Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
ASEE
Paper ID #48501

Title Air and Missile Defense Threat Scenario Variation to Reduce Pretest Sensitization, Video Games as a Case Study

Julie R Szekerczes, Indiana State University
Dr. M. Affan Badar, Indiana State University

M. Affan Badar, PhD is Professor, former department chair, and current Director of the PhD in Technology Management Program in the Bailey College of Engineering & Technology at Indiana State University. In 2016-18 he was on leave and worked as Professor and Chair of the Industrial Engineering and Engineering Management Department at university of Sharjah, UAE.

Dr. James Nevin McKirahan Jr., Indiana State University

JAMES MCKIRAHAN is currently an assistant professor in Applied Engineering and Technology Management in the College of Technology at Indiana State University. Dr. McKirahan earned an Associate of Applied science degree in Electronics Technology from Sout

Dr. A. Mehran Shahhosseini, Indiana State University

A. Mehran Shahhosseini is a Mechanical Engineering Professor in the Department of Applied Engineering and Technology Management at Indiana State University. He has published over 65 articles in different journals and conferences.

AIR AND MISSILE DEFENSE THREAT SCENARIO VARIATION TO REDUCE PRETEST SENSITIZATION, VIDEO GAMES AS A CASE STUDY

Abstract

This study uses fixed and variable video game types to measure pretest sensitization as a proxy for repeated and varied threat test scenarios in system performance testing of air and missile defense systems. The pretest sensitization phenomenon exists when repeated exposure to a test condition influences the participant's response. Research shows that air and missile defense development correlate with video games, resulting in similar interfaces and computer operating environments. This study demonstrates the need to vary the scenarios to provide accurate system performance results supporting programmatic decisions. The study uses a three-factor nested factorial design of experiments. Independent variables are: video game type (IV1), title nested within game type (NV1), and completion: first or replays (IV2). Time to complete the game is dependent variable. Five video game titles are sampled for each type, comparing average first completion(s) with replay(s) to detect a decrease in the time to complete the game with repeated exposure, indicating pretest sensitization. The data downloaded from www.HowLongToBeat.com represents a snapshot of results when downloaded. Using all completionist data with notes results in a sample size of 1,598. The findings confirm pretest sensitization with repeated exposure, impacting the accuracy of system performance and adding risk to programmatic decisions. Hence, the test scenarios in Department of Defense (DoD) or engineering management training must be varied to help determine system performance accurately. This is important because system performance results dictate future course of action in engineering management or DoD decision-making. Such results inform acquisition decisions such as further funding and development, program canceling, and fielding decisions.

Keywords
Test scenario variation, pretest sensitization, video game, nested factorial design.

1. **Introduction**

As part of the U.S. Department of Defense acquisition process, a program office develops a product per the needs/requirements defined by a service, such as the Army. Within the Army, the Army Test and Evaluation Command tests and evaluates the product to determine if it fills the capability gap(s), providing critical information to organizations that need it. Data-based results inform fielding decisions and oversight functions and provide the program office with results and areas that need improvement. Due to the cost and safety considerations of flying live targets against a system, modeling and simulation (M&S) are used to assess system performance against the required threat set when testing air and missile defense systems. Appendix A provides a list of acronyms used in the study). High-density threat simulations (scenarios) are presented to the soldier-operated system, providing a preponderance of the operational system's performance data. Scenario reuse has historically been prevalent due to the complexity of scenario development and time and cost constraints. Unfortunately, scenario reuse does not support complete system discovery learning and results in pretest sensitization, biasing system performance results. This bias can result in misinformed technology management decisions. If a scenario is static / non-changing, subsequent presentations of the same scenario lead to “gaming the system," and the outcomes are based on familiarity with the scenario. A first step in improving the test setup and resulting system performance data accuracy is to vary the scenarios presented during the test such that any supporting group of soldiers fights against a scenario only once, avoiding pretest sensitization. The goal is to isolate actual system performance for evaluation and remove any prior knowledge of threats presented in a scenario. Soldier performance is outside the scope of system evaluation (and this study). Hence, this study does not control human factors and measurement of metrics such as comprehension, absorption, and training effectiveness of training scenarios, which correlate to age and gender, are also outside the scope of this study.

Army doctrine for fratricide avoidance specifies that only certain types of targets will be automatically engaged and intercepted. Therefore, soldiers are in the loop to engage most target types in system operational testing. Soldier-in-the-loop engagements within the software/hardware-in-the-loop (S/HWIL) environment are the focus of this study. Soldiers are unaware of what will occur during testing and operate as if deployed with the system. When presented with the same scenarios several times, they begin to anticipate the threats, and become ready to engage them before they appear on the screen. Such knowledge of the scenario specifics would not be present operationally when deployed with the system. This prior familiarization with the scenarios biases the system performance results. This study investigates the presence of pretest sensitization in repeated exposure to video games as a proxy for repeated scenarios in S/HWIL testing requiring soldier-in-the-loop execution decisions. The study examines the need to vary the scenarios to avoid injection of soldiers' prior knowledge of scenario content into testing.

Video Games as a Case Study. Due to the complex nature of air and missile defense, soldiers engage the threats using a computer, similar to a video game interface. Hence, like a user attempting successive attempts against a fixed pattern video game, soldiers will begin to recognize the scenario pattern and target behavior with consecutive presentations of the same scenario. Soldiers may anticipate the threat and prepare to engage it before it appears on the screen. Given the similarities between air and missile defense and video game interfaces, the correlation between their development paths, and the classification of air and missile defense data, video games serve as a case study. Fixed and variable pattern video game results are analyzed and compared for pretest sensitization to determine the need for scenario variation in air and missile defense testing. Video games in the study need not be air and missile defense-related as the goal is to remove soldiers' prior knowledge of scenario content; the very nature of their digital computer interface provides the data needed, explicitly focusing on differences between the first completion and replay time(s) of the same game(s). The video game case study employing the three-factor nested factorial design of experiments offers similar results to repeated attempts against the same scenario.

Problem. The frequency of testing supporting agile software development, new development system integration, and continual emerging findings result in short suspense follow-on test events and very constrained timelines to generate scenarios. The scenario development process remains labor intensive, and analysts must test the scenarios after creation for integration with other M&S architecture tools and the defense design. The process usually requires modifications to the scenarios followed by a re-test. The process takes weeks and does not meet the demand, causing the community to reuse already developed and vetted scenarios. The community disagrees on whether repeated scenarios lead to pretest sensitization and the ongoing demand for scenarios under very constrained timelines results in pressure to reuse existing scenarios. The problem of this study is the lack of evidence concluding whether repeated scenarios lead to pretest sensitization, biasing the data, and potentially leading to inaccurate system performance results supporting high-level acquisition decisions. The community needs this information to confirm that scenario variation is required and to allow sufficient time for this process and/or to allocate funds for the improvement of the M&S tools; otherwise, report the test limitation with the system results.

Purpose. This study uses video games as a case study to detect pretest sensitization from fixed pattern video games compared to variable patterned ones and to detect any statistically significant impact on the response. Unclassified game completion time data readily available on the internet enables the analysis, focusing on the time difference(s) between first completion(s) and replay(s) of several fixed and variable pattern video games. The postulation anticipates a reduction in time from first completion(s) to replay(s) of fixed patterned video games due to prior knowledge of what will happen in the replays of those games, but not for the variable patterned ones.

RQ1: Does a fixed or variable video game battle type (IV1) affect the user's time to complete the video game (DV)?

$$H_{0_1}: \mu(t)_{0(fixed)} = \mu(t)_{1(variable)} \tag{1}$$
$$H_{1_1}: \mu(t)_{0(fixed)} < \mu(t)_{1(variable)} \tag{2}$$

RQ2: Does video game completion (first or replay) (IV2) affect the user's time to complete the video game (DV)?

$$H_{0_2}: \mu(t)_{0(first\ completion)} = \mu(t)_{1(replay(s))} \tag{3}$$
$$H_{1_2}: \mu(t)_{0(first\ completion)} > \mu(t)_{1(replay(s))} \tag{4}$$

RQ3: Does the video game title (NV1) nested under video game battle type (IV1) affect the user's time to complete the video game (DV)?

$$H_{0_{3_1}}: \mu(t)_{0(title1,fixed)} = \mu(t)_{1(title2,fixed)} = \mu(t)_{2(title3,fixed)} = \mu(t)_{3(title4,fixed)} = \mu(t)_{4(title5,fixed)} \tag{5}$$
$$H_{0_{3_2}}: \mu(t)_{0(title1,variable)} = \mu(t)_{1(title2,variable)} = \cdots = \mu(t)_{4(title5,variable)} \tag{6}$$
$$H_{1_{3_1}}: \mu(t)_{0(title1,fixed)} \neq \mu(t)_{1(title2,fixed)} \neq \cdots \neq \mu(t)_{4(title5,fixed)} \tag{7}$$
$$H_{1_{3_2}}: \mu(t)_{0(title1,variable)} \neq \mu(t)_{1(title2,variable)} \neq \cdots \neq \mu(t)_{4(title5,variable)} \tag{8}$$

RQ4: Are there interaction effects (IV1 × IV2) between video game battle type (IV1) and completion (first or replay) (IV2)?

$$H_{0_4}: \rho_{(IV1 \times IV2)} = 0 \tag{9}$$
$$H_{1_4}: \rho_{(IV1 \times IV2)} \neq 0 \tag{10}$$

RQ5: Are there interaction effects (IV2 × NV1) between completion (first or replay) (IV2) and video game title (NV1) nested under video game battle type (IV1)?

$$H_{0_5}: \rho_{(IV2 \times NV1)} = 0 \quad (11)$$
$$H_{1_5}: \rho_{(IV2 \times NV1)} \neq 0 \quad (12)$$

This study is important as the findings will help engineering management and Department of Defense (DoD) training and evaluation community on the path forward for test planning, execution, and reporting.

2. Literature Review

The literature review demonstrates the importance of M&S in the Department of Defense acquisition. It examines digital methods of delivery and impacts on retention. Similarities between air and missile defense systems and video games are delineated, supporting the applicability of the case study to the problem. The literature review explains the video game design types to discern which would be equivalent to repeated and varied scenarios. It identifies video game titles for each design type, indicating the needed datasets.

The www.HowLongToBeat.com site has the required dataset readily available. This site allows gamers to track what they are playing and identify new games that interest them. Players can catalog their gaming collection, research additional games to purchase, and estimate how long it will take them to complete any game or their entire backlog (collection of video games they own and have not played or finished yet). Gamers can also compare their game completion times with other players. The site provides statistics and tracking for its members. Data is collected, and this site summarizes statistics in several categories. The Main Story (required category) includes the main game objectives, being the minimum needed to get to the credits at the end of the game (much like the credits that roll at the end of a movie). The Main Story plus Additional Quests/Medals/Unlockables category contains the main story content plus additional tasks that are not required. The player reaches the 100% Completionist category when they earn every medal and complete everything the game offers. The Combined category considers all play styles in the estimation. The site also collects speedrun data for users attempting to complete the game as fast as possible (HowLongToBeat, 2023). The literature review revealed no study duplication investigating the effects of repeated digital games or scenarios on pretest sensitization.

Importance of M&S in Acquisition Testing. The Air and Missile Defense community uses M&S for robust system testing in a cost-constrained environment. There is limited availability of live surrogates and safety issues associated with flying specific targets, flying certain types of targets simultaneously in the same airspace, and flying many air objects concurrently at the range. Range time, support infrastructure, targets, and personnel are constrained and costly assets. Software and hardware in the loop (S/HWIL) M&S allow testing defense against particular aerial platforms and at higher track and network loading levels. Live testing does not permit such testing due to safety and cost considerations. The threat scenarios are part of the distributed interactive simulation used during testing to present the threat to the system. Such testing provides the system data needed in reduced timeframes, supporting required system performance analysis. The Department of Defense relies on M&S to support engineering, augment test and evaluation, support the acquisition process, and enhance training (Boyer, 1993). Fortanbary et al. indicate that the use of simulation with the design of experiments is needed to support acquisition operational testing using a minimum of resources (Fortanbary et al., 1996, pp. 81-90).

The Department of Defense uses M&S to test new systems at all phases of the design life cycle, reducing design costs and fielding schedules by identifying and fixing anomalies early. It enables integration testing of new and existing systems configurations, such as the case for Army Integrated Air and Missile Defense (Henderson et al., 2005, pp. 2-59). Terminal High Altitude Area Defense (THAAD) hardware in the loop (HWIL) missile simulations support missile flight tests, development, verification, and validation of embedded software and simulations, and assessment of performance against current and future threat environments (Buford, 1998). Ground-test capability helps to reduce system development risk by testing the latest technologies for space-based, interceptor, and airborne sensor platforms (Lowry et al., 2011, pp. 999-1011). Wind tunnel Virtual Flight Testing serves as a realistic and reliable method for testing and evaluating aircraft and missile flight control systems (Huang & Wang, 2015). Flight paths and route assessments are modeled for UAVs (Unmanned Aerial Vehicles) to reduce the risk of flight in a hostile environment (Wei et al., 2013, pp. 646-660) & (Sun et al., pp. 1-13). Simulation helps estimate an aircraft's vulnerability to the dual-band man-portable air defense system and aids in flare countermeasure model development that improves aircraft survivability (Smith, 2015). Historically, research has recognized the need for M&S to augment testing and evaluation in the acquisition process due to constrained costs and the increasing complexity of the system and requirements as early as 1990 (Atkinson, 1990). Interoperable M&S are used to analyze asymmetric threats in marine coastal environments and to conduct training against these threats (Tremori, 2013).

Digital Delivery Methods and Retention. Training threat scenarios supporting air and missile defense lose their value after more than three uses with the same group of soldiers. The soldiers recognize the scenario and begin to "game" the simulated air battle (Hawley, 2006) (Hawley & Mares, 2007). Routine training enhanced by scenario variability is needed to challenge repetitive skills. Adaptive expertise is developed over time by training using increasing scenario variability and challenge levels (Hawley & Mares, 2007). Their findings imply the existence of pretest sensitization from repeated exposure to training scenarios.

Well-scripted static video games can provide exhilarating events, challenging gameplay, and intense player immersion. However, repeated playthroughs of static games reduce the quality of immersion and enjoyment. Replay value is improved when the game is less predictable. (Snowdon & Oikonomou, 2011).

Similarities Between Air and Missile Defense and Video Games. With the air and missile defense and video game development pathways intertwining, it is not a coincidence that air and missile defense community personnel have described the interfaces as like playing a video game. The military has historically enabled the growth of video games, utilizing them for recruiting and training. The globalization of video games has impacted military operations in modern warfare.

The Cathode Ray Tube Amusement Device game developed in 1947 featured a radar screen similar to a submarine extant at the time. This game allowed players to simulate the firing of missiles at targets depicted on the screen. The Department of Defense funded Spacewar!, developed by graduate students of Massachusetts Institute of Technology (MIT) in 1962. With the advancement of video games, militaries have incorporated video game technology into various aspects of their organizations, utilizing it in military applications worldwide. The user interface and controls for piloting a UAV are like operating a video game (Story, 2018). Video games paralleled the Strategic Defense Initiative (SDI) in the 1980s. President Reagan identified that video games could strengthen America's military by the early 1980s. Video games such as Space Invaders, Defender, Battlezone, and Combat contributed to his perspective. Video games from the 1980s helped Americans envision the SDI, making missile defense more plausible. By the late 1980s, the SDI proponents used video games for propaganda. Critical video games supporting the growth of air and missile defense included Missile Command, Wargames, Strategic Defense Initiative, and High Frontier. A military recruiter reported that The Missile Command locater ball was just like one used for an air defense system. It presented enemy aircraft on the radar screen, using the ball to control the cursor and pushing a button to fire the interceptor and neutralize the enemy (Kapell & Elliott, 2013).

Air and missile defense weapons systems, such as the Army's Patriot system, continuously increase automation use. Automation will further proliferate in future weapons systems. Some systems are approaching the threshold of having autonomous operations. Patriot has two operating modes: semi-automatic and automatic. Semi-automatic mode is an operator-in-the-loop system and is the mode of discussion herein. Patriot system developers adapted its engagement algorithms from the Safeguard system, the first operational U.S. anti-ballistic missile system, eliminated as part of a treaty limiting U.S. and Soviet Anti-Ballistic Missile systems. The potential for misclassification in ambiguous situations, such as tracking in cluttered environments, requires human intervention for certain target types. The Patriot interface involves tracks on a screen being "hooked" and an engagement committed/authorized on hostile tracks from the computer. The soldier queries the system to confirm classification and identification results, comparing with track characteristics. There is continual growing interest in automation for weapon system autonomy driven by increased computing power, software engineering, and artificial intelligence (Hawley, 2017).

After a realistic commercial demonstration, Matrix Games developed "wargames" for the Navy and the Air Force—these "wargames" support training, education, and concept development. The Army allocated $400,000 for the initial phase of a similar effort with the potential to expand to $4 million for prototyping (solicitation fall 2018). Developers based the game on land warfare, and the prototype was based on publicly available data to simulate realistic combat, featuring actual terrain and layouts. The Army pursued One-World-Terrain, leveraging currently available virtual reality and gaming industry technologies to expedite development processes. (Kwiatkowski, 2018). An increasingly close relationship between the US military and the digital game industry has resulted in military interfaces similar to those encountered in video games. (Power, 2007).

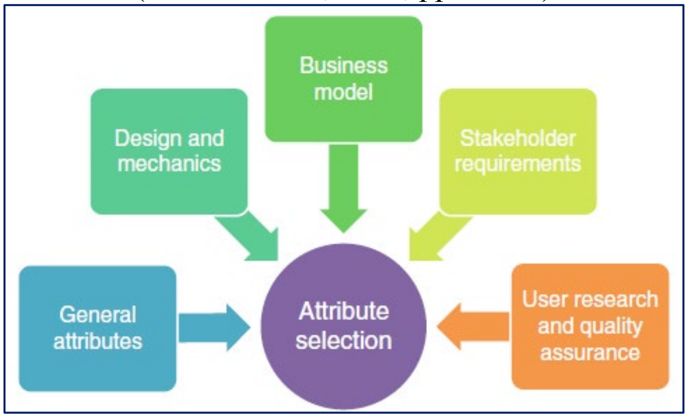
Video Game Design Types. Modern video games and simulation software tend to have either static and unchanging content or content randomized within a narrow set of parameters (Hastings et al., 2009, pp. 245-263). Developers design boss fights to test the player's abilities. They are generally dangerous foes that take longer to kill than regular enemies in the game and could likely end the player character's life within a few devastating attacks. They assess a subset of the player's skills learned in the game area, addressing new, thematically different playable regions (s) and/or new gameplay mechanics. The user must complete them before proceeding to the game's next area. Boss fights add to the playtime and longevity of a game, making it memorable. They require a strategy to defeat, can include multiple phases, and have a puzzle. They can surprise the player with different mechanics or theming. Bosses continually spark one's interest. Simplistic boss fights can be predictable, leading to complacency with repeated exposure. The boss fight should feel like a new experience at peak difficulty, even if testing mechanics previously used. Players will tend to recognize boss fight patterns with repeat encounters (Agriogianis, 2018).

Video game boss fights include fixed and variable pattern designs. Both design types involve fixed attack patterns. While evolving, AI used in video games has not yet progressed to being able to create original attacks. Fixed Pattern fights involve unique enemies having a prescribed means of defeat. Once the player discerns the pattern, it will no longer be challenging. Fixed pattern designs limit replay value and become more manageable with repeated attempts, reducing the players' interest level in the fight. Effectively, the enemy is repeating a set and unchangeable pattern. This design type is susceptible to exploiting the patterns or "gaming" the battle. Variable pattern designs randomly change the order in which the attacks are presented to the player and may incorporate conditions that determine which attack will be used based on player statistics, such as health or location proximity to the enemy. A player must be more reactive in a variable pattern design, which is more challenging. The number of varied attacks and their frequency determines the difficulty level. From Software, known for developing challenging video game battles, created multiple pools of attacks based on the player statistics mentioned above. The rate at which enemies attacked, and the rapid transition between attack pools resulted in a perceived effect of emergent behavior. (Bycer, 2018).

During gameplay, a person becomes involved in a continual loop of actions and corresponding responses from the game, resulting in the game's state changing. The time the user is engaged will depend upon the number of state changes and actions required of the player and the variation and complexity of the system responses to the user's actions (Drachen et al., 2013, pp. 13-40).

Gameplay Metrics Available for Study. The gaming industry uses game mechanics and underlying business model (story play-through or free-to-play) metrics to determine the types of player telemetry data to log and analyze. One can collect telemetry data on an event, such as when a user starts or finishes a game, or over a certain frequency, like the location of a character within the game at given time intervals. A game analyst can also initiate data collection by turning on or off tracking of a specific attribute, to check on the results of a software patch, for example. Not all metrics can be collected, stored, and analyzed due to bandwidth issues and the extensive resources required. A few key metrics that provide the most needed information are chosen and only collected on a percentage of game players. The data type necessary for focused improvements and cost-benefit analysis generally drives the data collected. Figure 1 depicts the drivers of attribute selection for data collection (Drachen et al., 2013, pp. 13-40).

Figure 1. Drivers of Attribute Selection for Game Data Collection.
(Drachen et al., 2013, pp. 13-40)



Given the required balance with resources needed, developers collect general attributes across all games, such as user, and when a player starts and finishes a game. The essential core gameplay, game mechanics, and business attributes are also generally collected. User research and quality assurance tend to use core metrics. Developers only collect a sample of non-core metrics as needed.

Dataset Needed for Study. The literature review identified candidate fixed and variable video game titles for the study as of April 2022. (Mitra, 2021), (Mota, 2023), (Top5Gaming, 2017), (WatchMojo.com, 2017), (Szpytek, 2022), and (theDeModcracy, 2020). Of these identified, game titles have been chosen within a recent decade (2010 to 2023) to minimize variation from technology evolution. Video games have different story lengths. Average completion times for the subject games are available on HowLongToBeat.com informing sampling from games of similar length. This study uses the 100% completionist playthroughs with notes data category to prevent variability based on which sides players choose to complete and due to play style, ensuring the data's normality. 100% completionists had sufficient data for sampling and generally had more data points available than in the main story category.

Data collection planning analysis indicated that the preponderance of the average completion times occur between 24.25 – 145.5h for fixed design video games and between 41.5 – 97h for variable ones, within the first through third quartiles, where time is in hours (h = hours). Only one fixed title has an average completion time between 41.5 and 97 hours. To ensure sufficient fixed title sample size, titles with average completion times between 24.25-145.5h were used for both game types. A few exceptions are made for fixed games to allow a sample size of five titles per game type. Too few replays for the required analysis disqualified Halo 4. The sample size needed for fixed video game titles included Undertale and Legend of Zelda: Breath of the Wild, even though their story lengths are shorter and longer than the selected range, respectively. Exhibit 2 provides the resulting sample set.

Time to complete is a valuable metric for the study, indicating whether replays occur faster than first completions. The data on the How Long to Beat website is updated regularly (HowLongToBeat, 2023). Hence, the results of this study focus on a single point of time when the data was collected. Each reported completion specifies the completion time. The analysis stratifies the data based on metadata flags indicating which completion occurred first versus replay(s). It compares these subsets for differences between average first completion(s) and average replay(s).

3. Methodology

This research compares first completion versus replay times for fixed and variable pattern games and measures for pretest sensitization as a case study for repeated and varied scenarios, respectively.

Research Design. Error rate setting: The FMEA and cost-risk-benefit analysis resulted in 0.05 for both alpha and beta. Titles per game type considered for the study as specified by the literature review, within release dates between 2010-2023, and with similar story lengths to the extent that provides the sample size, result in five fixed and twelve variable titles for the study. The analysis uses all five resulting fixed games, and randomly selects five of the twelve variable games using Microsoft Excel. Excel generated the random number set (3, 8, 11, 12, 1) applied to the variable game titles. The resulting video game titles for data collection include Borderlands 2, Dragon Age: Inquisition, Dying Light, Legend of Zelda: Breath of the Wild, and Undertale for fixed design type; and Batman Arkham City, Sekiro: Shadows Die Twice, Dark Souls, Elden Ring, and Doom Eternal for variable design type.

Exhibit 4 specifies the data type for each factor. Game type (IV1) and title (NV1) are nominal, and completion (IV2) is ordinal; the time to complete (DV) is continuous ratio data. Data collection indicates 1,128 more first completions than replays. Also, the HowLongToBeat.com website implementation of anonymous and private categories precludes the ability to correlate replay(s) with first completion(s)

Video game design type (IV1) has two levels, the video game title nested within game design type (NV1) has five levels, and completions (first and replay(s)) (IV2) has two levels. With the inability to determine which replays go with their respective first completions, this study uses all available completionist playthroughs with notes data points. Table 1 summarizes the counts of data points available for analysis. The resulting sample size is 1598. (Badar, Raman, Pulat, & Shehab, 2005), (Kashlak, 2019), (Montgomery, 2013).

Table 1. Sample Size.

Type	Title	First	Replay	All
Fixed	Borderlands 2	31	14	45
	Dragon Age	61	10	71
	Dying Light	29	6	35
	Legend of Zelda	214	36	250
	Undertale	118	23	141
Variable	Batman	34	11	45
	Dark Souls	72	28	100
	Doom Eternal	94	14	108
	Elden Ring	521	56	577
	Sekiro	189	37	226
	All	1363	235	1598

The subsequent design of experiments is a three-factor nested factorial design, as shown in Table 2. (Badar, Raman, Pulat, & Shehab, 2005), (Launsby & Schmidt, 2005), (Ryan, 2007) and (Montgomery, 2013). Data collection obtains time to complete the video game (DV) data for each video game design type (IV1), title (NV1), and completion (IV2) to comprise the sample size. As shown in the text box in the body of the table, each cell of the body of the data table contains the time to complete the video game (DV) data. There are 1,128 more first(s) data points than replay(s), and sample sizes vary between video game titles.

Table 2. Design of Experiments Construct.

IV1: Type	*NV1: Title*	*IV2: Completion*	
		First(s)	Replay(s)
Fixed	Borderlands 2 Dragon Age: Inquisition Dying Light Legend of Zelda: Breath of the Wild Undertale	*DV: Time to Complete Video Game* (data for each cell of this table)	
Variable	3- Batman Arkham City 8- Sekiro: Shadows Die Twice 11 - Dark Souls 12 - Elden Ring 1 - Doom Eternal		

Note. IV = Independent variable. NV = Nested variable. DV = Dependent variable. The number in front of the game title refers to the random number generated that resulted in the game title selection.

Table 3 summarizes the characteristics of the research variables. The independent variable types are as follows: video game design type (IV1) is nominal/categorical, video game title (NV1) nested under the video game design type is nominal/categorical, and game completion (first(s), replay(s), IV2) is ordinal. The dependent variable, the time to complete the video game (DV), is continuous ratio data (Badar, Raman, Pulat, & Shehab, 2005), (Kashlak, 2019), (Montgomery, 2013). The time to complete data accuracy is per the video game or computer clock and is in hours and minutes. The analysis process converts this time format to hours. While the analysis uses hours, HowLongToBeat.com supplies the data with the accuracy of minutes, so the analysis supports accuracy to 0.02 hours or 1 minute.

Table 3. Variable Characteristics.

Variable	*Game Type*	*Game Title*	*Completion*	*Time to Complete*
Code	Type	Title	Completion	t
Data Type	Nominal	Nominal	Ordinal	Ratio
Units	N/A	N/A	N/A	Hours
Accuracy	N/A	N/A	N/A	0.02 minutes
Range/Levels	• Fixed • Variable	Title 1→5	• First(s) • Replay(s)	Target: Constant
Source	LR informed	HLTB	HLTB	HLTB
Instrumentation	N/A	N/A	N/A	Computer clock
Importance to Research	High	Low	High	High
Easy or HTC	HTC	HTC	HTC	N/A
Experimental Control	IV1	NV1	IV2	DV

Note. t = Time. N/A = Not Applicable. LR = Literature Review. HLTB = HowLongToBeat.com. HTC = Hard to change. IV = Independent variable. NV = Nested variable. DV = Dependent variable.

The linear statistical model of the observed response for the design is as follows:

$$y_{ijkl} = \mu + (IV1)_i + (IV2)_j + (NV1)_{k(i)} + (IV1 \times IV2)_{ij} + \varepsilon_{(ijk)l} \tag{13}$$

Where the observed response is y_{ijkl}, μ is the overall main effect mean, the effect of the *i*th video game design type is $(IV1)_i$, the effect of the *j*th completion (first or replay) is $(IV2)_j$, the effect of the *k*th video game title within the *i*th level of the video game design type is $(NV1)_{k(i)}$, the *video game design type* × *completion* interaction is $(IV1 \times IV2)_{ij}$, and the random experimental error is $\varepsilon_{(ijk)l}$. The $(IV2 \times NV1)_{jk(i)}$ *completion* × *video game title* interaction is removed from the model, being not statistically significant (Badar, Raman, Pulat, & Shehab, 2005), (Kashlak, 2019), (Montgomery, 2013).

Data Collection. The data was downloaded from the How Long to Beat website (HowLongToBeat.com) using a personal computer. The data on www.HowLongToBeat.com is updated regularly (HowLongToBeat, 2023); hence, the results focus on a single point in time when the data was collected. Each reported completion specifies the completion time. The "completionist playthroughs with notes" dataset category was used to minimize variation due to play styles and to ensure the dataset was not modal. The analysis stratifies the data based on metadata flags indicating which completion occurred first versus replay(s).

4. Results and Discussion

The study findings confirm that completion (IV2) and title (NV1) both have statistically significant impacts on the time to complete the video game (DV). The fact that completion is statistically significant is an essential finding to the air and missile defense test and evaluation community that has implications for the planning and reporting processes. The title being significant is inherently intuitive, as the video games had different story lengths, but it is less critical of a finding. Results coupled with visual confirmation on the box, scatter, and contour plots indicate that the type and completion interaction (IV1 × IV2) is also significant.

Linear Mixed Model Assumptions. Required assumptions were met, except independence. The analytical planning process randomly chose video game titles (NV1) for the varied video game type(s); the study used all five fixed designs that met the selection criteria to ensure a sufficient sample size. Analysis revealed positively skewed raw timing data (DV), but a natural log transform improved the normality of the data. Type (IV1) and title (NV1) are inherently correlated, as the title is nested within the type per the design. The replay(s) data correlates with some of the first completion(s). However, with anonymous users, it is unknown which replay(s) go with which first completion(s). Had this information been known, the study would have focused on repeated measures analysis. Given the data limitation(s), the study employs the best alternative analysis method, comparing the average of the first completion(s) with that of the replay(s). The data met the homogeneity of the residuals assumption. Table 4 provides a summary of the assumption results. The study provides the information needed, given the randomly selected titles and the data meeting the normality and homogeneity of the residual(s) assumptions.

Table 4. Assumptions Summary.

Assumptions	Met	Notes
Random	✓	Title: Varied random; Fixed used all for required sample
Normality of Residuals	✓	
Homogeneity of Residuals	✓	
Independence	X	• Type & Title correlated by nested design • Levels of Completion (IV2) correlated.

Note. IV = Independent variable.

Normality of Residuals: Subgroup Distributions and Summary Statistics: The analysis produced subgroup distributions by type (IV1) and completion (IV2) to view the general shape of the data and identify outliers. Data is stratified in this manner as these factors are critical in detecting pretest sensitization from repeated exposure to the video game(s). The analysis revealed the positive skew of the raw data; the mean(s) being larger than the median(s). The study detected outliers in all subgroup datasets; however, the metadata did not highlight a special cause; some users take longer to complete the video game(s) than others. Recall that there are 1,128 more first-time game completion data points than replays. The analysis involved taking a natural log transform of the timing data, improving the normality, and removing the skew. The log-transformed mean(s) are close to the median(s), and only one subgroup has outliers. The transformation, having improved the normality as compared to the raw timing data, met the normality assumption and prepared the data for statistical testing.

Independence: Type (IV1) and title (NV1) are inherently correlated, as the title is nested within the type per the design. The replay(s) correlate with some of the first completion(s). With the user being anonymous, it is unknown which replay goes with which first completion. Had this been known, the study would have focused on repeated measures analysis. The best alternative analysis method is used, given the data limitation(s), averaging and comparing all first(s) and replay(s). Note that independence violations can lead to potential bias impacting the accuracy of type I and II errors, and some model fitting results may be uncertain.

Homogeneity Of Residuals Across Subgroups: Examination showed that the variance is centered vertically around zero and is similar across subgroups, verifying the homogeneity of residuals.

Linear Statistical Mixed Model Results. Completion (IV2), title (NV1) and the type and completion interaction (IV1 × IV2) have statistically significant impacts on the time to complete the video game (DV). Completion being statistically significant is an essential finding for the air and missile defense test and evaluation community. The linear mixed model results in Tables 7 and 8 show that the mean for replay(s) is less than for first-game completion(s). This difference in means indicates decreased time to complete the game with repeated exposure, suggesting that pretest sensitization is present. Game title (NV1) being significant is intuitive as different games have different story lengths and are not as crucial of a finding. Scatter and contour plots also provide a visual indication that the type and completion interaction (IV1 × IV2) is also significant.

Initial: The initial unbalanced nested mixed model fixed effect tests shown in Table 5 indicate that completion (IV2) and the type and completion interaction (IV1 × IV2) have statistically significant effects on the time to complete the video game (DV). Their probabilities are less than the 0.05 alpha chosen in the error rate setting for a 95% confidence level. F statistic initial model results for these factors is $F_{Completion}(1, 9) = 11.07$, $p = 0.008$, $F_{Type}(1, 8) = 0.04$, $p = 0.848$, and $F_{Completion*Type}(1, 9) = 7.53$, $p = 0.022$. The F statistic reporting format is F(numerator degrees of freedom, denominator degrees of freedom) = F ratio, followed by the p-value. For example, the F statistic for completion is 11.07, with numerator degree of freedom = 1, denominator degrees of freedom = 9, and a p-value of 0.008 confirming statistical significance. (P-value = $0.008 < 0.05 = \alpha$). Video game type (IV1) has no statistically significant effect. With varied types of video games randomly pulling from the same set of pre-programmed responses, the user(s) still recognized the response. Similarly, Table 6 further indicates that the game title (NV1) statistically significantly affects the time to complete the video game (DV), as its confidence interval does not contain zero. The statistical significance of title (NV1) makes sense as the different video games have different story lengths. The restricted maximum likelihood (REML) analysis using JMP version 17 estimated the title (NV1) variance component to be 0.60, with a standard error of 0.30 and a 95% confidence interval of (0.01, 1.18), where the confidence interval format is (95% lower confidence bound, 95% upper confidence bound). The completion and title interaction (IV2 × NV1) only contributed to 0.08% of the total random error, which is negligible, and its confidence interval contains zero, indicating it does not have a statistically significant effect on the response. Therefore, the final model is revised, removing this interaction, simplifying it, and improving its accuracy.

Table 5. Initial Mixed Model Fixed Effect Test Results Using F Statistic for Equality of Means and Correlation.

Source	N parm	DF	DF Den	F Ratio	Prob > F
Completion	1	1	9	11.07	0.008*
Type	1	1	8	0.04	0.848
Completion*Type	1	1	9	7.53	0.022*

**p < 0.05*

Note. N parm = Number of parameters. DF = Degrees of Freedom (numerator).
DF Den = Degrees of Freedom (denominator).

Table 6. Initial Model REML Variance Component Estimates.

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Wald p-Value	Pct of Total
Title [Type]	3.38	0.60	0.30	0.01	1.18	0.046*	77.09
Title*Completion [Type]	0.00	0.00	0.00	0.00	0.00	0.754	0.08
Residual		0.18	0.01	0.17	0.19		22.83
Total		0.77	0.30	0.41	1.98		100.00

**Wald P-Value < 0.05*

Note. Var = Variance. Std = Standard. Pct = Percent. -2 Log Likelihood = 1836.37 Total is the sum of the positive variance components.
Total including negative estimates = 0.77

Final: The results from the revised model effectively provide the same conclusion. Tables 7 and 8 provide the refined statistical test results. The F statistic for the final revised model results is $F_{Completion}(1, 1586) = 13.40$, p = 0.000, $F_{Type}(1, 8) = 0.04$, p = 0.851, and $F_{Completion*Type}(1, 1586) = 10.35$, p = 0.001. Figure 2 box plots visually depict that the mean is lower for replay(s) than first completion(s) (IV2) for fixed and variable type game(s) (IV1), confirming pretest sensitization with repeated exposure. The REML analysis estimated the title (NV1) variance component for the final revised model to be 0.59, with a standard error of 0.30 and a 95% confidence interval of (0.01, 1.18). Again, the confidence interval does not contain zero, indicating that the factor title (NV1) statistically impacts the time to complete the game(s).

Table 7. Final Mixed Model Fixed Effect Test Results Using F Statistic for Equality of Means and Correlation.

Source	N parm	DF	DF Den	F Ratio	Prob > F
Completion	1	1	1586	13.40	0.000*
Type	1	1	8	0.04	0.851
Completion*Type	1	1	1586	10.35	0.001*

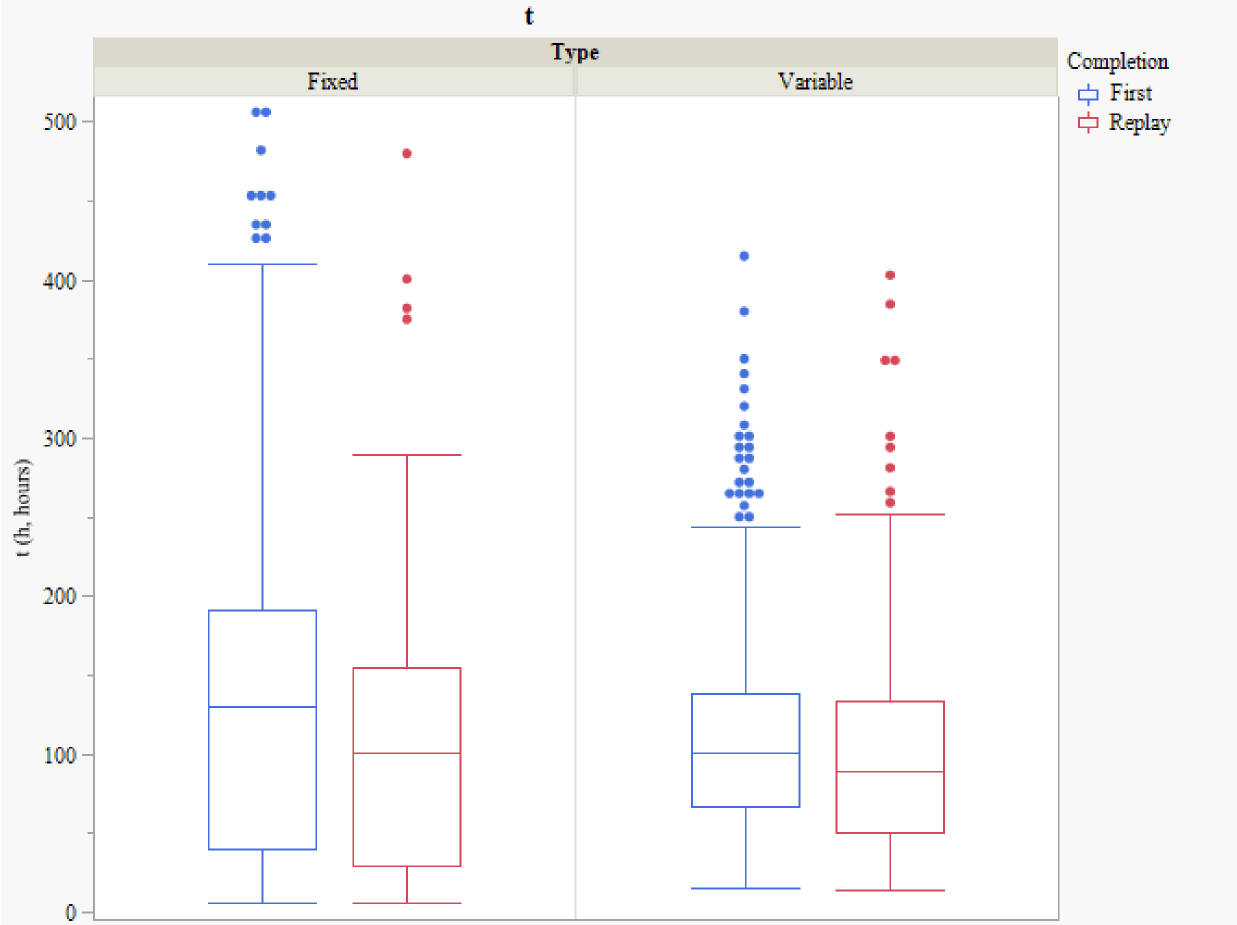
$*p < 0.05$
Note. N parm = Number of parameters. DF = Degrees of Freedom (numerator).
DF Den = Degrees of Freedom (denominator).

Table 8. Final REML Variance Component Estimates.

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Wald p-Value	Pct of Total
Title [Type]	3.35	0.59	0.30	0.01	1.18	0.046*	77.04
Residual		0.18	0.01	0.17	0.19		22.97
Total		0.77	0.30	0.41	1.96		100.00

**Wald P-Value* < 0.05
Note. Var = Variance. Std = Standard. Pct = Percent. -2 Log Likelihood = 1836.50. Total is the sum of the positive variance
components. *Total including negative estimates = 0.77*

Figure 2. Boxplots of t (DV) by Type (IV1) and Completion (IV2).



Note. t = Time in hours. DV = Dependent variable. IV = Independent variable.

Using the predicted value results converted back to hours, the scatter plots showed the mean time to complete the video game(s) (DV) is less for replay(s) than for first completion(s), confirming a decrease with additional exposure and the presence of pretest sensitization. The fixed-type games result in a more pronounced reduction in time from first completion(s) to replay(s) at about a 16-hour difference compared to variable ones with a one-hour decrease. The difference between these time reduction values provides evidence of the type and completion interaction (IV1 × IV2). The contour plot depicted a more significant difference in the shape of the data between first completion(s) and replay(s) for the fixed-type video games than for the variable ones, also visually confirming the type and completion interaction (IV1 × IV2). Without interaction effects, the contour plot(s) would show the same data shape throughout.

Results Consistent with Other Research. The literature review indicates that the study results agree with the findings from other research. Key findings detected by the study were that completion (IV2) and the type and completion interaction (IV1 × IV2) have a statistically significant impact on the time to complete the video game (DV), serving as a proxy for fixed versus varied test threat scenarios. These findings confirm the presence of pretest sensitization, which will bias system performance results. Similarly, Hawley and Mares' research found that training threat scenarios supporting air and missile defense lose their value after more than three uses with the same soldiers. They learned that soldiers recognize the scenario and begin to "game" the simulated air battle (Hawley, 2006) (Hawley & Mares, 2007). Test threat scenarios tend to have higher fidelity than training threat scenarios, but they are similar, and the effects identified are the same. Pretest sensitization present from repeated exposure to the training scenarios allowed the soldiers to recognize the pattern(s) and "game" the air battle. The decrease in the time to complete a video game with repeated exposure is consistent with their findings that soldiers "gamed" the air battle with repeated exposure to the training scenarios. Recognition of repeated patterns and prior knowledge of what would occur enabled these effects.

Video game-related research and development has identified the same phenomena. Snowdon and Oikonomou found that repeated playthroughs of static games reduce the quality of immersion and enjoyment. Players lost interest with repeated exposure and prior knowledge of what would occur (pretest sensitization present). However, they noted that the replay value is improved when the game is less predictable and the player is unaware of what will happen. (Snowdon & Oikonomou, 2011). Agriogianis also detected complacency with repeated exposure and found that players recognize boss fight patterns with repeat encounters (Agriogianis, 2018). Bycer explains that when an enemy repeats a set and unchangeable pattern, it is susceptible to exploiting the patterns or "gaming" the battle. (Bycer, 2018). These studies focused on game development and improvement, but the presence of pretest sensitization caused the effects that they observed.

The study findings agree with the observations and conclusions reported in the air and missile defense training and video game development and improvement research.

5. Conclusions

The study confirms pretest sensitization with repeated exposure to the video game(s) as a proxy for reusing test threat scenario(s). Such bias confounds and impacts the accuracy of system performance results. It is essential to vary the scenarios in testing or provide a caveat with the results, informing senior leaders of the bias present. This caveat would inform senior leaders of the risk(s) involved in making key program decisions based on the resulting performance results.

Table 9. Hypotheses, Results, and Research Question Answer(s) Summary.

RQ	Answer	Hypotheses	Results
6. Does game type (IV1), affect time to complete (DV)?	No	$H_{0_1}: \mu(t)_{0(fixed)} = \mu(t)_{1(variable)}$ $H_{1_1}: \mu(t)_{0(fixed)} < \mu(t)_{1(variable)}$	✓Failed to Reject Null
7. Does completion (IV2) affect time to complete (DV)?	Yes	$H_{0_2}: \mu(t)_{0(first\ completion)} = \mu(t)_{1(replay(s))}$ $H_{1_2}: \mu(t)_{0(first\ completion)} > \mu(t)_{1(replay(s))}$	*X* Null Rejected
8. Does title (NV1) affect time to complete (DV)?	Yes	$H_{0_{3_1}}: \mu(t)_{0(title1,fixed)} = \mu(t)_{1(title2,fixed)} = \mu(t)_{2(title3,fixed)} = \mu(t)_{3(title4,fixed)} = \mu(t)_{4(title5,fixed)}$ $H_{0_{3_2}}: \mu(t)_{0(title1,variable)} = \cdots = \mu(t)_{4(title5,variable)}$ $H_{1_{3_1}}: \mu(t)_{0(title1,fixed)} \neq \cdots \neq \mu(t)_{4(title5,fixed)}$ $H_{1_{3_2}}: \mu(t)_{0(title1,variable)} \neq \cdots \neq \mu(t)_{4(title5,variable)}$	*X* Null Rejected
9. Are there interaction effects (IV1 × IV2) between type and completion?	Yes	$H_{0_4}: \rho_{(IV1 \times IV2)} = 0$ $H_{1_4}: \rho_{(IV1 \times IV2)} \neq 0$	*X* Null Rejected
10. Are there interaction effects (IV2 × NV1) between completion and title?	No	$H_{0_5}: \rho_{(IV2 \times NV1)} = 0$ $H_{1_5}: \rho_{(IV2 \times NV1)} \neq 0$	✓Failed to Reject Null

Note. RQ = Research question. IV = Independent variable. DV = Dependent variable. NV = Nested variable
H0 = Null hypothesis. H1 = Alternate hypothesis. μ = Statistical mean.

The study detects pretest sensitization with repeated exposure to the same video game, serving as a case study for repeating test threat scenarios. Video game completion (IV2), game title (NV1), and the type and completion interaction (IV1 × IV2) have a statistically significant impact on the time to complete the video game (DV). Table 9 summarizes the hypothesis test results and answers to the research questions.

Research Question 1. Does a fixed or variable video game battle type (IV1) affect the user's time, μ(t), to complete the video game (DV)?

The linear mixed model fixed effect test compared the time to complete the video game(s) (DV) between game types (IV1). With an F statistic of $F_{Type}(1, 8) = .04$, $p = .851$ and $P > \alpha$ ($\alpha = 0.05$) failing to reject H_{0_1}, the outcomes of the two design types are similar. There is no significant difference in the mean (average) time to complete (DV) between fixed and variable video game battle types (IV1).

Both types use a fixed set of video game battle responses; the fixed repeats the same battle set, while the variable draws from a pool of battle responses. The pool has only a limited number of battle responses to draw. The user recognizes the battle responses in the variable design video games with repeat exposure. Additionally, some video game titles have lengthy stories. There are a variety of battle designs within a video game, causing battle-type similarity between the two design types over the entire game.

Research Question 2. Does video game completion (first or replay) (IV2) affect the user's time, $\mu(t)$, to complete the video game (DV)?

The fixed effects test comparing the time to complete the game(s) (DV) between completions (IV2) revealed an F statistic of $F_{Completion}(1, 1586) = 13.40$, $p = 0.000$. $P < \alpha$ indicates rejection of the alternate hypothesis (H_{1_2}). This effect is a critical finding of the study, detecting the presence of pretest sensitization with repeated exposure to a video game, serving as a proxy for repeating air and missile defense acquisition test and evaluation scenarios. This finding indicates that repeated scenario(s) result in pretest sensitization bias confounding the system performance results. This impact necessitates varying scenarios during testing to ensure accurate performance results support key program acquisition decisions. Otherwise, report results with a caveat so senior leaders understand the risk in making decisions based on the reported system performance results.

Research Question 3. Does the video game title (NV1) nested under video game battle type (IV1) affect the user's time, μ(t), to complete the video game (DV)?

The linear mixed model restricted maximum likelihood (REML) variance component estimates comparing the time to complete the game(s) (DV) between the random factor game titles (NV1) projected the title variance component to be 0.59, with a standard error of 0.30 and a 95% confidence interval of (0.01, 1.18). The confidence interval does not contain zero, indicating rejection of the null hypotheses ($H_{0_{3_1}}$ & $H_{0_{3_2}}$). While inherently intuitive because the various games have different story lengths, this finding is not particularly interesting. It effectively states that the time to complete an air battle during testing will depend on how long of an air battle the analyst(s) designs within the threat scenario. Indeed, time to complete fundamentally correlates with the video game story length or the length of the air and missile defense threat scenario.

Research Question 4. Are there interaction effects (IV1 × IV2) between video game battle type (IV1) and completion (first or replay) (IV2)?

The fixed effects test identified an interaction effect between design type and completion (IV1 x IV2) on the time to complete the game(s) (DV) with an F statistic of $F_{Completion}(1, 1586) = 13.40$, $p = 0.000$. $P < \alpha$ indicates rejection of the null hypothesis (H_{0_2}). In addition to the linear mixed model results, the box, scatter, and contour plot(s) provide visual confirmation of the interaction. This interaction effect is another important finding of the study. This discovery shows that repeat exposure to either battle type impacts the time to complete the video game (DV). Repeated exposure to the same static threat scenario influences the system performance results, confounding them with pretest sensitization. Repeated exposure to a scenario that can vary itself reduces the magnitude of the impact.

Research Question 5. Are there interaction effects (IV2 × NV1) between completion (first or replay) (IV2) and video game title (NV1) nested under video game battle type (IV1)?

The REML variance component estimates did not detect an interaction between completion and title (IV2 x NV1) affecting the time to complete the game(s) (DV), assessing the interaction (IV2 x NV1) variance component to be 0.0, with a standard error of 0.0 and a 95% confidence interval of (0.0, 0.0). The confidence interval contains zero, failing to reject the null hypothesis (H_{0_5}). Additionally, the interaction only contributed to 0.08% of the total random error, which is negligible, so it was removed from the final model, simplifying it and improving its accuracy. This finding confirms that the impact on the time to complete the video game (DV) response with repeated exposure is consistent across video game titles (NV1) nested within the same game design type.

6. Implications and Recommendations

This study has implications for engineering management community and researchers who deal with training in general and in Department of Defense (DoD) in particular. The findings of this study indicate that the test scenarios in training or air and missile defense acquisition testing must be varied. Scenario variation within the test design will reduce or avoid pretest sensitization bias confounding the results, providing more accurate system performance results supporting key acquisition decisions. If the scenarios are not varied, a caveat must be reported with the results indicating the presence of pretest sensitization bias impacting the accuracy, informing senior leaders of the risk involved in making key decisions based on the provided results.

Recommend funding the development of artificial intelligence (AI)-based scenarios designed to attack per an adversary's rules of engagement, modifying their attack based on real-time soldier response to further improve system performance results' accuracy. Current technology develops scenarios before tests; they are static and cannot change or evolve during the air battle based on soldier actions. Additionally, if a leaker gets through, there is no automatic way of removing the friendly asset that was "hit." (A leaker is an incoming threat that reaches its intended target because the system didn't successfully intercept it). A software application exists identifying friendly assets targeted by the threats in the scenario during an air battle. Then, if a threat is not engaged, a real-time casualty assessment team will attempt to manually remove friendly assets "hit" by the threat during the test, which can be very difficult in a high-intensity air battle. This method is not accurate or timely and is prone to human error. Removing assets that a scenario threat has impacted is vital because they would no longer be available to contribute to defense. A reduced number of sensors, shooters, and relays impacts system-of-systems effectiveness in continuing to fight the air battle. Automating real-time casualty assessment is recommended to achieve improved accuracy of the system performance results from software/hardware in the loop (S/HWIL) testing.

Future Research. With continual video game development and evolution, researching games released since 2023 to detect pretest sensitization would be informative. Analysts could study how much variation in battle type occurs within a game. Video game technology is advancing rapidly. When video game developers implement AI, it would be fascinating to learn whether this removes any residual pretest sensitization compared to the current variable patterned game design type(s). Such research would be beneficial to both gamers and video game developers.

Optimally, future air and missile defense scenarios may be artificial intelligence-based, designed to tactically attack per an adversary's rules of engagement and modify its attack based on real-time soldier response. M&S's future advancements could account for the destruction of simulated assets (real-time casualty assessment) and perhaps incorporate a red adversary team in the loop tactically coordinating attacks. Cooley and Oswalt have indicated the critical need for M&S artificial intelligence to adapt based on a response. (Cooley & Oswalt, 2021). These advancements are recommended for future research and are outside the scope of this study.

References

Agriogianis, T. (2018). *The Roles, Mechanics, and Evolution of Boss Battles in Video Games.* Long Island University. Long Island: Digital Commons @ LIU. Retrieved January 29, 2023, from https://digitalcommons.liu.edu/post_honors_theses

Atkinson, J. (1990). Modeling and simulation in the test and evaluation process [Abstract]. *SIMULATION, 54*(3), 127. Retrieved May 19, 2020, from http://tg5zw4at7r.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=Modeling+and+simulation+in+

Badar, M. A., Raman, S., Pulat, P. S., & Shehab, R. L. (2005). Experimental Analysis of Search-Based Selection of Sample Points for Straightness and Flatness Estimation. *Journal of Manufacturing Science and Engineering, 127*, 96-103. Retrieved February 15, 2023

Balci, O., Arthur, J. D., & Ormsby, W. F. (2011). Achieving reusability and composability with a simulation conceptual model. *Journal of Simulation, 5*, 157–165. Retrieved 10 16, 2022, from https://link.springer.com/content/pdf/10.1057/jos.2011.7.pdf

Bobbitt, Z. (2021). *The Three Assumptions of the Repeated Measures ANOVA*. Retrieved February 2023, from Statistics. Simplified. Statology: https://www.statology.org/repeated-measures-anova-assumptions/

Boyer, R. (1993). Joint modeling and simulation system [Abstract]. *Proceedings of SPIE, 1969*(1), 335–340.

Buford, J. J. (1998). Using hardware-in-the-loop (HWIL) simulation to provide low-cost testing of TMD IR missile systems [Abstract]. *Proceedings of SPIE, 3368*(1), 432-440.

Bycer, J. (2018). *The Evolution of Boss Designs in Video Games*. Retrieved Oct 18, 2022, from Game Developer: https://www.gamedeveloper.com/design/the-evolution-of-boss-designs-in-video-games

Clements, R. B. (1991). *The Experimenters Companion: A Guide and Reference to the Aspects of Research and Experimentation.* Milwaukee: ASQC Quality Press.

Cooley, T., & Oswalt, I. (2021). *Operationalizing Artificial Intelligence in Simulation-Based Training [2021 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)].* Retrieved from researchgate.net: https://www.researchgate.net/profile/Tim-

Cooley/publication/356773653_Operationalizing_Artificial_Intelligence_in_Simulation_Based_Training_Operationalizing_Artificial_Intelligence_in_Simulation_Based_Training/links/61aac84950e22929cd452038/Operationalizi
Department of Defense OUSD (AT&L). (2006). *RISK MANAGEMENT GUIDE FOR DOD ACQUISITION.* Retrieved 2012, from www.acq.osd.mil: http://www.acq.osd.mil/se/docs/2006-RM-Guide-4Aug06-final-version.pdf
Dondlinger, M. J. (2007). Educational Video Game Design: A Review of the Literature. *Journal of Applied Educational Technology, 4*(1), 21-31. Retrieved October 18, 2022, from http://hypermedia468.pbworks.com/w/file/fetch/82077116/Dondlinger2007EducationalVideoGameDesign.pdf
Drachen, A., El-Nasr, M. S., & Canossa, A. (2013). *Chapter 2 Game Analytics - The Basics.* Retrieved January 29, 2023, from cmps-people.ok.ubc.ca: https://cmps-people.ok.ubc.ca/bowenhui/game/readings/ch2-game-metrics.pdf
Fortanbary, M., Mansager, B., & Newberry, C. (1996). Supporting acquisition decisions through effective experimental design. *Mathematical and Computer Modelling, 23*(1/2), 81–90. Retrieved March 8, 2021, from Mathematical and Computer Modelling: https://www.sciencedirect.com/science/article/pii/0895717795002197
Hastings, E. J., Guha, R. K., & Stanley, K. O. (2009). Automatic Content Generation in the Galactic Arms Race Video Game. *IEEE Transactions on Computational Intelligence and AI in Games, 1*(4), 245–263. doi:10.1109/TCIAIG.2009.2038365
Hawley, J. (2017). *Patriot Wars: Automation and the Patriot Air and Missile Defense System.* Retrieved 10 16, 2022, from CNAS (Center for a New American Security) Report: Ethical Autonomy Series: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf?mtime=20170106135013&focal=none
Hawley, J. K. (2006). *Training for Effective Human Supervisory Control of Air and Missile Defense Systems.* Army Research Laboratory, Human Research and Engineering Directorate. Adelphi: Army Research Laboratory. Retrieved October 17, 2022, from https://apps.dtic.mil/sti/pdfs/ADA447995.pdf
Hawley, J., & Mares, A. (2007). *Developing Effective Adaptive Missile Crews and Command and Control Teams for Air and Missile Defense Systems.* Army Research Laboratory, Juman Research and Engineering Directorate. Adelphi: Army Research Laboratory. Retrieved October 17, 2022, from https://apps.dtic.mil/sti/pdfs/ADA464724.pdf
Hayden, M. (2008). *Multi-Factor ANOVA & Multiple Regression.* Terre Haute, IN: Quality Council of Indiana
Hayden, M. (2013a). *Spring 2013 Advanced Statistical Analysis in Technology (COT 703) Assignments (Assignment H: Error Rates).* Retrieved 2013, from ISU
Hayden, M. (2013b). *Spring 2013 Advanced Statistical Analysis in Technology (COT 703) Assignments (Assignment G: Assumptions).* Retrieved 2013, from ISU
Henderson, S., Wolter, J., & Kwinn, J. M. (2005). A Systems Engineering Approach to the Design of the Air Defense Artillery Simulation (ADASIM) Architecture. West Point, NY, USA. Retrieved from https://apps.dtic.mil/sti/tr/pdf/ADA436542.pdf
HowLongToBeat. (2023). *Completions (per video game title)*. (a Ziff Davis Company) Retrieved January 2023, from howlongtobeat.com: https://howlongtobeat.com/game/27100/completions

Huang, M., & Wang, Z.-w. (2015). *Hindawi Publishing Corporation: International Journal of Aerospace Engineering.* Retrieved May 19, 2020, from A Review of Wind Tunnel Based Virtual Flight Testing Techniques for Evaluation of Flight Control Systems: http://downloads.hindawi.com/journals/ijae/2015/672423.pdf

Indeed. (2021). *What Are Game Metrics and Do They Matter? (With Examples)*. Retrieved January 2023, from Indeed career advice and development: https://www.indeed.com/career-advice/career-development/what-are-game-metrics

Informa PLC Informa UK. (2013). *Intro to User Analytics*. Retrieved January 2023, from Game Developer: https://www.gamedeveloper.com/business/intro-to-user-analytics

jmp.com. (2022). *learning-library-all.pdf.* Retrieved February 2023, from jmp: https://www.jmp.com/content/dam/jmp/documents/en/academic/learning-library/learning-library-all.pdf

Kapell, M. W., & Elliott, A. B. (2013). *Playing with the Past: Digital Games and the Simulation of History.* Bloomsbury Publishing. Retrieved 10 16, 2022, from https://books.google.com/books?id=gZPFAgAAQBAJ&dq=are+air+%26+missile+defense+operations+similar+to+a+video+game&lr=&source=gbs_navlinks_s

Kashlak, A. (2019). *Design and Analysis of Experiments.* Edmonton, Canada: University of Alberta. Retrieved March 28, 2023, from https://sites.ualberta.ca/~kashlak/data/stat568.pdf

Kwiatkowski, M. (2018). Army takes new approach to developing training tools. *Inside the Army, 30*(46), 1-11. Retrieved 10 16, 2022, from https://www.jstor.org/stable/26587405

Launsby, R., & Schmidt, S. (2005). *Understanding Industrial Designed Experiments.* Colorado Springs: Air Academy Press

Lowry, H., Nicholson, R., Simpson, R., & Mead, K. (2011). Test and Evaluation of Space and Airborne Imaging Sensor Platforms. *JOURNAL OF SPACECRAFT AND ROCKETS, 48*(6), 999–1011

Mitra, R. (2021). *10 Easy Bosses That Still Cause Us Stress*. Retrieved February 7, 2023, from TheGamer: https://www.thegamer.com/easy-bosses-that-still-cause-us-stress-high-stakes/

Montgomery, D. C. (2013). *Design and Analysis of Experiments.* Hoboken: John Wiley & Sons, Inc.

Mota, L. (2023). *Top 10 Easiest Video Game Bosses*. Retrieved January 2023, from The Top Tens: https://www.thetoptens.com/games/easiest-video-game-bosses/

Power, M. (2007). Digitized Virtuosity: Video War Games and Post-9/11 Cyber-Deterrence. *SEcurity Dialogue, 38*(2), 271-288. Retrieved 10 16, 2022, from https://doi.org/10.1177/0967010607078552

Ryan, T. (2007). *Modern Experimental Design.* Hoboken: John Wiley & Sons, Inc.

Smith, L. (2015). Effective expendable countermeasure model against dual-band infrared and ultraviolet man-portable air-defence seeker systems [Abstract]. *Optical Engineering, 54*(8), 85104

Snowdon, J., & Oikonomou, A. (2011). Creating more entertaining and re-playable games by dynamically introducing and manipulating, static gameplay elements and events. *16th International Conference on Computer Games (CGAMES)*, pp. 94–100. doi:10.1109/CGAMES.2011.6000322

Steve, S. (2023). *How to Generate Random Numbers in Excel (Without Duplicates)*. Retrieved February 2023, from spreadsheetplanet.com, Amazon Services LLC: https://spreadsheetplanet.com/generate-random-numbers-without-duplicates-excel/

Story, C. M. (2018). *How the Globalization of Video Games Is Changing the Way Militaries Operate.* The University of Texas at Austin, Kinesiology. Austin: The University of Texas. Retrieved 10 16, 2022, from https://repositories.lib.utexas.edu/bitstream/handle/2152/66044/STORY-MASTERSREPORT-2018.pdf?sequence=1&isAllowed=y

Sullivan III, M. (2004). *Statistics: Informed Decisions Using Data.* Upper Saddle River, NJ, US: Prentice Hall

Sun, X., Cai, C., Yang, J., & Shen, X. (2014). *Hindawi: Mathematical Problems in Engineering: High-Performance Computing Strategies for Complex Engineering Optimization Problems.* Retrieved May 19, 2020, from Route Assessment for Unmanned Aerial Vehicle Based on Cloud Model: https://www.hindawi.com/journals/mpe/2014/518120/

Szpytek, P. H. (2022). *10 Of The Hardest Bosses Of All Time*. Retrieved January 2023, from Gameranx: https://gameranx.com/features/id/297195/article/10-of-the-hardest-bosses-of-all-time/

theDeModcracy. (2020). *Top 10 Best Bosses of the Decade (2010-2019)*. Retrieved February 7, 2023, from YouTube: https://www.youtube.com/watch?v=a6-4rb5hZlo

Top5Gaming. (2017). *10 of the Easiest Video Game Bosses!* Retrieved February 7, 2023, from YouTube: https://www.youtube.com/watch?v=A93i5BhFM1g

Tremori, A. (2013). Interoperable simulation for asymmetric threats in maritime scenarios: a case based on virtual simulation and intelligent agents [Abstract]. *International Journal of Simulation and Process Modeling, 8*(2-3), 160-167. Retrieved May 19, 2020, from http://tg5zw4at7r.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=proceeding&rft.title=International+Journal+of+

Truelist. (2023). *Gaming Statistics - 2023*. Retrieved January 2023, from True List: https://truelist.co/blog/gaming-statistics/

WatchMojo.com. (2017). *Top 10 Pathetically Easy Bosses In Video Games*. Retrieved February 7, 2023, from YouTube: https://www.youtube.com/watch?v=hdyganYr8Y4

Wei, L., Zheng, Z., & Kaiyuan, C. (2013). *Science Direct.* Retrieved 5 19, 2020, from Chinese Journal of Aeronautics: Adaptive path planning for unmanned aerial vehicles based on bi-level programming and variable planning time interval: https://www.sciencedirect.com/science/article/pii/S1000936113000927

Appendix A: ACRONYMS

AEC – Army Evaluation Center
AI – Artificial Intelligence
AMD – Air & Missile Defense
ATEC – Army Test & Evaluation Command
CI – Confidence Interval
CPU – Computer Processing Unit
DC – Detective Comics
DET – Likeliness of detection
DoD – Department of Defense

DV –	Dependent Variable
FMEA –	Failure Modes and Effects Analysis
HTC –	Hard to Change
HLTB –	HowLongToBeat.com
IV –	Independent Variable
M&S –	Modeling & Simulation
MIT –	Massachusetts Institute of Technology
NV –	Nested Variable
OCC –	Likeliness of occurrence
REML –	Restricted Maximum Likelihood
RPN –	Risk Priority Number
PO –	Program Office
SDI –	Strategic Defense Initiative
S/HWIL –	Software / Hardware In the Loop
SEV –	Level of severity
TTPs –	Training, Tactics, and Procedures
UAV –	Unmanned Aerial Vehicle
U.S. –	United States of America