to support Hypothesis #7 once again (Undergraduates only).

DISCUSSION

A number of beneficial results—related to the effective use of the Capsim simulation, and simulations in general, and for future research on using them effectively—can be gleaned from this study. First, a re-examination of the value of providing a “Footrace” Capsim environment for students before launching into team Competition Rounds (with scores provided by the accompanying Analyst Report as a measure of overall performance) was clouded when Undergraduate and Graduate scores were analyzed together. The collective data showed that the difference of mean Analyst report scores, while higher when including these “Footrace” practice rounds, was not statistically significant. Likewise, when the Undergraduate and Graduate classes were analyzed independently the results showed that there was no statistically significant difference in mean Analyst Report scores.

This result runs counter to what Forsyth & Anastasia (2016) observed in their research. Intuitively, one would think that immersing each student in the full simulation before creating teams would stimulate a quicker learning curve, and, as a result, higher scores. It is possible the lack of support for this hypothesis was impacted by the small post-pedagogy change sample size. Likewise, the overall quality of student interaction with the simulation during these practice rounds was not observed or

measured in any specific way—beyond merely proving that students had made decisions for four rounds. A more focused and earnest approach to introducing practice rounds might have led to score in support of Hypothesis #1.

As for Average Team Logins as a predictor of performance, the Pearson Correlation score ($r = .424$, $p = .004$) indicated that this relationship was fairly strong (Cohen, 1988). Thus, there was support for Hypothesis #2. As one might expect, teams likely perform better as a result of each individual learning about the simulation and engaging in it (as measured by average logins). It is also possible that these were teams of better students to begin with (regardless of the attempt to create even teams based on the “Snake Draft” team assignment approach), who, as a result, logged into the simulation more frequently. In hindsight, perhaps a survey of the students’ desires to get a good grade or perform well on the simulation could have been done to test the quality/intentions of each team. Comparing these “I care about this assignment” scores with team results might clarify this causal relationship between Average Login activity and Analyst Report scores.

The decision-making among teams involved in this study varied, as each team was free to determine that process for themselves. Some teams opted to make decisions for each functional area (R&D, Production, etc.) together, while other teams assign certain functional area responsibilities to individuals. Likewise, as observed anecdotally, these roles change over the course of the semester as well. Therefore, it is not unusual to expect better outcomes (specifically Analyst Report scores) with higher average team logins. These differences in decision-making style and simulation success also serve as another area for further research.

Likewise, as Kilburn & Kilburn (2012) highlight, it is possible that individual “stars” on teams create a unique competitive advantage, leading to higher scores. One way they propose to identify these individuals is through total logins by individuals (which the Capsim software tracks). Presumably students that care more, work harder, and login more

frequently will drive simulation teams to good performance. The data shown in Table #5 (Pearson Correlation of $r = .330$, $p = .029$) indicates support for this idea.

Of course, from a teaching perspective the goal of using computer simulations in class is to address the need for real-world experience for all students, not have one student take over the simulation. In this way, having simulation scores counts toward the class grade can be counterproductive. If one student on the team is excelling at making all the decision, then the scores will be high and grades for all students in the team will be as well. Having one individual lead to greater team Analyst Report scores goes against the goal of total team learning. As such, tracking individual login activity in team-based computer simulations might actually be a good way to guard against “star” leadership in the face of the typical team-based social loafing. Likewise, no comparison of individual learning, as tied to the simulation, was done in this study to see if the Analyst Report scores was indeed adequate measure of student learning. Future research on the effectiveness of simulation use in student learning could include both team-based results from the simulation and individual learning scores (i.e., on exams).

Improvement over time is likely (and hopefully) a result of the instructor continuing to get better at teaching the Capsim simulation. In all of these classes, the students were required to complete the Capsim tutorials and go through 4-6 practice rounds before launching into the Competition Rounds during which the Analyst Report scores were taken. The one anomaly is Class #1, as this was the only class where students completed the tutorials outside of class time. Since that time, students have been given time in class, with the instructor present, to complete the tutorials. This might help to explain the low scores in Class #1 and lead to another hypothesis concerning completion of tutorials in class versus out of class for future research. That said, while the visual depiction of the data shown in Graph #1 seems to show a relationship between the instructor

experience teaching with the simulation and Analyst Report score, the regression analysis did not end up supporting Hypothesis #4.

Further visual inspection of the scores from Graph #1 indicated that the later classes (post-pedagogy change) did not seem in line with the previous trend of scores. By taking them out and only focusing on the pre-pedagogy change classes, the regression analysis showed some support for Hypothesis #4. Just as we saw support for instructor experience leading to higher scores, it is possible that the pedagogical change influenced teaching effectiveness as the instructor began “re-learning” how to deliver the course using this new “Footrace” options.

Perhaps the most important predictor of student performance was the makeup of the industries between classes. Due to varying class sizes and a desire to have 3-6 students on each team, the number of student-run versus computer-run companies in the simulation varied from three to five student teams among class groups. In fact, the “pre-pedagogy change” teams in industries made up of five student teams performed nearly two times better, as measured by the Analyst Report scores, than those in three student team environments. The comparative data between the Three and Four Student Groups and the Four and Five Student Groups showed mixed support for Hypothesis #5 and #6. The aggregate analysis showed support for Hypothesis #5 and the Undergraduate-only analysis showed support for Hypothesis #6.

The computer-run companies don’t make mistakes (i.e., errors in forecasting sales, choosing the right financing mix) and are programmed to beat their competition based on the strategy the company is programmed to follow (i.e., focused differentiation, broad low cost). Therefore, more computer-run companies mean a more difficult competitive industry environment for the student teams. It is also important to note that the simulation does allow three performance setting for the computer-run companies—Easy, Moderate, and Difficult. All classes in this study were set to moderate, limiting any significant computer performance variation. Further research exploring the overall effect of student-run

companies' performance (as a measure of learning) with these varying levels of competitive difficulty is needed as well.

From a student evaluation perspective, it is necessary to "level the playing field" between classes when using computer simulation results as a part of the overall course grade. In this case, it is not fair to treat each class simulation the same when determining grades based on a team's company success. While attaching a grade to simulation results hopefully leads to an adequate amount of "skin in the game" for student focus and involvement, in this case the game is rigged based on class size (and as a result diminishing numbers of computer-run teams). One solution to this problem could be adjusting between the Easy, Moderate, and Difficult settings of the simulation. This would lead to more equitable and comparable grading based on Capsim performance between classes.

STUDY LIMITATIONS

This study focused only on team performance (as measured through their Analyst Report scores). Limitations of this study are connected to both team selection and team dynamics. While Kilburn, Kilburn & Faught (2010) found that pre-competition student assessment scores and average GPA were not predictors of final group ranking (i.e., success), the "snake draft" approach was still used to provide some measure of spreading out the talent among all of the teams based on prior class performance (i.e., grades). That said, this team assignment technique does not take into account the personalities, learning styles, communication styles, and leadership abilities of the students in these teams. A number of "soft skill" and personality-based traits can play a role in overall team success.

For instance, Schonecker, Martell, and Michlitsch (1997) found that strong individuals dominating the group can negatively influence performance. Gosenpud and Washbush (1996) found a connection between the Myers-Briggs personality type inventory and performance. One study showed that teams with low cohesiveness, low heterogeneity, high oppor-

tunistic practices, and high hypothesis-driven thinking performed better on simulations than groups that were alike and got along (Anderson, 2005). This research is supported by Adobor and Daneshfar (2006), whose research showed that task conflict in the team (i.e., idea exchange) increased individual team member learning and emotional conflict reduced learning. Each of these factors likely play a role in simulation performance beyond a mere change in pedagogy or delivery.

Likewise, team success using simulations can also be affected by team size. Wolfe and Chacko (1983) showed that teams with 3-4 students increased knowledge of theory and fact in strategic management more than those operating in smaller teams. However, no teams in this study had fewer than three students. Again, this study did not explore the differences in performance based on the number of students in the team. A further exploration of the data could be beneficial in testing this conclusion. Additionally, social interaction that is psychologically safe has been linked to better performance (Xu & Yang, 2010) as has framing the simulation in a way that students actually feel like the CEO (Tiwari, Nafees & Krishnan, 2014). Finally, gender has been shown to play a role in decision-making style, attitudes towards the simulation, and team dynamics (Garber & Clopton, 2004; Kaenzig, Hyatt, & Anderson, 2004; Sadler-Smith & Riding, 1999). Each of these elements of team selection, management, and dynamics certainly also may play some role in determining simulation success and overall student learning.

Finally, the Analyst Report scores serve as a measure for overall learning in the simulation. That said, there are potential weaknesses with this metric as noted previously (Frezzo, Behrens, & Mislevy, 2010; Neely & Tucker, 2012). While the link between online simulations and learning outcomes has been established (Higher Education, 2017), additional research should be done to assess the link between specific simulation performance scores and learning outcomes.

CONCLUSIONS

This research sought to explore the role that pedagogy choice plays in the overall effectiveness of computer simulation use. Specifically, one of the more widely used business strategy simulation tools, the Capsim Capstone, was used in this study. Given its features and previous research, this study also explored the roles of average team and individual login behavior in overall simulation performance. Additionally, instructor experience with the simulation and the simulation operating environment (i.e., the number of computer versus student-run teams) was also explored.

Online business simulations in general, and the Capsim simulation specifically, offer a variety of opportunities to connect theory with practice. With the pervasiveness of these teaching tools, it is worthwhile to explore how best to use them to enhance educational outcomes. With the Capsim simulation specifically, this research was not able to show evidence for the value of assigning the “Footrace” mode to individual students in early practice rounds to increase team-based Analyst Report scores in the competition rounds. However, while there was no statistically significant support, the limited data did show a slight increase in scores and could have been effected by the “newness” of the pedagogical change, warranting further research.

The same can be said for relative instructor experience with simulation usage. With experience comes better student results—hopefully, from improved teaching using these simulations. As such, any new faculty members assigned with classes using these types of simulation would benefit from “shadowing” or being mentored by faculty with experience using them. This would increase the learning curve, and thus, improve the student learning experience.

This research also has helped to identify the importance of using team logins and individual logins as an ongoing assessment tool and predictor for team performance. Specifically, when pulling apart the simulation results (i.e., Analyst Report scores) and student learning,

increased individual login activity may actually be an indicator that the whole team is leaning on one team member for decisions, and thus, not learning as they would with increased total team member logins. Capsim, in particular, allows the instructor to track these team login behaviors, but other gaming simulations should be managed in such a way to deter this type of individual login behavior.

Finally, the study has shown support for the hypothesis that fewer student teams (as opposed to computer-run companies) in the industry-operating environment lead to lower Analyst Report scores. Here, the use of these scores as a proxy for student learning, which is a common practice (Karriker & Aaron, 2014), may begin to fail as an ideal measure for grading and evaluation. While the industry-operating environment becomes more difficult with more mistake-free computer teams competing for share and profit, learning through failure remains a viable outcome for student teams. Lower scores do not equate a lack of learning. In fact, anecdotally, those teams that have made early mistakes in the simulation are often the ones that are most earnest in learning all they can to avoid continued failure (with a portion of the grade tied to their company performance).

That said, it is useful to be aware of these differences of performance in varying student and computer team environments if the Analyst Report scores are used in grading. Fairness would dictate some grading adjustments from team to team and from class to class given these conclusions. From an evaluation and grading perspective, it's important to understand the inequity of expecting student teams in three or four student team environments to score as well as student teams where there is only one computer-run company. For these classes the Analyst Report scores are counted for a portion of the grade, but there is no formal adjust made between three, four, and five student team environments. This type of adjustment, with this specific simulation tool, should be considered when simulation performance is included in the students' final grade. Similarly, the nuances of any computer simulation used in a

classroom, and for grading, should be fully understood before creating graded assignments tied to simulation results.

REFERENCES

- Adams, S. G. 2003. Building successful student teams in the engineering classroom, **Journal of STEM Education: Innovations and Research**, 4(3), 1-6.
- Adobor, H., & Daneshfar, A. 2006. Management simulations: Determining their effectiveness. **Journal of Management Development**, 25(2), 151-168. <https://doi.org/10.1108/02621710610645135>
- Ahn, J. 2008. Application of the experiential learning cycle in learning from a business simulation game. **E-Learning and Digital Media**, 5(2), 146-156. <https://doi.org/10.2304/elea.2008.5.2.146>
- Anderson, J. R. 2005. The relationship between student perceptions of team dynamics and simulation game outcomes: An individual-level analysis. **Journal of Education for Business**, 81(2), 85-90. <https://doi.org/10.3200/JOEB.81.2.85-90>
- Azriel, J. A., Erthal, M. J., & Starr, E. 2005. Answers, questions, and deceptions: What is the role of games in business education? **Journal of Education for Business**, 81(1), 9-13. <https://doi.org/10.3200/JOEB.81.1.9-14>
- Bell, R., & Loon, M. 2015. The impact of critical thinking disposition on learning using business simulations. **The International Journal of Management Education**, 13(3), 119-127. <https://doi.org/10.1016/j.ijme.2015.01.002>
- Brown, K. G., Arbaugh, J. B., Hrivnak, G., & Kenworth, A. 2013. Overlooked and underappreciated: What research tells us about how teaching must change. In B. C. Holtom & E. C. Dierdorff (Eds.), **Disrupt or be disrupted** (pp. 219-258). San Francisco, CA: Jossey-Bass.
- Cadotte, E., Riley, R., Bonney, L., & MacGuire, C. 2013. **The role that large scale, integrative business simulations can play in assurance of learning and assessment**. Retrieved from <https://wwwmkt.ilscdn>.

com/images/docs/The-Role-that-Large-Scale-Integrative-Business-Simulations-Can-Play-in-Assurance-of-Learning-and-Assessment.pdf

Capsim Professor Guide. n.d., Retrieved March 3, 2018 from http://ww3.capsim.com/permanent_links/Capsim_Professor_Guide.cfm

Capsim Team Member Guide. n.d., Retrieved March 6, 2018 from http://ww3.capsim.com/modules/GIA/files/2015C_0/0/Capstone/EN/PDF/2014_Capstone_Team_Member_Guide.pdf

DeMeglio, F. 2008. Virtual workplaces in the classroom. **Business Week Online**, January, p23.

Egereonu, F. O. 2011. **Business simulation: An exploration of the virtual enterprise program.** UMI Dissertations Publishing.

Faria, A. J. 1986. A survey on the use of business games in academia and business. **Simulation & Games**, 18, 207-224. <https://doi.org/10.1177/104687818701800204>

Forsyth, B., & Anastasia, C. 2016. The business simulation paradigm: Tracking effectiveness in MBA programs. **Journal of Management Policy and Practice**, 17(2), 85-100.

Frezzo, D., Behrens, J., & Mislevy, R. 2010. Design patterns for learning and assessment: Facilitating the introduction of a complex simulation-based learning environment into a community of instructors. **Journal of Science Education & Technology**, 19(2), 105-114. <https://doi.org/10.1007/s10956-009-9192-0>

Garber, L.L., Jr., and Clopton S. W. 2004. **Participation in market-ing games: An examination of the student experience including gender effects**, Appalachian working paper, Boone, NC: Appalachian State University.

Gosenpud, J., & Washbush, J. 1996. Total enterprise simulation performance as a function of Myers-Briggs personality type. **Simulation & Gaming**, 27, 184-205. <https://doi.org/10.1177/1046878196272004>

Gove, S. 2012. Increasing student engagement using client-based peer assessment in multi-role, whole enterprise simulations. In S. DuBravac (Ed.), **Increasing student engagement and retention using immersive interfaces: Virtual worlds, gaming, and simulation.**

Cutting-edge technologies in higher education (pp. 95-128). Bingley, UK: Emerald Group. 7 [https://doi.org/10.1108/S2044-9968\(2012\)000006C007](https://doi.org/10.1108/S2044-9968(2012)000006C007)

Gremmen, H., & Potters, J. 1997. Assessing the efficacy of gaming in economic education. **The Journal of Economic Education**, 28(4), 291. <https://doi.org/10.2307/1183414>

Kaenzig, R., Hyatt, E. M., & Anderson, S. E. 2004. **Exploring gender differences in college of business educational experience**. Paper presented at the Annual Meeting of the American Accounting Association, Orlando, FL.

Karriker, J. H., & Aaron, J. R. 2014. More than just fun and games: BSG and Glo-Bus as strategic education instruments. **Journal of Management Education**, 38(5), 768-775. <https://doi.org/10.1177/1052562914534245>

Kilburn, B., & Kilburn, A. 2012. The team vs. the individual: Login activity as a predictor of web-based simulation team success. **Academy of Educational Leadership Journal**, 16(3), 15-22.

Kilburn, B., Kilburn, A., & Freight, S. 2010. Web-based simulation team success: The role of pre-competition assessment, group size, and group average GPA. **Journal of Learning in Higher Education**, 6(2), 89-94.

Klassen, K. J., & Willoughby, K. A. 2003. In-class simulation games: Assessing student learning. **Journal of Information Technology Education**, 2(1), 1-13. <https://doi.org/10.28945/306>

Laerd Statistics. 2015. Independent-samples t-test using SPSS Statistics. **Statistical tutorials and software guides**. Retrieved January 4, 2021 from <https://statistics.laerd.com>

Ledman, R. 2016. **Using a simulation for a capstone course: lessons learned**. Proceedings of the 23rd Annual American Society of Business and Behavior Sciences Conference, Las Vegas, NV.

Lohmann, G., Pratt, M.A., Benckendorff, P. et al. 2019. Online business simulations: authentic teamwork, learning outcomes, and satisfaction. **Higher Education**, 77(3), 455-472. <https://doi.org/10.1007/s10734-018-0282-x>

- Miles, M. P., Franklin, G. M., Heriot, K., Hadley, L., & Hazeldine, M. 2014. AACSB International's 2013 accreditation standards. **Journal of International Education in Business**, 7(2), 86-107. <https://doi.org/10.1108/JIEB-12-2013-0047>
- Neely, P., & Tucker, J. 2013. Case study: An examination of the decision making process for selecting simulations for an online MBA program. **Education Training**, 55(2), 128-138. <https://doi.org/10.1108/00400911311304788>
- Poisson-de Haro, S., & Turgut, C. 2012. Expanded strategy simulations: developing better managers. **Journal of Management Development**, 31(3), 209-220. <https://doi.org/10.1108/02621711211208844>
- Rainie, L., & Anderson, J. 2017, May 03. **The Future of Jobs and Jobs Training**. Retrieved from <http://www.pewinternet.org/2017/05/03/the-future-of-jobs-and-jobs-training/>
- Sadler-Smith, E., & Riding, R. 1999. Cognitive style and instructional preferences. **Instructional Science**, 27, 355-371. <https://doi.org/10.1007/BF00892031>
- Saulnier, D. 2009. **Business solutions: An interview with Capsim's Dan Smith** [blog post]. Retrieved March 6, 2018 from <http://saulnier.typepad.com/blog/2009/03/business-simulations-an-interview-with-capsims-dan-smith.html>
- Schoenecker, T., Martell, K., & Michlitsch, J. 1997. Diversity, performance and satisfaction in student group projects: An empirical study. **Research in Higher Education**, 38, 479-495. <https://doi.org/10.1023/A:1024966627400>
- Snow, S. C., Gehlen, F. L., & Green, J. C. 2002. Different ways to introduce a business simulation: the effect on student performance. **Simulation & Gaming**, 33(4), 526-532. <https://doi.org/10.1177/1046878102238617>
- Tan, K., Tse, Y., & Chung, P. 2010. A plug and play pathway approach for operations management games development. **Computers & Education**, 55(1), 109-117. <https://doi.org/10.1016/j.compedu.2009.12.008>

- Umble, E. J., Umble, M., & Artz, K. 2008. Enhancing undergraduates' capabilities through team-based competitions: The Edward Jones challenge. **Decision Sciences Journal of Innovative Education**, 6(1), 1-27. <https://doi.org/10.1111/j.1540-4609.2007.00164.x>
- Wolfe, J. 2016. Assuring business school learning with games. **Simulation & Gaming**, 47(2), 206-227. <https://doi.org/10.1177/1046878116632872>

CITATION INFORMATION

Fletcher, Kevin A. "Maximizing Student Learning Outcomes When Using Business Strategy Simulation Tools: The Role of Pedagogy, Participation, and Practice (A Multi-Year Study)." *The BRC Academy Journal of Business* 11, no. 1 (2021): 167–198. <https://dx.doi.org/10.15239/j.brcacadjb.2020.11.01.ja07>

WEB APPENDIX

A web appendix for this paper is available at: <https://dx.doi.org/10.15239/j.brcacadjb.2021.11.01.wa07>