Gamifying M&S Transportation Education & Training to Improve Engineering Learning Outcomes

Rachel Su Ann Lim[1], Kevin F. Hulme, Ph.D., CMSP[1], Aaron Estes, Ph.D.[2], Presentacion Rivera-Reyes, Ph.D.[3], Jesse Hartloff, Ph.D.[4], Stephen E. Still, Ph.D.[1], and Mark Schiferle[2]
1: The Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL);
2: Department of Mechanical and Aerospace Engineering;
3: Department of Engineering Education;
4: Department of Computer Science and Engineering
University at Buffalo, Buffalo, NY

rlim2@buffalo.edu; hulme@buffalo.edu; aaronest@buffalo.edu; privera1@buffalo.edu;
hartloff@buffalo.edu; sestill@buffalo.edu; ms286@buffalo.edu

ABSTRACT

Engineering education practitioners continue to identify novel mechanisms to dramatically increase the effectiveness of next-generation training towards vocational preparation and workforce development. To this end, the discipline of Modeling & Simulation (M&S) remains a valuable resource that enables learners to bridge the gap between conceptual experimentation and real world implementation. Gaming experiences that incorporate physics-based modeling and high-fidelity simulation effectively blend synthesis and analysis in a manner that empowers students to modify parameters - and observe downstream impacts – all in real-time. Recent engineering education research efforts have incrementally focused on the advancement of serious play, including novel graphics tools and gaming engines that continue to emerge as powerful mechanisms for next-generation training. Likewise, Gamification (i.e., incorporating typical elements of game playing, including rewards to motivate, point scoring, and group competition) continues to gain recognition as a valuable resource for educators.

To accelerate engineering education training effectiveness for the improvement of learning outcomes, we describe the design and pilot deployment of an innovative Gamification-based experiential learning andragogy. We establish a targeted series of assessment materials based on the present state-of-the-art in the literature to reliably measure key outcomes (e.g., conceptual knowledge, learner engagement, and self-regulated learning) from our novel implementation. For this paper, the disciplinary focus area is Emerging Practices in Transportation Engineering, including an experimental pilot deployment within two graduate-level engineering courses, and one engineering club that has targeted relevance to this discipline. Our presentation will include preliminary results (including statistics and correlations), as well as a discussion of challenges encountered, and a summary of lessons learned. To conclude our paper, we propose a "Theory of Change" for planned best practices to authenticate program functionality and investigate long-term sustainability on a broader scale.

ABOUT THE AUTHORS

Rachel Su Ann Lim received her B.S and M.S. degrees from the Department of Biomedical Engineering at the University at Buffalo. Her contributions to the Motion Simulation Laboratory involve novel healthcare advancement and simulation analysis as a platform to extract features that encourage game-based simulation.

Kevin F. Hulme, Ph.D., CMSP is a Senior Research Associate with the School of Engineering and Applied Sciences at the University at Buffalo. His areas of research and teaching interests include the design, development, and deployment of experiential learning and game-based Simulation elements for next-generation Training. He is a Certified Modeling & Simulation Professional (CMSP).

Aaron Estes, Ph.D. is a Teaching Assistant Professor with the Department of Mechanical & Aerospace Engineering at the University at Buffalo. His teaching interests include dynamics and control (for both ground and flight vehicles), and system identification. Likewise, he has instructional and research-based interests in the novel application of high-fidelity Simulator-based Game-Based Learning within engineering higher education.

Presentacion Rivera-Reyes, Ph.D. is an Assistant Professor of Teaching with the Department of Engineering Education at the University at Buffalo. His areas of focus include curriculum design and innovation, problem-based learning, active learning, and problem solving in engineering education research.

Jesse Hartloff, Ph.D. is an Assistant Professor Teaching with the Department of Computer Science and Engineering at the University at Buffalo. His research areas of expertise include computing education, software systems, and computational theory/algorithms.

Stephen E. Still, Ph.D. is a Professor of Practice with the Department of Civil, Structural, and Environmental Engineering at the University at Buffalo. He is also an active researcher in Transportation Systems Engineering, teaches courses in applied transportation planning and technology, and has more than 30 years of experience in consulting and management of transportation systems.

Mark Schiferle has his B.S. and is currently pursuing his M.S. in the Department of Mechanical Engineering at the University at Buffalo. His primary area of interest includes the novel application of Gamification and Game-based Simulation to improve student training effectiveness in Post-secondary engineering education.

Gamifying M&S Transportation Education & Training to Improve Engineering Learning Outcomes

Rachel Su Ann Lim[1], Kevin F. Hulme, Ph.D., CMSP[1], Aaron Estes, Ph.D.[2], Presentacion Rivera-Reyes, Ph.D.[3],
Jesse Hartloff, Ph.D.[4], Stephen E. Still, Ph.D.[1], and Mark Schiferle[2]
1: The Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL);
2: Department of Mechanical and Aerospace Engineering;
3: Department of Engineering Education;
4: Department of Computer Science and Engineering
University at Buffalo, Buffalo, NY

rlim2@buffalo.edu; hulme@buffalo.edu; aaronest@buffalo.edu; privera1@buffalo.edu;
hartloff@buffalo.edu; sestill@buffalo.edu; ms286@buffalo.edu

INTRODUCTION

Educators continue to innovate and revolutionize mechanisms to improve the effectiveness of learning outcomes in engineering education. For more than a half-century, educators have evolved the fundamental principles of Dale's Pyramid ("Cone of Experience") (Dale, 1969), which theorized learners retain more information resulting from what they actually **do** as opposed to what they might hear/read/observe within an educational setting. More recently, and advancing these concepts, it has long been demonstrated that training effectiveness can be substantially enhanced through participatory activities and "experiential" learning (e.g., Kolb, 1984) – see Figure 1, whereby the trainee serves as an active participant in the process of skills acquisition. The effectiveness of experiential exercises is often attributed to its roots in inductive (active) learning – which begins with specific observations, experiments, or problems that students experience and analyze, and proceeds to the generalization of these observations within an abstract or theoretical framework (Prince & Felder, 2006).

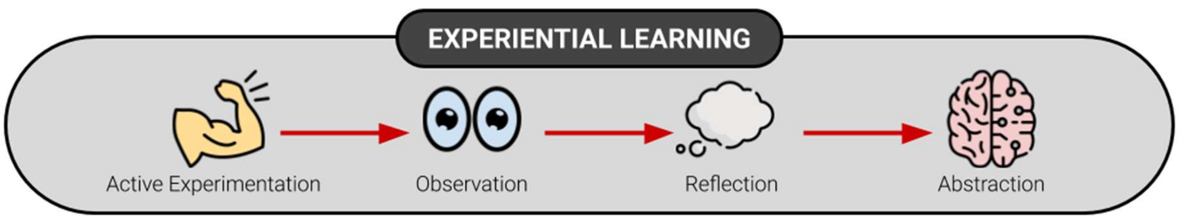

Figure 1: Common components of Experiential Learning

Games (digital or non-digital) have proven to be a convenient medium for creating a domain where students can interact with complex problems, much like in a real-life setting. In more recent times, we have seen the emergence of novel and innovative techniques - specifically geared towards higher education - to increase learning outcomes (e.g., learner engagement, knowledge retention). Namely, Serious Gaming (i.e., employing traditional game techniques and "gameplay" to enable a better understanding of a particular concept), Gamification (i.e., employing game mechanics in non-game situations to enhance motivation and positively influence behavior) (Hughes, 2017), and the overarching notion of Serious Play (Rieber et al., 1998) - specifically within the context of Modeling & Simulation (M&S) (i.e., employing task-specific interactions by simulating scenarios in risk-free physical/virtual environments) (DoD, 1998; Aldrich, 2005; Tolk, 2012) all serve to motivate learners and effectively couple the novelties of entertainment (e.g., badges, scoring, rewards) into an innovative context for skill acquisition.

Although there remains a consensus that serious games have significant potential as an instructional tool, their effectiveness in terms of **learning outcomes** is still understudied (Bellotti et al., 2013). Within the context of engineering education, social learning theory has gained prominence over the last two decades. Social learning can be loosely defined as participating with others to make sense of new concepts, and enabling activities translated into a real-world context. Collectively, these instructional merits implicitly enable students to become active members of a true "community of practice" (Anusha & Reddy, 2015). Within this same context, games have effectively been implemented for instructing interpersonal skills. Social simulation and digital gaming can empower young adults to quickly explore concepts within an engaging (experiential) environment that inherently promotes behavioral learning (Rayborn &- Waern, 2004). In this manner, the game (itself) can be implemented to assign the training goals, rather than a traditional learning context, where learning outcomes are defined by the subject matter expert (i.e., the teacher).

In this paper, we demonstrate the pilot deployment of an innovative Gamification-based experiential learning andragogy for engineering education. Figure 2 describes the study process of a lesson and illustrates a student learning new material with basic or no knowledge of the topic. During the lesson, the instructor delivers a conventional lecture, with mathematical formulas, text and videos to provide the students with a better grasp of the topic. Later, students are exposed to hands-on gameplay. Students are then assessed on the topic in the format of an exam or quiz. Upon successful completion, the student is awarded a badge to acknowledge their participation and achievement. Our disciplinary focus area is ***Transportation Engineering*** - an emergent educational domain that provides the foundational systems that support human endeavor throughout the world and have an extreme impact on our overall quality of life. This discussion includes an experimental pilot deployment within two engineering courses and a student engineering club, as well as preliminary results, data correlations, and lessons learned.

Figure 2: Incorporated Game Play in Lessons

DISCIPLINARY FOCUS AND BROADER IMPACTS

The experimental scenarios described in this paper have domain-specific relevance upon training and education, and with direct relation to the human-machine interface associated with a highly-complex engineering system -- a motor vehicle. Numerous broad engineering disciplines are directly impacted by these technologies, such as ***Emerging Practices in Transportation (STL)***, and ***Road Vehicle Dynamics (RVD)***, student cohorts from which are directly highlighted within the present work. As a companion to methodological demonstration and deployment within full engineering courses -- discipline-specific student clubs (***The Society of Automotive Engineers - SAE***) who have targeted interests in Transportation engineering -- can gain benefit from Gamification as a supplementary hands-on mechanism for concept reinforcement. Identifying innovative mechanisms to Gamify the learner experience is likely to improve practical skill sets (e.g., hand-eye coordination, multi-tasking, cognition, and general problem-solving) (McFadden, 2017; NTSA, 2019). Within academia, it remains a primary goal to develop new methodologies that will help to foster the transition into industry, and correspondingly - present scientific evidence on the added-value and the tangible benefits derived from deployment of these tools (Markopoulos et al., 2015).

EXPERIMENTAL DESIGN

In this section, we present a detailed overview of our Experimental Design: the race course that was designed, and deployed as the GBL/Gamification centerpiece for all training experiments that are highlighted in this work; a discussion of our major equipment; and an overview of our experimental cohorts and outcome measures.

Simulator Race Course: Triple Curve - Cornering Strategy with Tire Saturation

The Triple Curve (Hulme et al., 2021b; Hulme et al., 2019) is a racing simulation on a triangular track with three straight segments joined by tight corners. The students are given simple directions: complete as many ***legal laps*** (i.e., no barrier cones struck) as possible within a two-minute window. This task implies that the students maximize their speed - but quickly learn that braking is essential to maintain control. The conceptual Training link is “tire saturation”; to achieve an optimal lap time, the student needs to accelerate as much as possible, but only within the physical limitations of the tires. This is especially important in corners, where tire traction provides the centripetal force to hold the vehicle in the corner, and the tractional demands increase in proportion to the square of the speed of the vehicle. The length of the straight segments allows students to accelerate to approximately 80 mph before entering each successive corner, and students gradually learn to apply a smooth braking force as they enter corners. The simple geometry of the track, by design, requires students to assume the same (120°) corner over and over again, adjusting their approach with each entry, each time engaging with the physics of tire saturation.

Within the Triple Curve, we conceived and implemented a series of gaming elements to guide the learner towards their primary training objective (legal laps). We programmed the Simulator to collect the trainee's X/Y drive path for the entire exercise, as well as ancillary statistics (e.g., total and legal laps; maximum and average speed; cone strikes/spinouts). There are two primary variations of the Triple Curve: ***The "Gauge" trainer*** (which was leveraged for the present study) enables visualization of technical guidance to observe if certain learners might respond to on-screen overlays that would help participants navigate towards optimized performance. These gauges include spheres embedded in the roadway to suggest a physical drive (racing) line -- each coded with representative color for regions of acceleration (green), braking (red), and transition (yellow). Likewise, a vertical contour gauge was provided to indicate the maximum tire slip angles encountered during a turn. Note that the yellow region (middle of gauge) is the 6^o slip angle "sweet spot," and we encouraged vehicle passengers to coach drivers towards this zone on each turn. Refer to Figure 3, which illustrates the primary GBL components of the Gauge trainer environment.

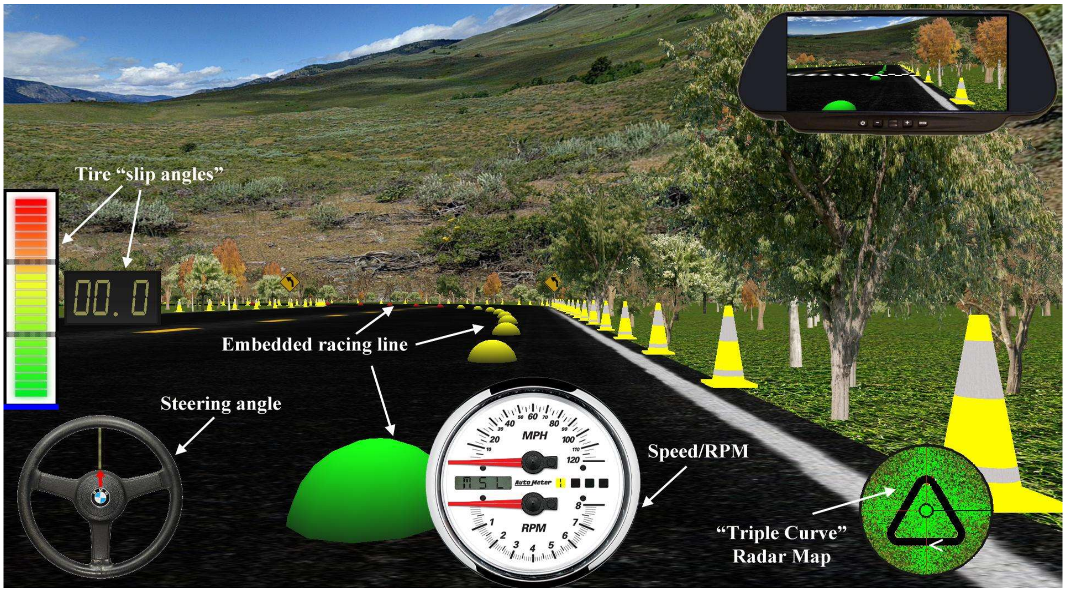

Figure 3: Triple Curve Game-based Learning Environment for (dynamic) Engineering Education

Note that in advance of experimental deployment upon our engineering student cohorts, the research team performed extensive testing internally in an effort to guarantee experimental reliability (i.e., the *consistency* of our measures) and validity (i.e., the *accuracy* of our measures). For the former, we executed the experiment using a wide series of test drivers and initial conditions to guarantee homogeneous model outputs, and for the latter, we implemented physics-based dynamics models of appropriate fidelity (e.g., the "Four-wheel Model", which implements a fully non-linear and load-sensitive tire model, and allows students to experience how changes in key vehicle parameters affect vehicle stability and limit behavior; Milliken & Milliken, 1995). For the purposes of our experimentation, these endeavors sufficiently guaranteed model performance – through simulation - that is suitably-reflective of real-world vehicle behavior, and served to affirm the baseline efficacy of our experimental implementation and methodology.

Simulator - Equipment Overview

The simRING simulator features a 360-degree display screen whose visual rendering provides the occupants of the simulator with a holistic depiction of a virtual driving world, complete with traffic, physical landmarks, traffic control elements and other roadway conditions pertaining to the training exercise. The system likewise features input navigation controls, realistic (6-DOF) motion cueing and washout filtering, as well as a stereo sound system that emulates sounds heard inside and outside the vehicle during a typical driving excursion. Figure 4 showcases our entire six-channel display system (each 1920x1080p; resulting in a 11520x1080p 50Hz composite image).

Figure 4: simRING simulator

Refer to Figure 5, which describes and illustrates our I/O pipeline (Hulme et al., 2021a, Hulme et al., 2016) that fosters our GBL training atmosphere and enables Gamification-based driver performance-rating.

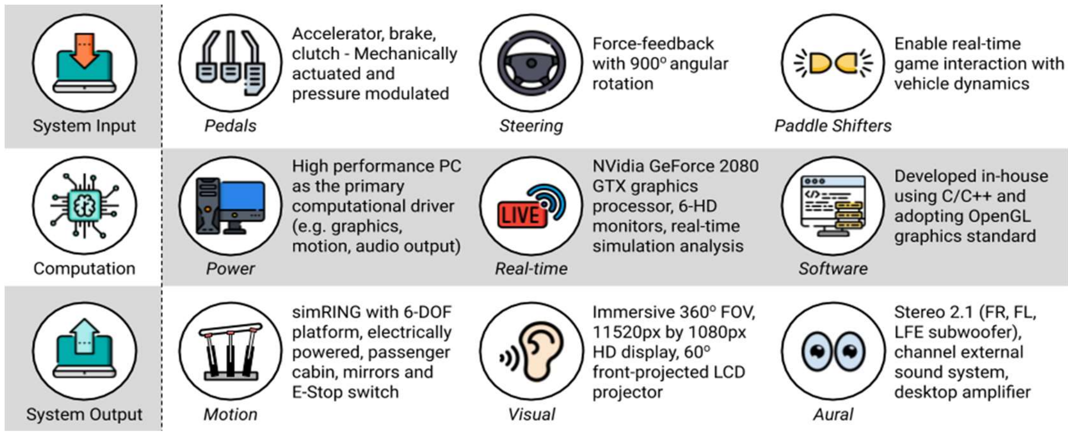

Figure 5: simRING Input/output Rendering Components

System input is provided by the on-board controls, which consist of a force-feedback steering wheel, steering wheel paddle shifters, and force-modulated (gas/brake) pedals. The input signals are then delivered to a high-performance PC workstation. On the computation side, the simulation has been developed using C++ and controls our entire framework, including real-time analysis and physics-based modeling. Lastly, system output rendering is performed, including graphics rendering, calculation of vehicle state outputs to approximate the real-time motion of the platform/vehicle, and aural event cueing which creates a sound "source" and plays the sound when an event (e.g., vehicle ignition, engine idle, squealing tires, cone strikes, crash events, vehicle shutdown) takes place.

Cohort Sizes and Demographics

Our experimental cohort consisted of a total of n=82 participants (74 males, 8 females). Unfortunately, these cohorts are gender-disproportionate - these demographics were based upon course/club enrollment, and therefore, not a variable over which we had direct control. As described previously, all participants were enrolled in a transportation course/club at the time of participation - either RVD, STL, or SAE. Such a disproportionate gender distribution could have impacted the experiment through previously published gender effects in young drivers regarding attitudes, behaviors, and risk perception (Cordellieri et al., 2016) during a simulated driving task – a noted limitation of our study design. Each participant had one opportunity to drive the simRING using the Triple Curve track with the Gauge-trainer GBL-training approach, and to complete either a condensed Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1990; Artino, 2005), where the sub-sample size is denoted as n_1 = 35 (31 males, 4 females), or, a tailored pre- and post- conceptual quiz (CQ), where sub-sample size is denoted as n_2 = 47 (43 males, 4 females). The surveys for these sub-cohorts are described in greater detail, as follows.

Motivated Strategies for Learning Questionnaire (n_1 =35)

Participants in this sub-cohort were from the STL and SAE courses, and were tasked to complete a 12-question MSLQ, used to measure student engagement. The MSLQ consists of two primary components. The first, ***Part A – Motivation***, consists of 31 items that assess student goals, value beliefs for a course, their beliefs about their skill to succeed in a course, and their anxiety about taking tests in a course. The second, ***Part B – Learning Strategies***, consists of 50 items that assess a student's approach (e.g., strategies and study skills) when learning new educational content. The MSLQ questions are rated on a Likert-scale ranging from 1 (not applicable) to 7 (highly applicable). For our abbreviated adaptation, we carefully selected six questions each, from Part A and Part B, from the original 81-question MSLQ.

Conceptual Quiz (n_2 =47)

Participants in this sub-cohort were from a Road Vehicle Dynamics (RVD) course, exclusively. The course topic was aligned with the driving task, utilizing the key components of understanding oversteer and understeer of a vehicle. Participants were tasked to complete a pre- and post- conceptual quiz (CQ) to gauge their knowledge before and after the drive, and to assess if the GBL-experiment enhanced their overall understanding of these key vehicle dynamics concepts. A well-designed CQ should target the assessment of conceptual knowledge rather than procedural knowledge (Sands et al., 2018) - and therefore - should not require the recollection of definitions, but instead should examine a student's ability to make meaningful connections between core concepts. This can typically be accomplished with a series of multiple-choice questions; the familiar Mechanics Baseline Test (Hestenes & Wells, 1992) is an example (validated) measure of conceptual knowledge. In our implementation, each CQ had 4 multiple choice questions, and were explicitly based on the (oversteer/understeer) lesson.

GAMIFICATION METHODOLOGY FOR QUANTIFYING DRIVER PERFORMANCE

The GBL drive assignment was to complete as many "legal laps" as possible given a 2-minute (120 second) experimental timeframe. A legal lap is defined as one where no barrier cones are struck - the driving task is therefore a critical balance of "going fast" (speed) and "using caution" (accuracy). Upon collecting the results, the instructors determined that the granularity of legal laps - taken alone - was insufficient to assess the overall quality of trainee driving performance. Accordingly, we employed a weighted Gamification system to rate and quantify overall driving performance, and better discern drivers on a more comprehensive excellent/very good/average/fair/poor scale.

The quantitative scoring tabulated by the Simulator ranges from 0 to 100, where 0 is the lowest possible score and 100 represents the highest possible score. A total of eight essential factors are taken into consideration when evaluating the holistic performance of an experimental drive: (1) number of total laps attempted, (2) number of legal laps achieved (***primary rating category***), (3) total distance travelled (miles), (4) fastest legal lap (mph), (5) average travel speed (mph), (6) maximum travel speed (mph), (7) number of crashes, and (8) total number of cone strikes. Each of these factors are weighted using the scoring metric based explicitly on this particular simulator training assignment. The total weighted sum provides a more holistic final score. As there are many data collected during each excursion, a careful pre-evaluation was performed - based largely on expert driver data - to heuristically determine the optimum weighting components for our current Gamification system. Table 1 illustrates the final weightings implemented.

Table 1: Weighted Quantitative Scoring Table Evaluation

Total Laps	Pts.	Legal Laps	Pts.	Total Distance (miles)	Pts.	Fastest Legal Lap (s)	Pts.	Average Speed (mph)	Pts.	Max Speed (mph)	Pts.	Spinout	Pts.	Cone Strikes	Pts.
5	**100**	5	**100**	>= 2.2	**100**	<= 20	***100***	>= 71	***100***	>= 83	**100**	0	**100**	0	**100**
				>= 2.1	***90***	<= 22	***90***	>= 68	***90***	>= 78	**90**				
4	**80**	4	**80**	>= 2.0	**80**	<= 24	**80**	>= 65	**80**	>= 73	**80**				
				>= 1.9	***70***	<= 26	***70***	>= 62	***70***	>= 68	**70**				
3	**60**	3	**60**	>= 1.8	**60**	<= 28	**60**	>= 59	**60**	>= 63	**60**				
				>= 1.7	***50***	<= 30	***50***	>= 56	***50***	>= 58	**50**				
2	**40**	2	**40**	>= 1.6	**40**	<= 32	***40***	>= 53	***40***	>= 53	**40**	>= 1	**0**	>= 1	**0**
				>= 1.5	***30***	<= 34	**30**	>= 50	**30**	>= 48	**30**				
1	**20**	1	**20**	>= 1.4	**20**	<= 36	***20***	>= 47	***20***	>= 43	**20**				
				>= 1.3	***10***	<= 38	**10**	>= 44	**10**	>= 38	**10**				
0	**0**	0	**0**	> 0.0	**0**	*> 40*	***0***	< 44	***0***	< 40	**0**				
Weight	**0.1**	Weight	**0.4**	Weight	**0.075**	Weight	**0.1**	Weight	**0.025**	Weight	**0.05**	Weight	**0.2**	Weight	**0.05**

PRELIMINARY RESULTS AND CORRELATES

In this section, we present an overview of our preliminary results of the implementation of our Gamification-centric methodology towards improving learning outcomes in engineering education. This section is decomposed into three primary subsections. First, we will discuss quantitative results collected from the driving simulator that provide a Gamified assessment of driving performance during the Triple Curve training. Next, we will investigate lesson-pertinent survey results collected in a self-report format, and based on peer-reviewed (and validated) mechanisms informed by the literature. Finally, we conclude this section by investigating correlations between these two data types, in hopes of identifying patterns/tendencies among our cohort participants. Prior to presenting our results, we offer a brief paragraph outlining the limitations of our data reporting, and potential confounding factors.

Study Limitations

The *Study Limitations* of our experimental design include previously-noted assumptions associated with our vehicle dynamics model (e.g., simplified tire behavior), which could impact simulated driver performance. Likewise, we must take in to account the previously-noted limitations of the Gamification model (i.e., designed such that it is difficult to obtain a "high" score, particularly for first-time or inexperienced simulator participants). Finally, we recognize the previously-noted limitations of the male-dominated cohort demographics (i.e., students weren't recruited for this data collection, and therefore, the authors had to utilize data from whoever enrolled in the three courses/clubs). *Potential confounding factors*: since practical matters associated with experimental implementation did not allow for a period of acclimation (i.e., due to the brief experimental duration, and due to the large cohorts that had to participate during short 50-minute college course timeslots), we were not able to allow study participants to properly adapt to the simulated environment. Such a period of acclimation is customarily successful in reducing potentially confounding factors such as Simulator Adaptation Sickness or unfamiliarity with handling (e.g., input controls; dynamic output response) the simulator vehicle during the actual testing session (Michaels et al., 2017).

Simulator Results - Rating Gamification

In this subsection, we present the quantitative observations from our driving simulator cohorts. We begin by discussing how we collected and analyzed our various data directly from the driving simulator. These data were Gamified to quantify student driver performance upon the Triple Curve raceway.

In Table 2, we summarize the statistics from the Gamification algorithm for the driving simulator scoring. Overall across all three courses, we have an average driving score (out of 100) at 60.76 with an average standard deviation of 17.97. The (overall) lowest score is 13.25, and the (overall) highest score is 88.00. Note that the sub-cohort statistics are comparable.

Table 2: Drive Score Statistics

Course	Mean	Std Dev	Min	Max	n
RVD (n1)	60.40	17.90	13.25	88.00	47
SAE (n2)	61.21	17.94	24.50	87.50	28
STL (n2)	61.32	21.21	30.00	86.00	7
Total (N)	**60.76**	**17.97**	**13.25**	**88.00**	**82**

As a mechanism to pre-validate our Gamification scale, and to illustrate baseline feasibility of our chosen weighting scale, we performed calculations based on a series of "idealized" datasets generated by an experienced driver. Although the theoretical fastest legal lap time is unknown, we speculate that a lap of 20.0 seconds is achievable. The "expert" drive score was posted by a seasoned simulator driver whose composite rating score (99.00) is, by design, close to the maximum achievable Game Rating. Based on the (cohort n_1) RVD mean and standard deviation scores, five diverse sets of driving performance data were selected, including the very best (*excellent*; 88.00) and worst (*poor*; 13.25) among the trainees. We note that this approach skewed the cohort results to be lower than might be expected on a traditional bell curve (i.e., a cohort mean score of ~60 rather than a traditional 70; a cohort maximum score of less than 90), but we accepted these limitations as acceptable to demonstrate the feasibility of our methodology.

Likewise, for purposes of demonstration, and among the same (n_1) cohort, we selected data from an "average" score (60.75) achieved from a participant that was close to the calculated cohort mean (60.40; see Table 2), as well as a positive standard deviation for an "above average" (*good*; 81.50) student score and a negative standard deviation for a "below average" (*fair*; 43.25) student score. Table 3 shows the individual statistics of the driving performance and weighted scores for each selected example drive, and this presentation enables analysis of Gamified scoring contributor's side-by-side across a range of simulator drivers.

Table 3: Five (5) Drive Performance Selected Based on Score

Drive	Total Laps	Pts.	Legal Laps	Pts.	Total Distance (miles)	Pts.	Fastest Legal Lap (s)	Pts.	Average Speed (mph)	Pts.	Max Speed (mph)	Pts.	Spinouts	Pts.	Cone Strikes	Pts.	Score
Poor	3	**60**	0	**0**	1.59	**30**	0	**0**	47.26	**20**	78.16	**90**	2	**0**	9	**0**	**13.25**
Fair	2	**40**	2	**40**	1.37	**10**	40.02	**0**	41.22	**0**	60.76	**50**	0	**100**	3	**0**	**43.25**
Average	3	**60**	3	**60**	1.47	**20**	36.35	**10**	44.09	**10**	64.06	**60**	0	**100**	0	**100**	**60.75**
Good	4	**80**	4	**80**	1.95	**70**	26.4	**60**	58.41	**50**	74.06	**80**	0	**100**	0	**100**	**81.50**
Excellent	4	**80**	4	**80**	2.37	**100**	22.22	**80**	71.1	**100**	86.52	**100**	0	**100**	0	**100**	**88.00**
Expert	5	**100**	5	**100**	2.48	**100**	21.55	**90**	74.34	**100**	86.71	**100**	0	**100**	0	**100**	**99.00**

Based directly on the scoring categories determined in Table 3, Figure 6 presents a frequency spectrum for the entire n_1=47 RVD cohort (with the resulting exponential data trend line shown across the five columns) of cohort driving performance for our chosen Gamification (rating) scale. This plot can be interpreted in such a manner that many participants (n=29/47; the rightmost two columns) were rated above the calculated average threshold, and a vast majority of the class cohort (n=41/47; the rightmost three columns) were rated at/near a threshold that was at or above the calculated average threshold.

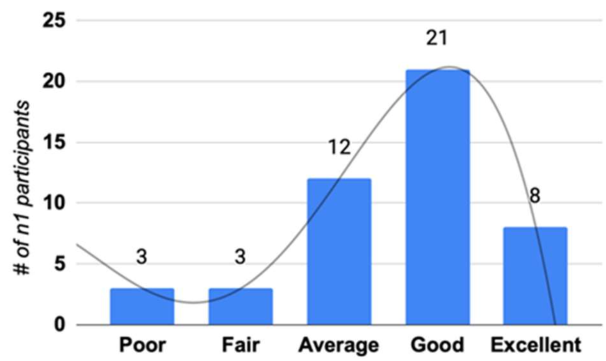

Figure 6: Drive Performance Categories

Survey Results - MSLQ (*n_1=35*) and CQ (*n_2=47*)

In this subsection, we discuss the self-report data that was collected by each cohort to supplement observations from the simulator (quantitative) data discussed previously. Note that the assignments of the sub cohorts are not random; but rather, they were dictated by the specific training objectives of the engineering course/club within which our experiment was ultimately deployed. In other words, sub cohort n_1 (RVD) answered the MSLQ questionnaire, and sub cohort n_2 (STL and SAE) responded to the pre- and post- CQ.

Recall that the MSLQ is a key metric to gauge student engagement through our GBL implementation; motivation and positive attitudes towards relevant course content is a critical factor towards achieving overall success in training performance. In Figure 7, we observe the self-reported MSLQ statistics for the n_1 cohort. Based on the seven-point Likert scale, *Self-Motivation* (Part A) saw a self-reported mean value of 5.54 ± 0.88, while *Learning Strategies* (Part B) saw a self-reported mean value of 4.78 ± 0.82.

Notionally, we interpret these results as meaning that cohort students rated a moderate preference towards goals/values/beliefs over learning content/style with relation to instructional presentation. Furthermore, it may be worth noting that based on the self-reported MSLQ questionnaire data, it was observed that of the n_1=35 cohort, n = 29 (82.9%) are more inclined as self-motivated learners (i.e., their individual Part A scores were self-reported as higher than their Part B scores) and the remainder n = 6 (17.1%) are more inclined towards learning strategies and overall learner approach when absorbing new content.

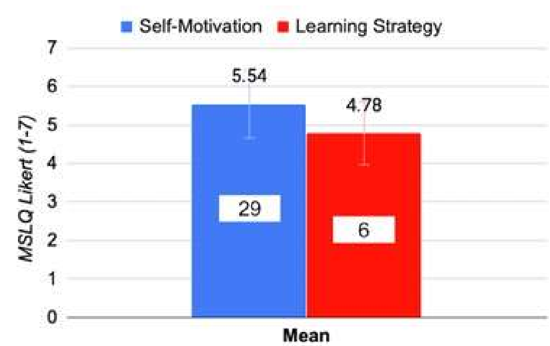

Figure 7: MSLQ Statistics (n_1)

Similarly, the CQ (pre- and post- experiment) serves to assess concept reinforcement through GBL. In our implementation, the CQ was issued both pre-experiment and post-experiment. Figure 8 illustrates response frequencies for each question before and after the GBL drive, where the selected answer either remained correct (shown in yellow), remained wrong (shown in green), modified for the better (shown in blue), or, modified for the worse (shown in red). It is promising to observe that for three of the four CQ questions (Q1, Q2, and Q3), many respondents either remained correct (pre- and post-), or, if they changed their answer, a majority modified for the better (i.e., the blue region is larger than the red region).

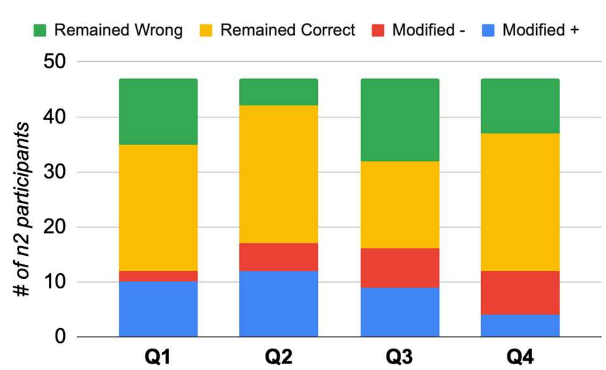

Figure 8: CQ Statistics (n_2)

Furthermore, we also investigated the overall accuracy of the responses; 46.8% (22 participants) had an improved post-CQ score, and 76.6% (36 participants) were compelled to modify at least one answer after the GBL drive, as shown in Figure 9. Our optimistic takeaway is that with hands-on GBL experience to supplement passive engineering dynamics courseware, the students achieved disciplinary knowledge improvement and an enhanced understanding of the lesson's primary concepts. The GBL-exposure was influential enough to sway most participants to augment their response (for better or worse) post-experiment based on their initial responses.

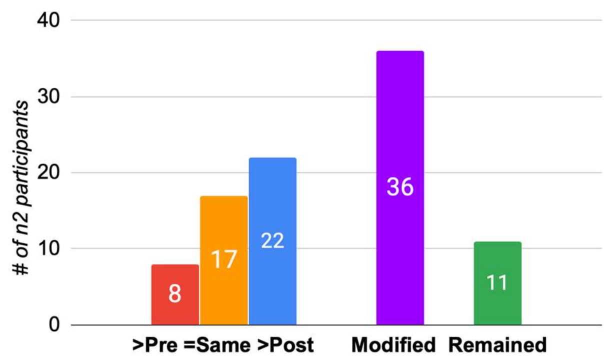

Figure 9: CQ response frequency (n_2)

Observed Correlates

To conclude our analysis, we correlated the drive performance (quantitative) and the assessments (self-report) for each of the two sub cohorts (n_1 and n_2). The sigma was calculated based on the mean and standard deviation of cohort drive scores for n_1 (60.40 ± 17.9) and cohort n_2 (61.24 ± 18.3) respectively.

For the n_1 cohort, we measured correlation between simulator score rating and self-reported "pairing" per the two primary components (***A: Motivation; B: Learning Strategies***) of the MSLQ. To determine each student pairing, we observed their individual scores relative to the cohort mean. If their self-report score was more than one standard deviation above the cohort mean, they were rated A+/B+ respectively; if more than one standard deviation less, A-/B- respectively; otherwise, they were classified as just A/B. The resulting Figure 10 displays a bubble graph of the nine (9) different probable preferential learners from Part A and B. In the Figure, observe that the Y-axis is the simulator score mean for that portion of the cohort, and the relative size of each bubble represents the simulator score standard deviation for that portion of the cohort, while the X-axis represents the total number of participants/observations representative of that A/B pairing. Note that eight of the nine total combinations were represented among our n_1=35 cohort, and as expected, the highest frequency pairing is the A/B concentration (n=20/35) who exhibited a mean simulator score of 65.33 ± 16.01. These learners were slightly more than 5 points above the overall cohort average (solid red line). The authors hypothesized that the A+/B+ concentration might be prime candidates for being strong performers in a GBL-based simulator environment. Indeed, there were (n=4/35) participants among that self-reported concentration, with a mean simulator score of 67.13 ± 14.83; nearly 7 full points above the cohort average. Likewise, the "B+" learners were particularly high performers; the aforementioned (A+/B+) concentration ranked third overall, and the (A-/B+) concentration was highest, with however only one student who self-reported that pairing.

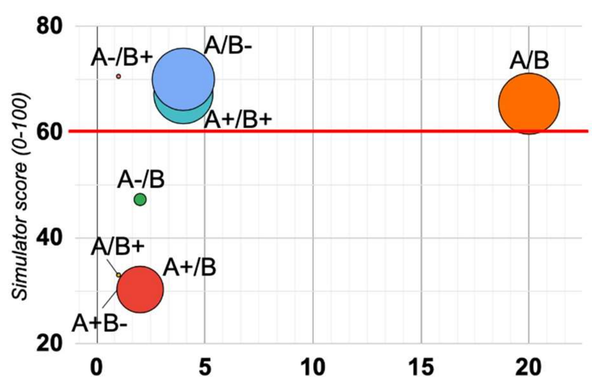

Figure 10: Simulator-MSLQ Correlation (n_1)

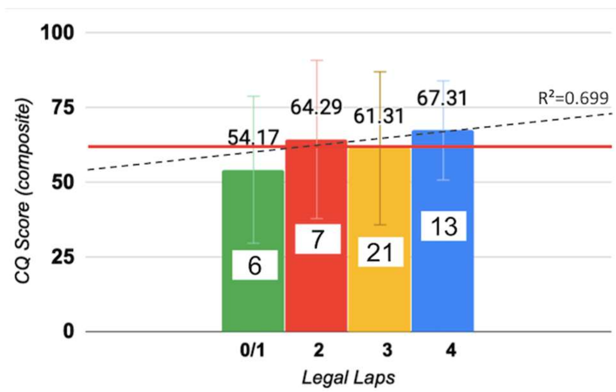

Figure 11: Simulator-CQ correlation (n_2)

For the n_2 cohort, we measured correlation between simulator score rating and pre- and post- performance on a discipline-specific conceptual quiz (CQ). Our instructional objective was to determine if measured student performance on the GBL course implementation would correlate with academic performance relative to underlying theory. The resulting Figure 11 displays the composite CQ score (pre-score plus post-score, normalized to a 100-

scale) versus the number of Legal Laps obtained during the experimental drive. Each X-axis series indicates mean, standard deviation (as error bars), and textually-embedded are the frequencies of drivers that fall in to each category (and sum to 47) – to comprise the (n_2) cohort total. The cohort average (composite) CQ score was approximately 61.25, shown as a solid red line. From these data, numerous preliminary observations can be made. There is a noteworthy (R^2 ~0.7) trend that correlates academic performance (CQ) and GBL-experimental objective. Those students (n=6) that did not perform well within the Triple Curve (0 or 1 legal laps) also scored lowest on the pre- and post-CQ (*mean: 54.17 ± 24.5*). Simultaneously, those students (n=13) who performed the best (achieving 4 legal laps) also scored the highest on the pre- and post-CQ (*mean: 67.31 ± 16.6*). The intermediate performers on The Triple Curve (i.e., either 2 or 3 laps) also displayed moderate academic performance on the CQ, with means of 64.29 (± 26.4) and 61.31 (± 25.6), respectively. Perhaps these general trends warrant future downstream analyses that would be enhanced by: a) a larger overall cohort size, b) a more substantive experimental duration on the simulator, and c) a more comprehensive granularity of GBL scoring subcomponents than those demonstrated in this work.

SUMMARY AND PRIMARY CONCLUSIONS

In this paper, we described the design and pilot deployment of an innovative Gamification-based experiential learning andragogy for Transportation Engineering education. Primary novelties included an intended emphasis on Gamification and Game-based Learning (GBL), areas which continue to emerge in young adult education and training, and have shown to couple the novelties of entertainment into an innovative, engaging, and experiential context for knowledge and skill acquisition. Our presentation has included: an experimental pilot deployment within two graduate-level engineering courses as well as a targeted exercise within a discipline-targeted engineering club, and preliminary results culled from multiple sources of data, including quantitative (i.e., collected in real-time from the simulator), literature-established and lesson-appropriate surveys (i.e., Likert-style self-report), and statistical correlations between these related data sources. Our innovative and emergent methodology of incorporating GBL serves to better prepare our next-generation of students and warfighters. Furthermore, our implementation promotes the application of M&S to support engagement, repeatability in task execution, development of cognition skills, and a "hands-on" teaming atmosphere, all of which serve to enhance current and future workforce development. An executive summary of our primary outcomes is provided below:

1) Based on targeted quantitative metrics collected from our high-fidelity driving simulator, we developed, verified, and deployed a Gamification-based rating system to holistically quantify driver performance within the Triple Curve GBL environment. The weighted-sums algorithm accounted for the primary experimental objective (number of legal laps), as well as ancillary performance metrics.
2) Using (5) representative datasets from the n_1 cohort, we decomposed the entire cohort into simulator performance rating categories, with a majority of the class cohort performing at an average, or better threshold. These observations serve to provide baseline validity to our model.
3) The MSLQ was issued (pre- GBL experiment) to allow students to self-report preferences related to student engagement, decomposed into A. Motivation and B. Learning Strategies. Our results indicate a cohort average that was moderately higher for Part A (Motivation), with nearly 83% of students reporting higher (average) ratings for A relative to B survey categories.
4) The CQ was issued (both pre- and post- GBL experiment) to gauge student conceptual knowledge. Our results indicate that for three of the four CQ questions, if students changed their answer, a majority modified for the better. Overall, 76.6% (36 participants) were compelled to modify at least one answer after the drive, and as a result, 46.8% (22 participants) had an improved post-CQ score.
5) We investigated correlation between measured simulator scores and self-reported MSLQ. We decomposed all students into pairwise combinations of A/B self-reported preference, based on "strength" of responses. We observed numerous trends, including above average simulator scores from the pairing (A/B) that had the highest frequency, as well as particularly strong simulator scores from those among the "B+" cohort.
6) We investigated correlation between measured simulator scores and relative demonstrated performance (improvement) between pre- and post-CQ. We observed a general correlation between academic performance on the pre- and post- CQ and the number of Legal Laps achieved within the Triple Curve game, recalling that as being the primary experimental objective.
7) Overall impact of result correlations in future implementations could be improved with larger (and more gender-diverse) cohort sizes, and GBL experiments with greater depth/duration and Gamification granularity for scoring/rewards.

As previously suggested, the current work suitably serves as a guidepost towards widening the scope of Gamifying engineering education and training across a number of other candidate M&S disciplines, both civilian (i.e., beyond solely Transportation), as well as military training applications. Notably, our aspirations for this work are to influence a broader adoption of gaming experiences in multidisciplinary engineering course curricula. Specifically, incorporation of physics-based modeling and high-fidelity simulation to enable real-time observation of downstream impacts (cause/effect). To this end - in the final section of this paper, we formalize an outline of a logic model for planned best practices - to authenticate program functionality and investigate long-term sustainability - on a much more comprehensive scale.

FUTURE WORK: PROPOSED "THEORY OF CHANGE" FOR GBL IN ENGINEERING EDUCATION

Historically, games are not automatically effective in educational settings; they must be rooted in effective andragogical practices. In other words, the implementation of "surface characteristics" of gamification (e.g., badges and rewards), without employing rigorous game design elements, can actually inhibit (rather than enhance) student engagement (Deterding, 2012). The innovative research methodology and preliminary findings presented in this paper are intended to demonstrate a foundation to revolutionize the form/function of content delivery within future engineering education. Our strategic vision is to conceive, develop, and deploy a wide series of GBL-courseware, leveraging advanced modeling & simulation (M&S) and emergent Gamification approaches to improve learner engagement, training effectiveness, and success rates for vocational preparation. Currently in development is a revised engineering education taxonomy (courseware and assessment standards), that will rigorously incorporate game design characteristics that have been consistently associated with positive learning outcomes (Stott & Neustaedter, 2013).

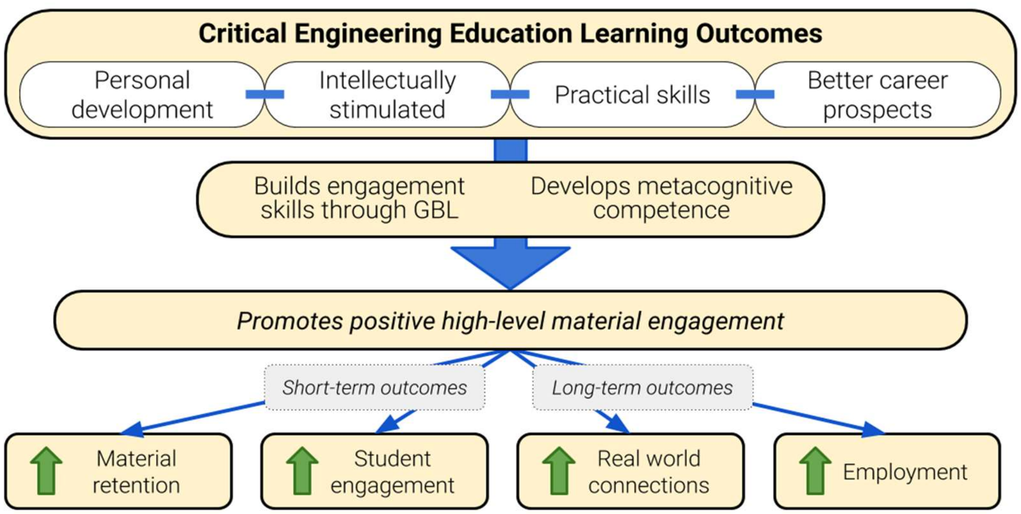

Figure 12: Proposed GBL-centric "Theory of Change" for transforming engineering education

Refer to Figure 12, which illustrates the primary (notional) components of our proposed Logic Model (i.e., a "Theory of Change") for using Gaming to improve engineering education - whose primary goals are to develop extensible principles that will engage critical thought processes for solving "real world" problems. This model explicitly incorporates elements of GBL to enhance self-confidence, independence, pushing boundaries (i.e., outside-the-box thought processes) and critical thinking. Our model is explicitly designed to build specific skills in engaging GBL activities and metacognitive competence - to increase active learner participation, and promote positive engagement with instructional materials. We hypothesize that through the execution of these model elements, we will enable an optimal atmosphere to enhance and improve short-term and long-term learning outcomes.

ACKNOWLEDGEMENTS

The authors wish to extend thanks to Moog, Inc. (East Aurora, NY), for generously donating the 2000E motion simulation platform to the School of Engineering and Applied Sciences (SEAS) at the University at Buffalo, and for their ongoing technical and fiscal support. The authors are likewise grateful for the ongoing support of the University at Buffalo Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL).

REFERENCES

- Aldrich, C., (2005). Learning by Doing: A Comprehensive Guide to Simulations, Computer Games and Pedagogy in E-learning and Other Educational Experiences. San Diego: Pfeiffer. ISBN 978-0-7879-7735-1.
- Anusha, S., & Reddy, N.S., (2015). Social Learning in Engineering Education, *Journal of Engineering Education Transformations*. 292. https://doi.org/10.16920/ijerit/2015/v0i0/59749.
- Artino, A.R. (2005). Review of the Motivated Strategies for Learning questionnaire. *Retrieved from*: http://www.eric.ed.gov/PDFS/ED499083.pdf.
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R., (2013). Assessment in and of Serious Games: An Overview, *Advances in Human-Computer Interaction*, (special issue: User Assessment in Serious Games and Technology-Enhanced Learning), Volume 2013, https://doi.org/10.1155/2013/136864.
- Cordellieri, Pierluigi & Baralla, Francesca & Ferlazzo, Fabio & Sgalla, Roberto & Piccardi, Laura & Giannini, Anna maria. (2016). Gender Effects in Young Road Users on Road Safety Attitudes, Behaviors and Risk Perception. Frontiers in Psychology. 7. https://doi.org/10.3389/fpsyg.2016.01412
- Dale, E., (1969). Audio-Visual Methods in Teaching, 3rd ed., Holt, Rinehart & Winston, New York, 1969.
- Department of Defense (DoD), (1998). Modeling and Simulation (M&S) Glossary, DoD 5000.59-M, January
- Deterding, S., (2012). Gamification: designing for motivation, *Interactions* 19.4: 14-17.
- Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, 30(3), 159-166.
- Hughes, A., (2017). Gamification Versus Serious Games, (*online article*), https://trainingindustry.com/articles/learning-technologies/gamification-versus-serious-games/.
- Hulme, K.F., Androutselis, T., Eker, U., & Anastasopoulos, P., (2016). A Game-based Modeling and Simulation Environment to Examine the Dangers of Task-Unrelated Thought While Driving", MODSIM World Conference, Virginia Beach, VA, April, 2016.
- Hulme, K.F., Estes, E., Schiferle, M., & Lim, R., (2019). Game-based Learning to Enhance Post-secondary Engineering Training Effectiveness, The Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) (*Training subcommittee*), Orlando, FL, December, 2019.
- Hulme, K.F., Rachel Su Ann Lim, Meghan Bauer, Nailah Hatten, Helena Destro, Brenden Switzer, Jodie-Ann Dequesnay, Rebecca Cashmore, Ian Duncan Jr., Anand Abraham, Jacob Deutsch, Nichaela Bald, Gregory A Fabiano, & Kemper E Lewis, (2021a). "Blueprint for a Simulation Framework to increase driver training safety in North America: Case Study", *Safety* (***accepted, publication pending***).
- Hulme, K.F., Mark Schiferle, Rachel Su Ann Lim, Aaron Estes, & Matthias Schmid, (2021b). "Incorporation of Modeling, Simulation, and Game-Based Learning in Engineering Dynamics Education to enhance Learning Outcomes towards improving Vehicle Design and Driver Safety", *Safety* (***accepted, publication pending***).
- Kolb, D., (1984). *Experiential Learning as the Source of Learning and Development*. Prentice Hall, NJ.
- Markopoulos, A., Fragkou, A., Kasidiaris, P., and Davim, J. (2015). Gamification in engineering education and professional training. *International Journal of Mechanical Engineering Education*. 43. 118-131. https://doi.org/10.1177/0306419015591324.
- McFadden, C., (2017). 9 Everyday Skills Video Games Can Improve, Interesting Engineering, (*online article*), https://interestingengineering.com/9-everyday-skills-video-games-can-improve, November 15, 2017.
- Michaels J, Chaumillon R, Nguyen-Tri D, Watanabe D, Hirsch P, Bellavance F, Giraudet G, Bernardin D, Faubert J. Driving simulator scenarios and measures to faithfully evaluate risky driving behavior: A comparative study of different driver age groups. PLoS One. https://doi.org/10.1371/journal.pone.0185909
- Milliken, W. F. & Milliken, D. L., (1995). Race Car Vehicle Dynamics, SAE, 1995.
- National Training and Simulation Association (NTSA), (2019). US Army: The Future is Now, I/ITSEC Showdaily, https://www.iitsec.org/-/media/sites/iitsec/showdaily/18iitsec_daily_day4/showdaily19_day4.ashx
- Pintrich, P.R., & De Groot, E.V., (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance, *Journal of Educational Psychology*, Vol. 82, No. 1, p. 33-40.
- Prince, M.J., & Felder, R.M., (2006). Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases, *Journal of Engineering Education*, Volume 95, Issue2, pp. 123-138.
- Raybourn, E., & Waern, A. (2004). Social learning through gaming. Conference on Human Factors in Computing Systems - Proceedings. 1733-1734. https://doi.org/10.1145/985921.986206.
- Rieber, L.P., Smith, L., Noah, D., (1998). The Value of Serious Play, *Educational Technology*, (38), pp 29-37.
- Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R., (2018). Using concept inventories to measure understanding, *Higher Education Pedagogies*, 3:1, 173-182, https://doi.org/10.1080/23752696.2018.1433546.
- Stott, A., & Neustaedter, C., (2013). Analysis of gamification in education, Surrey, BC, Canada 8: 36.
- Tolk, A. (2012). *Engineering Principles of Combat Modeling and Distributed Simulation*: John Wiley & Sons.