Implementation of Game-based M&S Tools to Enhance K9-12 STEM Learning Effectiveness

Kevin F. Hulme, Ph.D., CMSP; Qian Wang, Ph.D.; Gongda Yu; Aaron Estes, Ph.D.; Irina Benedyk, Ph.D. Presentacion Rivera-Reyes, Ph.D.; Vidhi Solanki; Emeric Humbert
University at Buffalo / The Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL)
Buffalo, NY 14260
hulme@buffalo.edu; qw6@buffalo.edu; gongdayu@buffalo.edu; aaronest@buffalo.edu; birina@buffalo.edu; privera1@buffalo.edu; solankiv@buffalo.edu; emerichu@buffalo.edu

ABSTRACT

The mandate to measurably improve learning effectiveness remains an ongoing challenge within Education. Despite numerous limitations, teacher-centered didactic instruction remains the most prevalent mode of instruction in K9-12 Science, Technology, Engineering, and Mathematics (STEM) courses. In these settings, teachers promote a passive classroom environment where students are less engaged and less motivated to learn. For many students, such docile training environments frequently result in substandard learning outcomes. Particularly, for underrepresented minority (URM) students, passive learning environments are not effective at addressing performance gaps. Accordingly, educators have been exploring the application of instructional best practices for *active* learning - including advanced Modeling & Simulation (M&S) tools to cultivate outcomes within both individual and team-based settings.

The effectiveness of Game-based Learning (GBL), physics-based modeling and high-fidelity driving simulation to convey critical dynamics principals to a high-school STEM cohort was evaluated. Our implementation included hands-on analysis of two key vehicle parameters: 1) front-to-rear weight distribution, and 2) front-to-rear tire stiffness distribution. The experiment served as a primary instructional component for the National Summer Transportation Institute (NSTI), sponsored by the Department of Transportation (DOT) during the summer of 2023 at the University at Buffalo. Experimental content was deployed upon two additional college-age student cohorts: CSTEP (Collegiate Science and Technology Entry Program) and WiSE (Women in Science and Engineering), respectively. To supplement quantitative simulator data, questionnaires and self-report surveys were issued pre- and post-experiment to gauge conceptual understanding and to better understand individual learning preferences.

Primary outcomes of our experimental deployment include: i) nuanced performance increases on the simulator (i.e., depending on vehicle type); ii) cohort improvement on the conceptual quiz (comparing pre- to post-); iii) comparable cohort preferences between theory/simulator implementations; and iv) mild correlations between simulator performance and simulator preference, which could reveal the potential benefits of a simulator-centric pedagogy in amplifying driving performance.

ABOUT THE AUTHORS

Kevin F. Hulme, Ph.D., CMSP received his Ph.D. from the University at Buffalo, specializing in multidisciplinary analysis and optimization of complex systems. He is the Program Manager for The Stephen Still Institute for Sustainable Transportation and Logistics and provides oversight for its Motion Simulation Laboratory. His current areas of focus include applied modeling and simulation (M&S), simulation science, transportation and vehicle simulation, game-based approaches in engineering education, human factors, and advanced air mobility. Dr. Hulme is a Certified Modeling and Simulation Professional (CMSP).

Qian Wang, Ph.D. received her Ph.D. from RPI (Troy, NY) in Transportation Engineering, and is an Associate Professor of Teaching in Civil, Structural, and Environmental Engineering at the University at Buffalo. She instructs classes related to transportation systems analysis, travel demand forecasting, and transportation network analysis.

Gongda Yu is a third-year Ph.D. candidate in Transportation Systems Engineering at the University at Buffalo. His research interests encompass travel behavior under extreme weather conditions, human factors, intelligent information systems, and driving phobia.

Aaron Estes, Ph.D. is a Teaching Assistant Professor with the Department of Mechanical and Aerospace Engineering at the University at Buffalo. His teaching interests include vehicle dynamics and control, and system identification, and he has research and instructional experience with Game-Based Learning (both physical and digital).

Irina Benedyk, Ph.D. is an Assistant Professor with the Department of Civil, Structural, and Environmental Engineering at the University at Buffalo. Her areas of focus include travel behavior, human factors in transportation engineering, driver cognition, sustainable transportation, and engineering 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.

Vidhi Rasik Solanki is a PhD student in the department of Civil, Structural, and Environmental Engineering at the University at Buffalo. Her work focuses on system identification, seismic response prediction, and control system development for model-in-the-loop frameworks.

Emeric Humbert received his M.S. in Geography in France (Universite de Lorraine), and he recently received his M.S. at the University of Buffalo in the Sustainable Transportation and Logistics (STL) program. He is currently an Engineer trainee with the New York State Department of Transportation.

Implementation of Game-based M&S Tools to Enhance K9-12 STEM Learning Effectiveness

Kevin F. Hulme, Ph.D., CMSP; Qian Wang, Ph.D.; Gongda Yu; Aaron Estes, Ph.D.; Irina Benedyk, Ph.D.
Presentacion Rivera-Reyes, Ph.D.; Vidhi Solanki; Emeric Humbert
University at Buffalo / The Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL)
Buffalo, NY 14260
hulme@buffalo.edu; qw6@buffalo.edu; gongdayu@buffalo.edu; aaronest@buffalo.edu; birina@buffalo.edu; privera1@buffalo.edu; solankiv@buffalo.edu; emerichu@buffalo.edu

INTRODUCTION AND BROADER IMPACTS

The mandate to measurably improve Science, Technology, Engineering, and Mathematics (STEM) learning effectiveness remains an ongoing challenge within Education, and recent performance gaps suggest the ongoing need for curriculum re-evaluation. Didactic instruction (i.e., teachers giving lessons to students) remains the most prevalent mode of teaching in STEM courses (Ha et al., 2023). However, the result is a passive classroom environment where students are less motivated and less engaged, which often leads to substandard learning outcomes (Sandoval et al., 2022). This dilemma has been amplified for underrepresented minority (URM) students, where passive learning environments are not effective at addressing performance gaps (Matthews et al., 2020).

Accordingly, educators in STEM fields have been investigating more active instructional practices (Nguyen et al., 2021), where students learn experientially (Staehle et al., 2023), by doing rather than by observing (Nadelson et al., 2019). Active learning can enable meaningful connections to the real-world (Ješková et al., 2022), and implemented practices have demonstrated substantial benefit, including improved examination scores and passing rates (Ballen et al., 2017). A specific construct of active learning that has been rapidly gaining interest in high school STEM (Wang et al., 2022) is Game-based Learning (GBL); a form of (digital or non-digital) gameplay to target specific learning outcomes (Jääskä et al., 2022). A related concept within GBL – "gamification" – is a framework that often involves ranking or rating a learner for a task (Balci et al., 2022). GBL implementations focus less on content (Fromm et al., 2021), and are most effective when spatial/temporal decisions are required simultaneously (Reynaldo et al., 2021).

In this study, we evaluated the effectiveness of Game-based Learning (GBL), physics-based modeling and high-fidelity driving simulation to convey critical vehicle design and motion dynamics principals to a high-school STEM cohort. The experiment served as a primary instructional component for the DOT-sponsored National Summer Transportation Institute (NSTI). Quantitative simulator data, self-report questionnaires and brief surveys were issued to gauge conceptual understanding and learning preferences. Our implementation addresses numerous priorities from the Education subcommittee, including instruction strategies/best practices, STEM implementations to accelerate learning, and training strategies staged within individual and team-based settings.

Forecasted Broader Impacts to the M&S and military training communities are numerous. By incorporating game elements, traditional training programs become more engaging and enjoyable, which often leads to elevated participation and performance levels and other training outcomes among soldiers and warfighters. (Warfighter Digital, 2023). Likewise, our methodology suggests conceptual parallels to other I/ITSEC subcommittees, including GBL implementations to improve driver education (*Training/PSMA*), novel gaming approaches to advance the science of learning (*Education/ECIT*), and applied M&S to improve and enhance driver/vehicle safety (*Simulation/HPAE*).

EXPERIMENTAL DESIGN AND METHODOLOGY

Our primary objective is to assess the utility of GBL and motion-based vehicle simulation as a critical tool for STEM (9-12) training and education. The experiment has been designed such that active learning occurs for both the driver and passenger of the driving scenarios, and for their class peers observing the simulations (offboard) in a large group format upon a large projection screen. Targeted data collected before, during, and after the simulator experiment help confirm our central hypothesis that applied M&S can enhance conceptual understanding of vehicle performance.

Vehicle Dynamics background theory

The key concepts of ***understeer and oversteer*** are used to describe the overall sensitivity of a ground vehicle to steering input. They are defined by an "understeer gradient" that serves as a measure of how the steering required for a constant turn varies as a function of lateral acceleration. Steering at a steady speed is compared to the steering that would be required to follow the same circular path at low speeds. The low-speed steering for a given turning radius is referred to as Ackermann steer (e.g., Milliken & Milliken, 1995). The vehicle is understeer (i.e., has a positive understeer gradient) if the difference between required steer and Ackermann steer increases with respect to incremental increases in lateral acceleration. Alternatively, the vehicle is oversteer (i.e., has a negative understeer gradient) if the difference in steer decreases with respect to incremental increases in lateral acceleration. If the understeer gradient is zero, the vehicle is classified as neutral steer. Refer to Figure 1, which compares these three conditions.

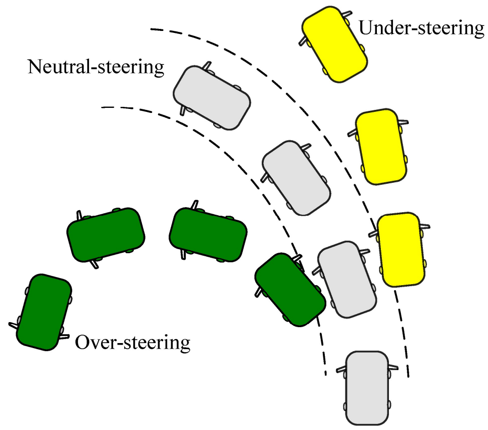

Figure 1 – Steering responses within a curve (Yin et al., 2017)

Defined more plainly, understeer occurs when a car steers less than the driver's input; oversteer occurs when a car steers more than the driver's input; and neutral steer is when the vehicle output closely matches the commanded input and provides a balance of stability and maneuverability. **Our experiment sought to employ hands-on GBL for discovery of oversteer/understeer and the two primary variables that influence its impact**.

1) **Weight distribution (center-of-gravity - CG)** dictates how much of the vehicle's weight distribution is toward the front tires. A front-heavy vehicle is more stable (i.e., understeer), easier to drive, but also less maneuverable. For reference – a typical sedan might have a 60% front (40% rear) weight distribution.
2) **Tire stiffness distribution (TSD)**, otherwise known as cornering stiffness, determines the vehicle's response on a turn. A vehicle with stiffer tires towards the front of the vehicle is more stable (i.e., understeer), and is therefore less maneuverable.

Most critical is the key **inter-relationship between these two variables**. The general trend is that a CG that is more towards the front of the vehicle than the TSD results in a more stable (i.e., understeer) configuration. Refer to Figure 2, where the CG marker is shown in yellow/blue, and the TSD marker is shown in black/white. If the (% front) CG is greater than the (% front) TSD, the vehicle will understeer; if the (% front) CG is equal to the (% front) TSD, the vehicle is neutral steer; and if the (% front) CG is less than the (% front) TSD, the vehicle will oversteer.

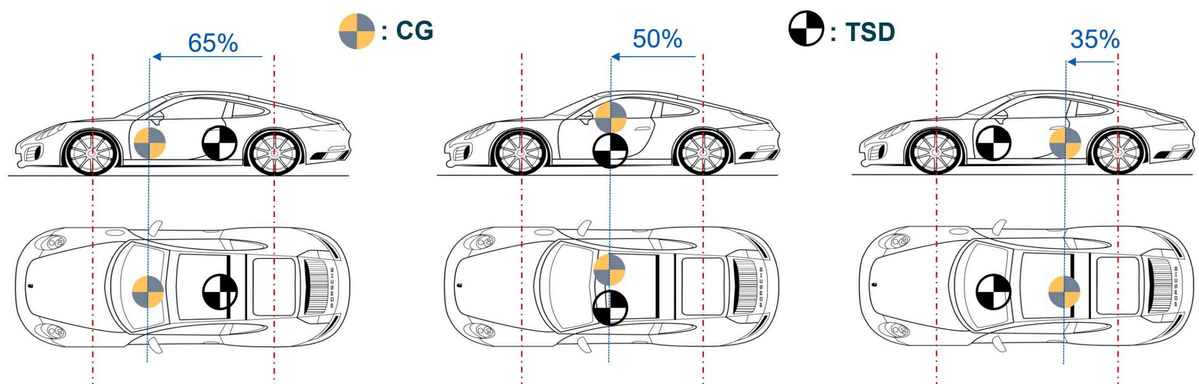

Figure 2 – CG/TSD interrelationship (*left-to-right: understeer | neutral steer | oversteer*)

For the K9-12 experiment that was conducted for this effort, we examined nine vehicle configurations that varied the CG and TSD systematically. Each of the two variables ranged from 35% to 65% front, resulting in the combinations displayed (in sequence) within Table 1. For each combination, the associated steering condition is listed, alongside a brief comment on the expected handling characteristics associated with that coupling of variable settings.

Table 1 – Experimental Grid for simulated vehicle variable settings

Drive #	% front CG	% front TSD	Condition	comment
1	65	35	Understeer	Stable, but underperforming
2	65	50	Understeer	Stable, and more easily driven by novices
3	65	65	Neutral steer	Partially stable; challenging to drive
4	50	35	Understeer	Stable, easily driven by novices
5	50	50	Neutral steer	Partially stable, but easy to drive
6	50	65	Oversteer	Not stable, challenging to drive at elevated speeds
7	35	35	Neutral steer	Partially stable; challenging to drive
8	35	50	Oversteer	Not stable, challenging to drive at elevated speeds
9	35	65	Oversteer	Not stable, incredibly challenging to control

Note that there are three pairwise conditions of each of the three steering conditions: understeer (drives 1, 2, and 4); neutral steer (drives 3, 5, and 7); and oversteer (drives 6, 8, and 9). The expected within-condition variances observed across the understeer and oversteer conditions are easy to explain intuitively. That is, the larger the assigned difference between % front CG and TSD (i.e., positive for understeer; negative for understeer), the more pronounced the observed steering characteristics. Furthermore, elevated stability is to be expected the closer the parameters are towards the front of the vehicle. The latter of which is to say that a 65/50 (drive #2) vehicle has more stability than a 50/35 vehicle (drive #4), even though the CG/TSD differential is +15% across both configurations.

Comparison of the three neutral steer conditions is more nuanced. In these configurations, only the front tire is steered, so when steering input is applied, a lateral force is produced at the front tire that is proportional to the cornering stiffness of the front tire. In the 65/65 case (drive #3), the front tire has massive cornering stiffness, and the CG is right next to the front tire. Conversely, in the 35/35 case (drive #7), the lateral force (and front cornering stiffness) at the front tire is reduced for the same steering input, however, there is a larger moment arm to the CG which is now located by the rear axle. Further technical nuances lie outside the intended scope of this experimental deployment.

Supporting Hardware and Software

The K9-12 GBL experiment was performed on a high-fidelity driving simulator featuring a 360-degree, 16-foot diameter, 6' high, six-channel display system (each HD channel is 1920x1080p, resulting in a 11520x1080p 50Hz composite image). Visual rendering surrounds the vehicle and driver/passenger and provides occupants with a full-surround depiction of the virtual driving environment. The system features a true Ford Contour vehicle cabin (see Figure 3), USB input navigation controls (see Figure 4), 6-DOF motion cueing and washout filtering, as well as a stereo system for aural cues that emulate sounds heard both inside and outside the vehicle during a driving excursion.

Figure 3 – simulator cabin

Figure 4 – control console

Figure 5 – K9-12 GBL Training area

Our experimental setup was designed to support instructional best practices for *active* learning (i.e., using M&S and GBL). Figure 5 shows the multi-purpose student seating area, which enabled viewing of a theoretical presentation, the live simulator itself (i.e., including the 360-degree display screen), and an off-board display screen that displayed the forward cockpit view during the simulator drives. The latter infrastructure enabled classroom trainees (i.e., those outside the simulator vehicle) to continuously observe and learn from their peer performances and engage in the instruction within a more "team-based" experience.

Simulator driving task and scoring mechanism

The racetrack that was used for our K9-12 GBL experiment was designed in-house (using the C++ OpenGL API) and is called The *Quad-radial speedway*. It begins with a short straight segment into Turn #(1), a large swooping left turn; after a brief straight segment, next comes Turn #(2), a sharp 90-degree left turn; this is followed quickly by Turn #(3), a challenging hairpin right turn; this transitions immediately into Turn #(4), a moderate radius left turn, which leads into a long straightaway back to the start line. Refer to Figure 6 where the Quad-radial speedway is illustrated. Each driver was given a vehicle with varying (% front CG, and % front tire stiffness) parameters per Table 1 and were afforded 2.5 minutes on the speedway.

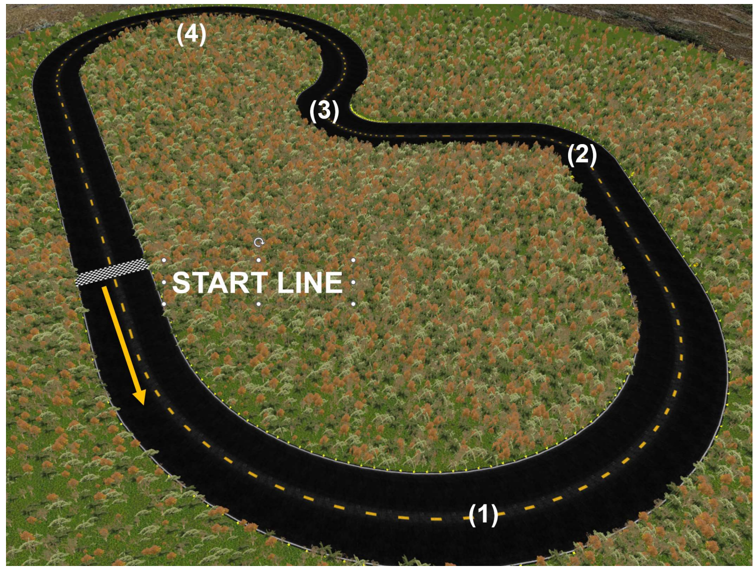

Figure 6 – The Quad-radial speedway

The primary experimental goal was to maximize the number of (legal) laps achieved during this period. Note that a legal lap implies that the driver remains on the asphalt (i.e., without striking any cones that border the speedway), and maintains control of the vehicle to avoid a spinout or crash. Quantitative rating of driver performance was achieved through the implementation of a heuristic weighted sums scoring (“gamification”) model; refer to Table 2. Note that the component weightings are parameterizable, and prioritized the primary driving goals (i.e., number of legal laps and fastest legal lap, measured in seconds) alongside the settings difficulty (i.e., vehicle handling) per each assignment of experiment variables. The readily quantifiable categories (i.e., fastest legal lap, average travel speed, crash metric, and number of legal laps) are rated on an interval (rank-ordered) scale, while the handling difficulty is rated on an ordinal scale.

Table 2 – Scoring model (rating components)

Component	Description	Rating type	Weighting
Legal laps	How many legal laps achieved in allotted time?	Interval (x/9)	30%
Fastest lap	What was the fastest (legal) lap achieved?	Ordinal (x/12)	30%
Handling	How difficult were the assigned handling characteristics?	Ordinal (x/12)	30%
Crash metric	How many critical mistakes were made in allotted time?	Interval (x/9)	5%
Travel speed	How consistent was the overall rate of travel?	Ordinal (x/9)	5%
GAMIFICATION RATING		**TOTAL (x/51)**	**100%**

Conceptual Quiz

To supplement quantitative data acquired from the driving simulator, a brief conceptual quiz was issued. The quiz contained four questions on the underlying theory relating to steering, handling, performance, and the primary variables that influence these characteristics. The multiple-choice questions each had a range of three to five possible responses. The (pre-/post-) quiz was issued to gauge overall improvement of conceptual understanding. Table 3 summarizes the questions from the conceptual quiz and the correct responses. Upon reflection, questions 3 and 4 both had candidate responses that deserved partial (0.5) credit and were rated as such during the final tally.

Table 3 – Conceptual Quiz

Question #	Conceptual Query	Correct response
1	Which vehicle configuration is less maneuverable, but more stable?	Understeer
2	Which vehicle configuration is favored for balanced performance?	Neutral steer
3	Does a more front CG make a car more stable or less stable?	Depends on tires*
4	Do stiffer tires towards the front make a car more stable or less stable?	Depends on CG*

*: partial credit was issued for a response: “more stable,” which is (generally) a true statement.

STUDY COHORT

The K9-12 STEM GBL simulator experience comprised one training module associated with the National Summer Transportation Institute (NSTI) at the University at Buffalo. The week-long NSTI enables high school students to explore the current state-of-practice associated with Transportation Science, and a critical opportunity to investigate careers and educational opportunities in the Transportation industry. Broad support for the NSTI program is provided by the Federal Highway Administration through the Department of Transportation.

Experiment activities and sequence details

The NSTI experience was designed to intersperse theory with simulator exposure in a progressive manner. All students in the cohort were given note sheets and were instructed to take as many (or as few) notes as they chose during the entire experiment and were collected and observed for downstream analysis. We began with a general experiment briefing, which was followed by entrance surveys and a conceptual quiz to gauge preliminary concept comprehension. This was followed by a general introduction to the simulator and a basic theoretical overview of vehicle oversteer/understeer. We then staged simulator drives #1-3, which was followed by conceptual reinforcement (i.e., vehicle weight distribution | CG). This was followed by simulator drives #4-6, and further conceptual reinforcement (i.e., tire stiffness distribution, and interrelationships with the CG). This was then followed by simulator drives #7-9. We concluded with exit surveys, including re-taking the conceptual quiz, a statement of learning preferences, a sickness questionnaire, and a general satisfaction survey. Refer to Table 4 for the experiment grid.

Table 4 – K9-12 GBL simulator experiment grid

Activity	**Description**	**Duration**
Overview/briefing	Description and overall purpose of experiment	5 minutes
Entrance surveys	Conceptual Quiz, basic driver demographics	10 minutes
Simulator introduction	Description of Lab, Simulator operation, Quad-radial track	10 minutes
Conceptual overview	Oversteer, understeer, neutral steer theory	10 minutes
Simulator drives (Part I)	Simulator Drives #1-3 (***see Table 1***)	15 minutes
Conceptual reinforcement (Part A)	Weight Distribution (CG); vehicle performance/stability	10 minutes
Simulator drives (Part II)	Simulator Drives #4-6 (***see Table 1***)	15 minutes
Conceptual reinforcement (Part B)	Tire (cornering) stiffness distribution; relationship to CG	10 minutes
Simulator drives (Part III)	Simulator Drives #7-9 (***see Table 1***)	15 minutes
Exit surveys	Conceptual Quiz, learning preferences, satisfaction survey	10 minutes
Q/A and Conclusion	Lab walkthrough, concluding remarks	10 minutes
TOTAL EXPERIMENT DURATION		**120 minutes**

Content demonstration sub-cohorts

Our experimental content was pre- and post-deployed by two URM college-age cohorts. The former was leveraged <u>as a preliminary deployment of the drive sequence</u> (Table 1) and consisted of undergraduate students from CSTEP (Collegiate Science and Technology Entry Program). Sponsored by the New York State Department of Education, CSTEP provides hands-on research experiences in critical STEM content areas. The latter was used <u>to re-evaluate the interspersion of the simulator drives with theoretical concepts</u> and consisted of undergraduate students from WiSE (Women in Science and Engineering). Sponsored by regional private sector partners, WiSE provides extracurricular research opportunities with a goal of improving success of women in STEM fields. Note that for sake of brevity, and to maintain focus on the NSTI cohort, data from these sub-cohorts has not been included in this dissemination.

RESULTS AND DISCUSSION

In our results presentation, our intention is to determine the educational effectiveness of our implementation – specifically if student learning patterns and behaviors were observable on the simulator. After a brief presentation on our study cohort, we offer quantitative insights on data captured by the driving simulator, including a demonstration of our scoring model. This is followed by a presentation of the survey/self-report/quiz data that was collected, to gauge overall improvement of conceptual understanding, and to better understand individual learning preferences. Finally, the section concludes with a brief discussion of supplemental analyses that attempted to correlate these findings and ascertain how knowledge transfer was reflected on the theoretical takeaways.

Cohort demographics

A total of N=23 participants (16 male, 7 female) were enrolled in the summer 2023 NSTI program. Note that there was no additional selection criteria for this experiment; that is, all students who enrolled for the week-long Institute were automatically recruited for this experiment. This is to say that no preference was given to any students based on gender, driving experience, simulator experience, gaming experience, nor any additional factors that might have unduly influenced the outcome of our deployment.

The average age for the cohort was 15.05 years (σ=1.20), with just over half (N=12) self-reporting some experience operating an actual vehicle. As a component of their entrance survey, we queried video-game experience and risk-taking propensity, both on a 4-point Likert scale. The cohort self-reported as being moderately high (μ=3.41/4.0, σ=0.67) in terms of video game experience (i.e., the talents of which could lend themselves to successful operation of a simulator), and moderate (μ=2.59/4.0, σ=0.85) in terms of risk-taking (i.e., which could be indicative of general driving habits/tendencies).

Simulator data

Refer to Figure 7, which illustrates the simulator scoring model across the NSTI experiment. Towards the bottom of the plot are the (five) individual rating components (for each driver) that were introduced in Table 2, and towards the top of the plot is the cumulative simulator rating (dark blue curve). The Y-axis represents the scoring (on a 0.0 to 1.0 scale), and the X-axis displays the 9 individual NSTI student drives, with vehicle parameter assignments first defined in Table 1. The green horizontal line is the calculated cohort average (μ=0.58/1.0; σ=0.17).

From the plot, numerous general observations can be made. There was a wide variance in rated simulator performance, with three drivers rated above, and six drivers rated below the overall cohort average. Of the three drivers who drove "very well", two (#2 and #4) were understeer vehicles, and surprisingly, one (#8), drove exceptionally well with a moderate oversteer vehicle. The remainder of the cohort rated between 40.0 (low) and 52.1 (high), including one understeer vehicle (#1), two oversteer vehicles (#6 and #9), and (surprisingly), all three neutral steer vehicles (#3, #5, and #7).

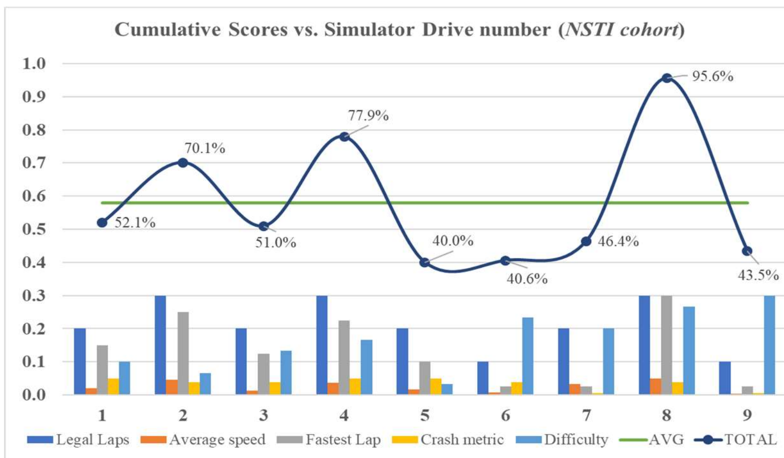

Figure 7 – Simulator drive results (*NSTI cohort*)

To compare the relative performance of different vehicle types (i.e., understeer, neutral steer, oversteer), in the forthcoming plots, we present position traces for Drive #1 (65/35 | understeer), drive #5 (50/50 | neutral steer), and drive #8 (50/65 | oversteer), respectively. On these plots, the track geometry of the Quad-radial speedway is denoted by the yellow circles (cones), and the student position X/Y trace across their 150 second drive duration is shown in red. We have also depicted an X/Y trace of an experienced "expert driver" shown in green, to illustrate a more ideal "racing line" for reference. However - illustrating the entire speedway at scale, it is difficult to observe nuances that influence performance to enable comparison of the key differences between the three primary vehicle configurations. Of particular interest is the most challenging segment of the speedway, which is the Turn #2/Turn #3 hairpin sector. For this reason, in Figures 8-10, we have magnified this area of the racetrack, which typically distinguishes drivers across skill levels.

When we observe the **<u>recovery from Turn #2 into Turn #3</u>**, with the understeer vehicle (Figure 8), the student is consistent (while the expert has more variance, likely due to higher attempted entry velocities), and the exit bands of both student and expert driver appear highly linear and undamped (i.e., no oscillatory behavior). For the neutral steer vehicle (Figure 9), both the student and the expert driver apex at the exact same point on each lap, and their exit bands appear as slightly damped oscillations. With the oversteer vehicle (Figure 10), there is a wider band of variability for both the student and expert driver, and their exit bands appear as moderately damped oscillations.

These general trends further accentuate as we observe the **recovery from Turn #3 heading towards Turn #4**. With the understeer vehicle (Figure 8), both the student, and expert driver are late exiting Turn #3, with hard braking required to slow down prior to the barrier cones. Again, the expert driver patterns are more pronounced due to elevated velocity. For the neutral steer vehicle (Figure 9), exiting Turn #3 happens much sooner, and appears controlled and consistent across both student and expert drivers. Finally, for the oversteer vehicle (Figure 10), exiting Turn #3 trends earlier than understeer, but there is much more variance and less control during this critical transition.

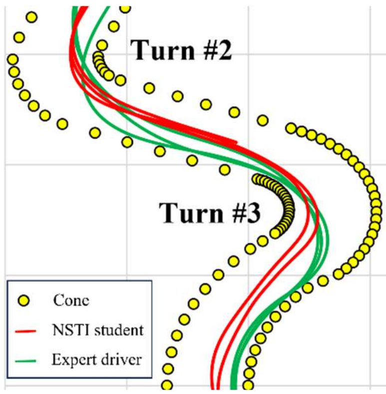

Figure 8 – Drive #1 (*close-up*)

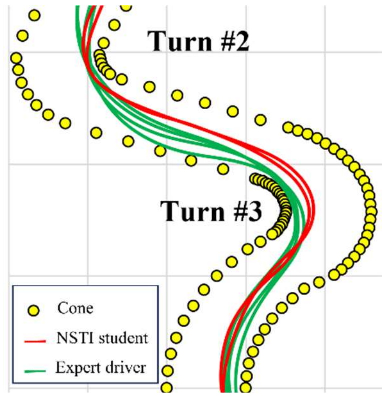

Figure 9 – Drive #5 (*close-up*)

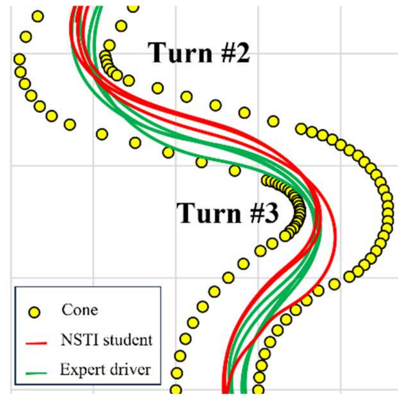

Figure 10 – Drive #8 (*close-up*)

From an education perspective, the takeaways from the simulator portion of our experiment serve to confirm theoretical first principals. The NSTI student cohort discovered that with the ability to control the vehicle and regulate speed during transitions, increased performance can be achieved – with both caution and skill - from a more oversteer vehicle. From high school physics, students' sense that centripetal acceleration is proportional to velocity squared divided by the turn radius (i.e., $a = v^2/r$). Accordingly, an oversteer vehicle can achieve elevated speeds and more naturally maneuver racing lines on turn sequences by widening the turning arc. After completing the turn, an oversteer vehicle can then recover more quickly to maneuver towards an optimal racing line on the straightaways, whereas an understeer vehicle lumbers for a prolonged period on its previous turning arc. As expected, the neutral steer vehicle provided more moderate and "idealized" driving characteristics. NSTI students (i.e., with little or no previous real-world nor simulator driving experience) discovered that this stratagem is easier stated than achieved.

Conceptual quiz

The conceptual quiz was issued both pre- and post-simulator and contained four questions (see Table 3) that related directly to the primary vehicle dynamics theoretical underpinnings for our GBL implementation. Refer to Figure 11 for a concise summary of the cohort results. For each of the N=22 students who completed both surveys, the red series are the PRE scores (range: 0 to +4), the blue series are the POST scores (range: 0 to +4), and the corresponding green overlay illustrates the improvement for each student (range: -4 to +4).

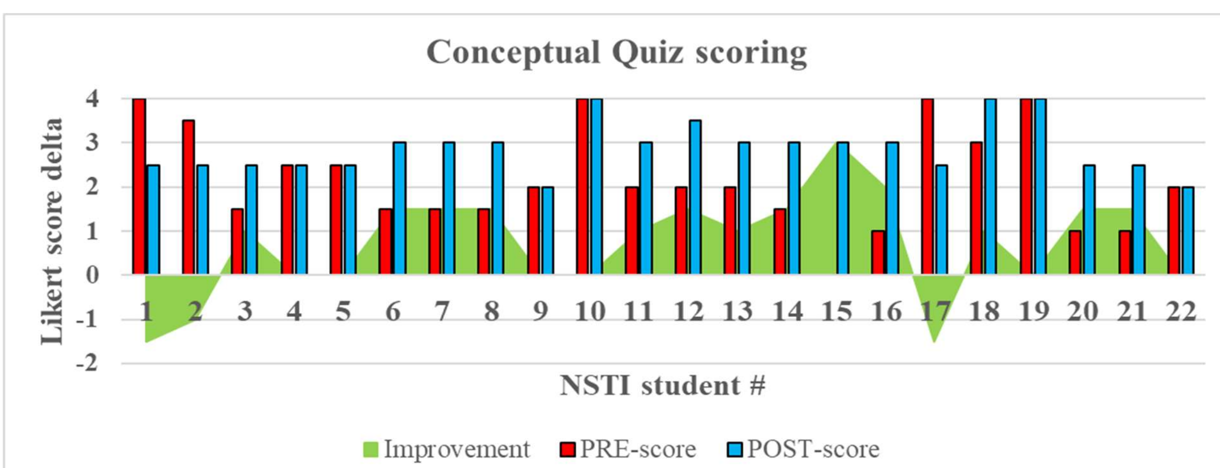

Figure 11 – Conceptual Quiz results (pre/post)

Accompanying statistics can be viewed in Table 5. The cohort mean increased from μ=2.18 to 2.89 from Pre to Post, with a companion reduction in standard deviation from σ=1.14 to 0.58, respectively. This is a general indicator that the NSTI conceptual understanding of oversteer/understeer, as well as vehicle weight and tire stiffness distribution, might have improved due to GBL exposure on the driving simulator. Score improvement averaged μ=0.71 per student, with a wide variance of σ=1.14. Overall, of the N=22 NSTI students, 13 student demonstrated improvement, 2 students remained the same, 3 students exhibited a reduction in score, and 2 students scored perfectly (both Pre and Post).

Table 5 – Conceptual quiz statistics

Metric	Mean (μ)	Standard deviation (σ)
PRE score	2.18	1.14
POST score	2.89	0.58
Improvement	0.71*	1.14*
*: *N=2 scored perfect both pre/post* | *N=13 improved* | *N=4 remained constant* | *N=3 decreased*		

In reviewing the data, we were concerned by the four students whose scores remained constant, and most concerned with the three students who performed worse after the theoretical and simulator segments. After a deeper analysis, we made an interesting discovery. All three students whose scores decreased – and two of the four students whose scores remained the same – **had the exact same responses on their POST conceptual quizzes**: Questions #1 and #2, correct; Question #3, partial credit; and Question #4, incorrect. This observation seems to imply that the experiment amply conveyed differences between vehicle conditions (i.e., over/under/neutral steer), while for some in the NSTI cohort -- confusion was elicited regarding the parameters that influence these conditions (i.e., weight and tire-stiffness distribution). The partial credit responses on Question #3 are not overly concerning, because these students recognized that a forward CG (generally) increases vehicle stability – whereas ultimately - stability is also interdependent on tire stiffness. However, the incorrect responses on Question #4 are concerning, as the exact same trend was misunderstood for tire stiffness: that is, forward tire stiffness (generally) implies increased stability, whereas ultimately, stability is also interdependent on weight distribution. Upon reflection, this confusion may have been because in the drive scenarios where stiff tires were nearest to the front of the car (i.e., drives #3, 6, and 9), the first was neutral steer, and the latter two were oversteer, all resulting in extremely erratic driving performance (and the lowest ratings/scores) by the young and inexperienced cohort. As such, it is possible that through experimentation, we inadvertently induced negative training (Hulme et al., 2021) with respect to this parameter, and the NSTI cohort would have benefited from additional conceptual reinforcement.

Learning preferences

At experiment completion, we queried the NSTI students to determine which mechanism was preferred to better understand vehicle dynamics concepts. The results are visualized in Figure 12. Of the N=22 students who completed the entire experiment, a majority of the cohort either preferred viewing slides (i.e., theory) that were interspersed with the simulator segments, or the simulator itself as being the MOST effective instructional mechanism.

A minority of the cohort preferred passively viewing others on the simulator, which for some, was an ideal opportunity to observe, analyze, and learn from peer performance. It is interesting to note that all three mechanisms were rated equally as the LEAST effective training mechanism. Anecdotally, we learned that many who rated the simulator in this manner commented that they did so simply because they did not have the opportunity to drive the simulator during the allotted experimental period. This was a noted weakness of our implementation which will be discussed further in the concluding remarks.

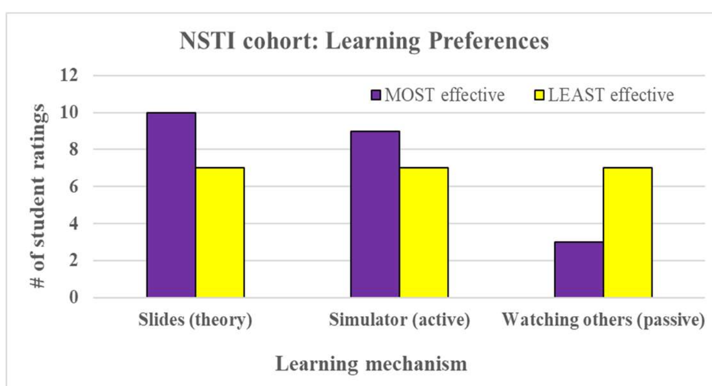

Figure 12 – NSTI student learning preferences

Overall experiment compliance

Post-experiment, the NSTI cohort rated the simulator demonstration positively. On a 4-point Likert scale, the average cohort rating was a 3.65 (±0.49), which is a strong indicator that the students appreciated and found value in the overall experience. Likewise, students in the NSTI cohort tolerated the simulator positively from a simulator sickness perspective. Post-experiment, we issued the Motion Sickness Assessment Questionnaire (MSAQ) (Gianaros et al., 2001), which is a 16-question survey (across four sickness categories) that rates on a 0–144-point scale. The average cohort score was 7.12 (±8.80), with a max score of 23, and a minimum score of 0 (frequency: N=8). This result was not unexpected, as younger adults are more compliant to simulators (Gálvez-García, 2015).

Correlation Analyses

Finally, we conclude this section by investigating correlations between the quantitative (simulator) and self-report (survey) data types to identify tendencies among the NSTI cohort participants. The following four analyses were attempted, with varying degrees of statistical success:

<u>Correlation #1: Conceptual Quiz rating vs. simulator driving performance</u>

We hypothesized that those in the cohort who drove well might also score high on the conceptual quiz (pre/post), and hence expected to observe an elevated positive correlation. However, the Pearson correlation (2008) between driving performance and the pre-Quiz score is approximately -0.054, which suggests a weak negative linear relationship between the two variables. For the post-Quiz scores, the Pearson relationship is -0.343, which indicates a moderate negative linear relationship. Both findings are therefore counter to our preliminary hypothesis.

<u>Correlation #2: Self-rated propensity for risk-taking vs. simulator driving performance</u>

Our intake surveys included one question regarding self-rated (1-4) propensity for risk-taking. The cohort (taken as a whole) reported a moderate average score of 2.59/4.0 (σ=0.85); this finding led us to speculate that there might be a relationship between calculated simulator performance (0 to 1.0) and risk tolerance. The Y-axis displays the normalized (average) simulator score for those who attested each of the 4 risk-propensity levels, which are plotted on the X-axis.

Figure 13 – Risk-taking correlation

Refer to Figure 13, which applies linear regression to analyze the relationship, and a small (non-significant) positive correlation is observed.

<u>Correlation #3: Learning preference vs. simulator driving performance</u>

Student respondents (N=9) who identified "simulator learning" as their primary mode of content assimilation demonstrated an elevated average driving score of approximately 0.68; higher than the overall cohort average (0.58), which is shown as the green line on Figure 7. In contrast, the mean score of those (N=10) students who favored traditional lecture-based learning was 0.51. The Pearson correlation coefficient {0.527} indicated a moderately positive linear relationship between simulator learning preference and driving performance. These findings suggest that a preference for simulator-based learning could be linked to improved (simulated) driving performance. <u>Implications</u>: This observation reveals the potential benefits of simulator-centric pedagogy in amplifying driving performance. The positive correlation suggests that driving education platforms might consider integrating or enhancing GBL and advanced M&S to potentially yield superior outcomes.

<u>Correlation #4: Stated learning preference for lecture vs. note-taking propensity</u>

Although the (N=10) NSTI students who preferred lecture-based instruction did not perform as well on the simulator, we wanted to observe and analyze their academic tendencies a step further. As noted previously, we allowed NSTI students to (optionally) take notes during the experiment, and we collected these note sheets – and codified the extent of each student's note-taking - at experiment completion. To ascertain relationships between these experiment parameters, a Chi-Square test (e.g., McHugh, 2013) was conducted. The derived p-value of 0.078, although surpassing the conventional threshold of 0.05, was noteworthy at $\alpha=0.1$, hinting at a potential statistical association. Implications: While this correlation may not be deemed significant under traditional standards, our findings postulate that individuals gravitating towards more passive, lecture-based learning might simultaneously benefit from an active strategy for independent engagement with the instructional material. This insight suggests the potential need for Education reforms that emphasize traditional lecture delivery with proactive learning techniques (e.g., note-taking).

CONCLUSIONS AND FUTURE WORK

Didactic instruction remains the most prevalent mode of teaching, however, the result is often a passive classroom environment where students are less motivated and not fully engaged, which historically results in substandard learning outcomes. The ongoing transition in Education is focusing on designing curriculum experiences to stimulate interaction with the world, where spatial and temporal decisions are required simultaneously. In this study, we evaluated the effectiveness of implementing Game-based Learning (GBL), physics-based modeling and high-fidelity M&S to convey critical vehicle design and motion dynamics principals to a high-school STEM cohort. Through hands-on participation, students observed vehicle steering conditions (oversteer/understeer) alongside two key design parameters that influence these conditions: 1) the percent front-to-rear weight distribution, and 2) the percent front-to-rear tire stiffness distribution. Our methodology addresses numerous priorities in Education, including novel instruction strategies, STEM implementations to accelerate learning, and training strategies staged within a variety of constructs. An executive summary of our primary experiment outcomes is provided below:

- We developed a five-category rating model to quantify (gamify) driver performance across the NSTI study cohort. Nine drivers participated in the experiment across vehicles with different assignments of the primary variables and results unsurprisingly suggested a general preference towards understeer.
- Delving deeper into the observed nuances between vehicle types, the NSTI cohort discovered that increased performance can be achieved – with both caution and skill - from a more oversteer vehicle, while the neutral steer vehicle provided more moderate driving characteristics.
- The four-item conceptual quiz cohort mean increased from $\mu=2.18$ to 2.89 from Pre to Post. Of the N=22 NSTI students, 13 students demonstrated improvement; a general indicator that the NSTI conceptual understanding was enhanced due to experiment participation. Note however that because there was no control group for the present examination, additional experimental testing is necessary to compare performance differences between the GBL implementation and conventional didactic (i.e., front-of-class) instruction.
- There was some lingering concern that for those students whose scores reduced on the conceptual quiz, and trends that were observed in the post-score responses, we may have inadvertently induced negative training on the simulator regarding tire stiffness behavior, specifically.
- Based upon post-experiment responses, we learned that about an equal number of NSTI cohort students preferred to learn experiment concepts from theory and the simulator. Furthermore, those that scored the simulator unfavorably admitted to doing so because not all among the cohort were afforded the opportunity to drive. This was a known weakness in our experimental design (due to time limitations); future implementations would address this demerit by having each member of a study cohort drive all nine combinations of vehicle parameter settings.
- Correlation analyses indicated that i) a preference for simulator-based learning could be linked to improved (simulated) driving, which reveals the potential benefits of simulator-centric pedagogy in amplifying driving performance; ii) those that preferred lecture-based learning might benefit from an active strategy for independent engagement with instructional material – i.e., traditional lecture delivery with proactive learning techniques (e.g., note-taking). (***note: neither finding was statistically significant***).
- Overall enjoyment of the experiment was rated high ($\mu=3.65/4.0$) post experiment, with simulator maladaptation rated as minimal, the latter of which was not unexpected.

Future Goals

While recent literature portrays an optimistic view of the potential for GBL to improve learning in higher education (Vlachopoulos & Makri, 2017), randomized controlled trials are uncommon, and supporting assessments (Ifenthaler et al., 2012) are often overlooked. It has been suggested that in-game (quantitative) data can be leveraged to improve ***retention rates***, ***conceptual knowledge, learner engagement***, ***performance measurements (e.g., gamification) and learner assessment*** among GBL students in engineering education. We will leverage the experimental outcomes from this study to continue to advance these urgent areas of interest in STEM education.

The discipline of Engineering Education continues to rapidly emerge, and researchers have yet to ascertain the degree to which GBL can positively impact learning systems, student performance, and trainee diversity and inclusivity (Wang et al., 2022). For a longitudinal analysis of concept reinforcement, it would be beneficial to include exposure to additional real-life applications using similar GBL approaches. The associated analyses would demonstrate the degree to which students have applied the learning principals garnered from the original exposure to GBL.

Ideas for improving the current study are being considered. In our next implementation, we plan to initiate a mixed cohort, where the students are divided in two groups. The first would receive the GBL component first (<u>and then evaluated</u>) and compared to the second (control) group, who would receive the theoretical component first (<u>and then evaluated</u>). After the formal evaluations, the groups would transition to the alternate training mechanism so that all students receive both training types; this approach would yield a more direct assessment of the effectiveness of GBL.

ACKNOWLEDGEMENTS

The authors extend extreme thanks to the University at Buffalo Stephen Still Institute for Sustainable Transportation and Logistics (SSISTL), including direct support towards constructing Game-based Learning (GBL) Modeling & Simulation (M&S) educational experiences described in this paper. The authors extend their gratitude to Moog, Inc. (East Aurora, NY), for their ongoing support for STEM Education.

REFERENCES

Balci, S., Secaur, J. M., & Morris, B. J. (2022). Comparing the effectiveness of badges and leaderboards on academic performance and motivation of students in fully versus partially gamified online physics classes. *Education and information technologies*, 27(6), 8669–8704. https://doi.org/10.1007/s10639-022-10983-z

Ballen, C.J., Wieman, C., Salehi, S., Searle, J.B., & Zamudio, K.R. (2017), Enhancing Diversity in Undergraduate Science: Self-Efficacy Drives Performance Gains with Active Learning. *CBE Life Sci Educ*. Winter;16(4):ar56. doi: 10.1187/cbe.16-12-0344. PMID: 29054921; PMCID: PMC5749958.

Fromm, J., Radianti, J., Wehking, C., Stieglitz, S., Majchrzak, T.A., & Brocke, J. (2021). More than experience? - On the unique opportunities of virtual reality to afford a holistic experiential learning cycle. *The Internet and Higher Education*. 50. 100804. 10.1016/j.iheduc.2021.100804.

Gálvez-García, G., (2015), "A comparison of techniques to mitigate Simulator Adaptation Syndrome," Ergonomics, Vol. 58, no. 8, pp. 1365-1371, 10.1080/00140139.2015.1005168.

Gianaros, P., Muth, E., Mordkoff, J., Levine, M., & Stern, R., (2001), "A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness," Aviation, Space, and Environmental Medicine, Vol. 72, pp. 115-9, PMID: 11211039; PMCID: PMC2910410.

Ha, V., Chung, L., Hanh, N. & Hai, B. (2023). Teaching Science Using Argumentation-Supported 5E-STEM, 5E-STEM, and Conventional Didactic Methods: Differences in the Learning Outcomes of Middle School Students. *Education Sciences*. 13. 247. 10.3390/educsci13030247.

Hulme, K.F.; Schiferle, M.; Lim, R.S.A.; Estes, A.; Schmid, M. (2021), Incorporation of Modeling, Simulation, and Game-Based Learning in Engineering Dynamics Education towards Improving Vehicle Design and Driver Safety. *Safety*, 7, 30. 10.3390/safety7020030

Ifenthaler, D., Eseryel, D., & Ge, X. (Eds.), (2012). Assessment in game-based learning: Foundations, innovations, and perspectives. New York: Springer

Jääskä, E., Lehtinen, J., Kujala, J., & Kauppila, O. (2022). Game-based learning and students' motivation in project management education. *Project Leadership and Society*. 3. 100055. 10.1016/j.plas.2022.100055.

Ješková, Z.; Lukáč, S.; Šnajder, Ľ.; Guniš, J.; Klein, D.; Kireš, (2022). M. Active Learning in STEM Education with Regard to the Development of Inquiry Skills. *Educ. Sci*., 12, 686. https://doi.org/10.3390/educsci12100686

Matthews, A.K., Allen-Meares, P., Watson, K., Crooks, N., Smith, A., Hart, A., Estrella, M. L., & Kim, S. (2020). The use of strategies from the social sciences to inform pipeline development programs for under-represented minority faculty and students in the health sciences. *Journal of clinical and translational science*, 5(1), e73. https://doi.org/10.1017/cts.2020.566

McHugh M. L. (2013). The chi-square test of independence. *Biochemia medica*, 23(2), 143–149. https://doi.org/10.11613/bm.2013.018

Milliken, W.F., and Milliken, D.L., (1995), "Race Car Vehicle Dynamics", SAE 1995 ISBN 1-56091-526-9

Nadelson, L., & Villanueva, I., & Bouwma-Gearhart, J., & Lanci, S., & Youmans, K., & Lenhart, C. A., & Van Winkle, A. K. (2019), Knowledge in the Making: What Engineering Students are Learning in Makerspaces Paper presented at 2019 ASEE Annual Conference & Exposition, Tampa, Florida. 10.18260/1-2--33039

Nguyen, K.A., Borrego, M., Finelli, C.J., DeMonbrun, M., Crockett, C., Tharayil, S., Shekhar, P., Waters, C., & Rosenberg, R., (2021). Instructor strategies to aid implementation of active learning: a systematic literature review. *IJ STEM Ed* 8, 9. https://doi.org/10.1186/s40594-021-00270-7

Pearson's Correlation Coefficient (2008). In: Kirch W. (eds) Encyclopedia of Public Health. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5614-7_2569

Reynaldo, C., Christian, R., Hosea, H., & Gunawan, A. (2021). Using Video Games to Improve Capabilities in Decision Making and Cognitive Skill: A Literature Review. *Procedia Computer Science*. 179. 211-221. 10.1016/j.procs.2020.12.027.

Sandoval, M., Oviedo, G., & Rodriguez-Torres, M. (2022). From preconceptions to concept: The basis of a didactic model designed to promote the development of critical thinking. *International Journal of Educational Research* Open. 3. 100207. 10.1016/j.ijedro.2022.100207.

Staehle, M.M., Childers, R.C. & Guilford, W.H. From Theory to Practice: Unraveling the Impact of Experiential Learning in Biomedical Engineering Education. *Biomed Eng Education* 3, 111–112 (2023). https://doi.org/10.1007/s43683-023-00122-6

Vlachopoulos, D., & Makri, A., (2017). The effect of games and simulations on higher education: a systematic literature review. *Int J Educ Technol High Educ* 14, 22. https://doi.org/10.1186/s41239-017-0062-1

Wang, L.H., Chen, B., Hwang, G.J., Guan, J-Q., & Wang, Y-Q., (2022). Effects of digital game-based STEM education on students' learning achievement: a meta-analysis. *IJ STEM Ed* 9, 26. https://doi.org/10.1186/s40594-022-00344-0

Warfighter Digital, (2023), The Exciting Frontier: What is the Future of Gamification in Military Training?, https://www.warfighterpodcast.com/blog/the-exciting-frontier-what-is-the-future-of-gamification-in-military-training/, published April 20, 2023.

Yin, D., Shan, D., & Hu, J.-S., (2017), A Study on the Control Performance of Electronic Differential System for Four-Wheel Drive Electric Vehicles. *Appl. Sci.*, 7, 74. https://doi.org/10.3390/app7010074