Spring 2022

Effects of Gamification on Knowledge Acquisition: Aviation Weather Online Training

Jacqueline Mcsorley
mcsorlej@my.erau.edu

Effects of Gamification on Knowledge Acquisition: Aviation Weather Online Training

By

JACQUELINE ELIZABETH MCSORLEY

A Dissertation submitted to the

Department of Human Factors

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Human Factors

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Spring 2022

Abstract

Due to the high popularity of gaming, researchers have begun to implement aspects of these games into real life (known as gamification). Motivation and engagement may be influenced by certain game mechanics and user types, which then could also impact learning. A gap still exists as to whether leaderboards and narratives impact motivation or engagement when implemented in a non-game context (i.e., an online training program). Studies indicate that aviation weather training for general aviation pilots is underdeveloped, particularly for NEXRAD. In considering the competitive/goal-oriented nature of pilots, gamification elements such as Achievements and Stories could best motivate pilots-in-training through elevated engagement and motivation. Therefore, the purpose of the current study was to determine effectiveness of two types of gamification mechanics—narrative and leaderboard—on motivation, engagement, and learning in a Next Generation Radar (NEXRAD) online training program. In order to test this intervention, the study used a 2x2 between-subjects experimental design. It was hypothesized that there would be a significant effect of narrative presence and leaderboard presence on the multivariate dependent variable (engagement, motivation, and knowledge acquisition). Participants (n = 41) took part in an online NEXRAD training program—administered through Qualtrics—and received either the narrative intervention, leaderboard intervention, both, or neither. Participants were measured on their NEXRAD knowledge acquisition before and after the training. Following the training program, participants reported their motivation, engagement, and reactions to the training. Results from this study did not support the hypotheses. Gaps in this realm of training still exist and closing them is imperative in improving pilots' understanding of the material and overall safety of flight.

Acknowledgments

I first want to say thank you to my advisor, Dr. Beth Blickensderfer, for inspiring my passion for training research and the aviation weather domain. Working with you has given me a world of possibilities for my future endeavors. I am also incredibly grateful for my other committee members—Dr. Thomas Guinn, Dr. Joseph Keebler, and Dr. Shawn Doherty—for supporting me both through the doctoral program and this dissertation. Each of you provided an invaluable piece of expertise that helped me grow as a researcher. Further I want to thank Dr. Eric Muth, my master's thesis advisor, for cultivating my passion for human factors and giving me the opportunity to begin my academic career.

I also want to thank my Applied Cognition and Training Science lab mates, John Kleber and Cassandra Domingo. John, I would have never made it this far in the program without your constant support and calm demeanor. Cass, working with you was a wonderful and unforgettable experience. Another big thank you to Jessica Derby—my roommate, colleague, and friend—for (quite literally) always being there for me. I am grateful to Emily Anania for always inspiring me, both professionally and personally. I also owe gratitude to all my other Embry-Riddle colleagues who encouraged, helped, and annoyed me every chance they got.

Finally, I would like to thank my family. Thank you to Bill, Jeannine, and Gean McSorley for being my proofreaders, life coaches, and cheerleaders. Thank you to my grandfather, Dr. Joseph Bellanti, for awakening and encouraging my pursuit of higher knowledge. And thank you to Pablo Almada for your love and support through every step of graduate school, no matter the distance.

Table of Contents

Table of Figures

Table of Tables

Chapter 1

Introduction

Purpose Statement

The purpose of this dissertation is to determine if one or more gamification elements influence knowledge gain from an aviation weather training program. This chapter will provide a brief overview of the dissertation including the rationale, background, and problem statement.

Background and Rationale

Training is a necessary part of skill development and knowledge acquisition in any domain. In order to improve engagement in the training, gamification is sometimes implemented. Gamification is “the use of game design elements in a non-game context” (Deterding et al., 2011). Elements of games--such as points, badges, leaderboards, narratives, and avatars--have shown to increase knowledge acquisition in some training programs. Mixed evidence exists on what contexts work for gamification, which game elements are effective, and what interactions between elements (if any) are most effective. The current study aims to evaluate two different elements of gamification in the context of an aviation weather training program. Pilots represent a highly motivated population when it comes to learning about flight operations, but they may have low motivation on topics like weather. Recently, a study found that pilots (particularly low flight hour, low-rated pilots) have low scores for interpreting a number of weather products including radar (Blickensderfer et al., 2019). For this reason, elements of gamification will be applied to an aviation weather training program to determine the effects of gamification on knowledge acquisition.

Chapter 2

Review of Related Literature

Introduction

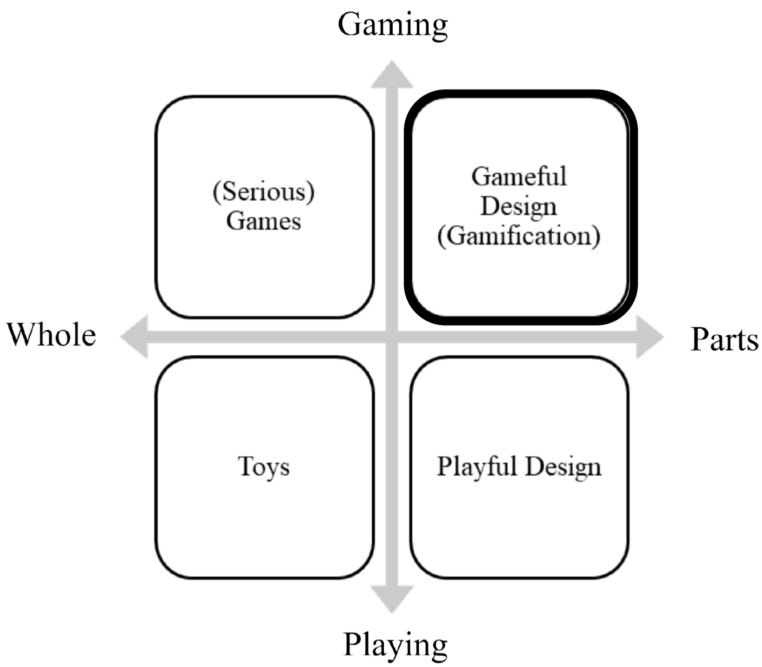
This chapter begins with definitions of the concept of gamification. Subsequently, this section delves into the theory behind motivation and its role in gamification. Once the motivation behind gamification is established, the chapter reviews common gamification implementation strategies and evidence of their effectiveness. The next section provides an overview of the training program development process.

Concept of Gamification

Games, particularly video games, are one of the most common recreational pastimes in America (Bureau of Labor Statistics [BLS], 2015). Due to the high popularity of gaming, researchers have begun to implement aspects of these games into real life. Gamification is commonly defined as "the use of game design elements in a non-game context" (Deterding et al., 2011). The term "gamification" was first seen twelve years ago in the digital media industry (Paharia, 2010); however, this concept has been around since the 1980s—when consumer video games began to increase in popularity—as seen in studies of heuristics for creating more entertaining interfaces (Carroll, 1982). To understand the concept of gamification, it is essential to first break down the distinctions between serious games, toys, playful design, and gamification (see Figure 1). As shown in Figure 1, gamification can be thought of as falling on two continuums: gaming (i.e., goal-oriented, structured) to playing (i.e., non-goal oriented, non-structured) and whole (i.e., an entirety) to parts (i.e., components/pieces rather than an entirety).

Figure 1

Distinction between game/play and whole/parts (Deterding et al., 2011)



Together, the framework produces four quadrants. The "whole-gaming" quadrant consists of serious games. A serious game is "a game in which education (in its various forms) is the primary goal, rather than entertainment" (Michael & Chen, 2005, p. 17). Serious games have been used for centuries, particularly in the realms of military, education, and business. The digital age has added new breadth to the realm of serious games through console or computer-based gaming. Digital serious games are "any form of interactive computer-based game software for one or multiple players to be used on any platform and that has been developed with the intention to be more than entertainment" (Ritterfeld et al., 2009, p. 6). Learning and full game immersion provide the drive for serious games (Deterding et al., 2011). In comparison, the "whole-playing" quadrant consists of toys. Toys also have a full immersion property, but no elements of gaming. Toys do not require rules or competition; their purpose is exploratory recreation (Deterding et al., 2011). Playful design (the "parts-playing" quadrant) implements aspects of playing in order to create a fun atmosphere or structure. Playful design uses parts of entertaining elements instead of a fully engaging condition, which distinguishes it from toys (Deterding et al., 2011). Finally, we arrive at the concept of gameful design, or gamification (the "parts-gaming" quadrant). In order to count as gamification, a design must include the following: (1) characteristics of a game (rather than of play), (2) parts or elements of a game (rather than an entire game), and (3) non-gaming context (rather than an entertainment-centered framework) (Deterding et al., 2011).

Another definition of gamification is the "process of enhancing services with motivational affordances in order to invoke gameful experiences and further behavioral outcomes" (Hamari et al., 2014). Affordances are the properties of an environment with which a person can interact (Kroemer et al., 2001). For example, a doorknob affords turning due to its size and position off the ground. Stairs afford stepping if the person's leg height and physical capabilities allow this activity. Even on a flat touch screen, well-defined shapes afford touching due to their resemblance of analog buttons. In the realm of gamification, it is important that the motivational affordances implemented provide the user with the same psychological outcomes as games (Huotari & Hamari, 2012). Deterding et al. (2011) argues that these affordances must also resemble the ones in games, not just create the same outcome. This is where the idea of gamification elements, aspects of games in a different context, comes into being (Deterding et al. 2011).

In sum, gamification is the application of game-like elements to a non-game situation (Deterding et al., 2011). While we have a succinct definition of what gamification entails, there are still uncertainties of which mechanics are unique to gaming and the level of involvement that constitutes gamification (Groh, 2012). For example, does adding a leaderboard automatically work as a gamification mechanic that motivates individuals to perform better? In considering this question, it is important to examine how gamification relates to research on motivation, particularly the aspects and implementation strategies of gamification that are necessary to motivate users.

Motivation

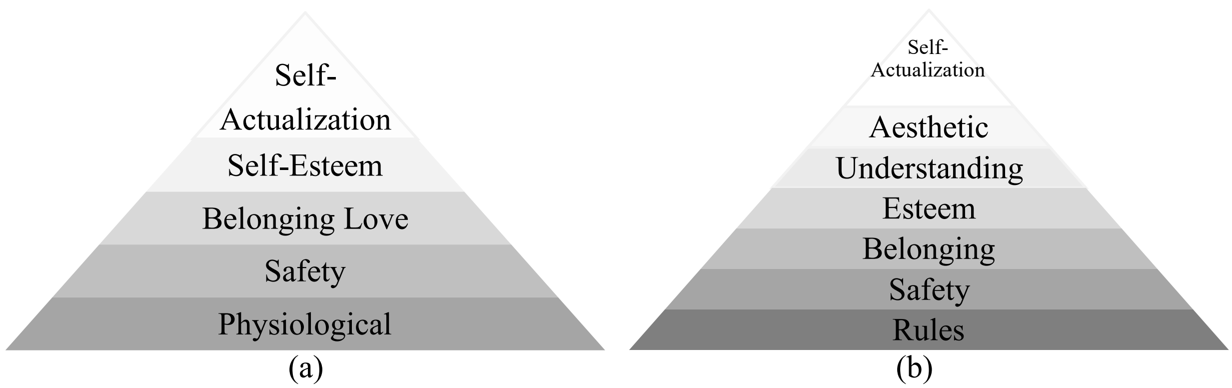
Hierarchies of Needs

Humanity stands out from other beings through our motivation toward self-actualization, a concept that has spanned human understanding since the ancient Greek philosophers (Deci & Ryan, 2002, p. 3). Abraham Maslow (1943) cultivated a hierarchy of needs (Figure 2a), one of the earliest modern theories of motivation. According to Maslow's theory, humans are unmotivated to seek higher tiers of the hierarchy until the lower levels are satisfied (Maslow, 1943). For example, a person would not try to achieve a higher status and recognition (self-esteem) within their job until they have established a sense of connection with their boss or colleagues (belonging love).

In the realm of gaming, users aim to fulfill lower levels of achievement before striving for higher levels (Siang et al., 2003). Siang and Rao (2003) developed a hierarchy of players' needs (Figure 2b) in order to illustrate the progression of factors that motivate players within a game. The lowest level of achievement in a game is understanding the rules that guide basic gameplay. Next, the player requires a sense of safety that allows for persevering through the game. The third level allows players to strive for a sense of belonging categorized by a feeling of comfort and eventual achievement of the goals. After winning becomes an attainable goal, the player can play the game in order to feel good about themselves and their progress. Once the player achieves a sense of esteem, they can develop helpful strategies beyond the basic rules of the game in order to truly understand the game mechanisms. Next, the player desires a set of aesthetics (e.g., sound effects, graphics, and musical score) that positively influence gameplay. Finally, the player attempts to attain near-perfection within the rules of the game. To exemplify this hierarchy, a novice player would first try to grasp the rules of a game before pursuing the goals. Similarly, a proficient game player would require a sense of esteem in the game before further exploring the strategies of successful gameplay.

Figure 2

(a) Maslow's Hierarchy of Needs and (b) Siang & Rao's Hierarchy of Player Needs

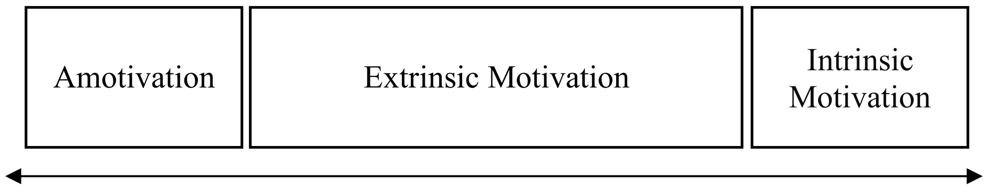


Self-determination theory

The hierarchy of needs and hierarchy of player needs characterize the idea of self-motivation. Self-motivation is influenced by both intrinsic and external sources. Intrinsic motivation stems from engaging in an activity because the person finds it interesting and enjoyable; extrinsic motivation derives from a separate incentive for engagement outside of the actual task (Ryan & Deci, 2000). Intrinsic motivation promotes higher knowledge acquisition, satisfaction, and persistence since the motivation stems from the natural connection of the person and task (Deci & Ryan, 1991). The self-determination theory examines motivation as a spectrum rather than a dichotomous choice (Figure 3). Behaviors toward the right side of the spectrum (intrinsically motivated) are more self-determined than those on the left (amotivated).

Figure 3

Continuum of Motivation According to Deci and Ryan (2002)



Multiple levels exist within the category of extrinsic motivation: external regulation, introjected regulation, regulation through identification, and integrated regulation (Deci & Ryan, 2002). *External regulation* occurs when all motivation stems from outside influences, often in the form of rewards and punishments. Next, *introjected regulation* begins to internalize these influences, usually to avoid feelings of guilt or achieve self-worth. This type of regulation is still primarily motivated through external sources; the motivation, though internalized within the individual, is not fully assimilated with their sense of self. *Regulation through identification* places value on a particular goal. Since the person is endorsing a particular behavior, regulation through identification requires them to rationalize the action as self-determined. Finally, *integrated regulation* occurs when the identifications are joined with a person's core values and goals, but they are performed to achieve outcomes external to one's self worth.

Ryan and Deci (1985) also suggest that we have certain mental needs required for intrinsic motivation: relatedness, competence, and autonomy. *Relatedness* refers to the need to connect with other people. *Competence* is the need to solve problems successfully in a given scenario. *Autonomy* is the need to exercise control over one's life. Game mechanics may fulfill some of these needs (Rigby & Ryan, 2011). For example, having teams or a social network may satisfy the need for relatedness by allowing players to connect with other users and compare their progress to the community. Puzzles, challenges, and goals—often with increasing levels of difficulty—may fulfill the need for competence. Game players may experience feelings of autonomy when they are able to choose their path within a game or even choose the type of games in which they take part.

RAMP framework of Intrinsic Motivation in Gaming

Marczewski (2013) built on the idea of relatedness, competence, and autonomy to develop the RAMP framework specifically for gaming. The RAMP framework consists of four elements: relatedness, autonomy, mastery, and purpose. While the needs for autonomy and relatedness remain the same as per the self-determination theory, the needs for mastery and purpose represent new elements. *Mastery* refers to the need to become skilled at a task (Marczewski, 2013). Successful mastery requires a person's skill level to increase proportionately to the difficulty of the game, which can put users in a flow state (Marczewski, 2013). Flow is "the state in which action follows upon action according to an internal logic which seems to need no conscious intervention on our part" (Csikszentmihalyi, 2014, p. 136-137). While this definition of flow may sound similar to automatic processing from the cognitive psychology literature, flow has the added characteristic of engagement. In other words, during flow, the person experiences an unconscious momentum to continue an activity because of the level of engagement. Flow requires an equilibrium between task difficulty and user skill in order to reach optimal levels of arousal (Csikszentmihalyi, 2014, p. 232). Thus, when attempting to master a skill, it is important to find a middle ground between high difficulty (anxiety) and low difficulty (boredom) (Csikszentmihalyi, 2014, p. 232; Marczewski, 2013).

The final component of the RAMP framework, *purpose*, is the need for our actions to have meaning (Marczewski, 2013). Purpose is similar to the concept of relatedness, but the former focuses more on helping others and less on connecting with others. Altruism, the unselfish regard for the welfare of others, can attribute to the need for purpose within a game (Marczewski, 2013).

Marczewski argues that some element of intrinsic motivation (including relatedness, autonomy, mastery, and purpose) is necessary for gamification to be successful in a system or training program (Marczewski, 2013). The notion theorizes that the existence of certain game mechanics and user types influence user motivation and engagement, which then could also impact learning.

Engagement

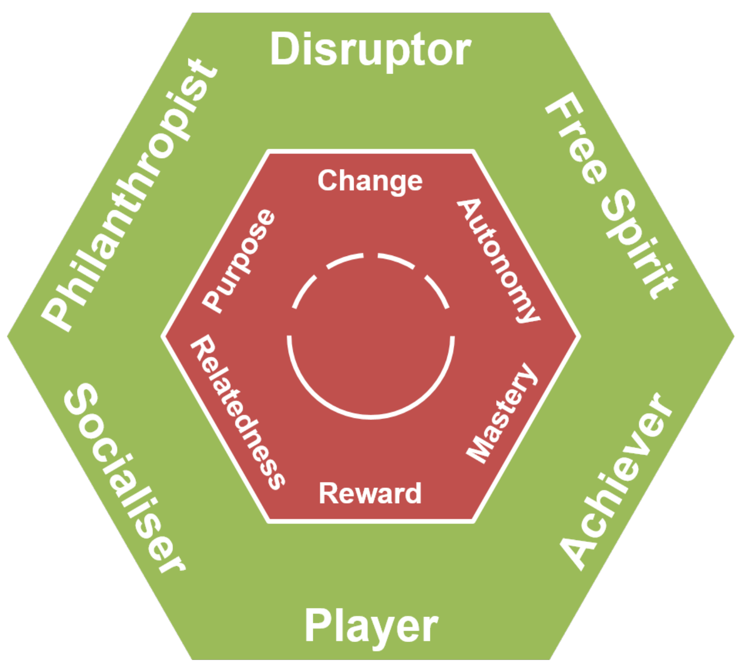
The concept of engagement differs from motivation, but the distinction varies in current literature (Martin et al., 2017). In the *Handbook of Research on Student Engagement*, Christenson and colleagues (2012) propose that motivation derives power from inner drives to fulfil psychological needs (e.g., mastery, relatedness, competence, autonomy), but engagement seems to emerge from external influences and becomes internalized over time. As previously stated, successful mastery can put users in a flow state (Marczewski, 2013). Mastery represents the motivational factor that may lead to a flow state (i.e., engagement) (Csikszentmihalyi, 2014, p. 136-137). Martin and colleagues (2017) define motivation as “the inclination, energy, emotion, and drive relevant to learning, working effectively, and achieving” and engagement as “the behaviors that reflect this inclination, energy, emotion, and drive.” Similarly, the *Handbook of Research on Student Engagement* also posits that motivation consists of affective factors whereas engagement reflects the behavioral factors (Christenson et al., 2012). While there are no “best” game mechanics for increasing engagement and knowledge acquisition, identifying user types could help guide gamification applications (Tondello et al., 2016).

User Types and Motivation

The type of user may influence the effectiveness of a motivation strategy (Tondello et al., 2016). Previous literature suggests that personality impacts the motivational effects of the type of games or gamification elements (Jia et al., 2016; Nacke, Bateman, & Mandryk, 2014; Johnson et al., 2012). Tondello and colleagues developed a 24-item survey to map user preferences to the Hexad of User Types (Figure 4). The scale was validated through internal reliability, test-retest reliability, and factor analysis. According to the model, different user types--based on personality traits within games or systems--could react more positively to different elements of gamification. Figure 4 illustrates the model such that each user type surrounds the exterior border (i.e., disruptor, free spirit, achiever, player, socializer, philanthropist) and motivational attributes align the interior border (i.e., change, autonomy, mastery, reward, relatedness, purpose) (Marczewski, 2016). This model comprises a Hexad of User Elements, whereas a later iteration of the model resembles a dodecad (Tondello et al., 2016). Autonomy, mastery, and relatedness represent the intrinsic motivational elements whereas change and reward are extrinsic motivations.

Figure 4

Hexad of User Types (Marczewski, 2016)



Intrinsic user types—Philanthropists, Achievers, Socializers, and Free Spirits—are driven by an inward desire to perform the activity as dictated by the RAMP framework. *Philanthropists* seek a sense of purpose through helping others and expecting nothing in return. *Achievers* have a need for mastery within a game and constantly try to improve themselves without needing others' validation. *Free Spirits* seek autonomy in a game and feel limited by rules and restrictions. They want to explore, build, and design things without boundaries. Finally, the *Socializers* want to interact with others within a game and are motivated by relatedness to other players (Tondello et al., 2016).

Player (extrinsic) user types (i.e., Self-seekers, Consumers, Networkers, and Exploiters) are not motivated intrinsically and rely on rewards within the game. Player user types resemble the Intrinsic users in their actions, but not the motivation behind these actions. *Self-Seekers* act similarly to Philanthropists in that they assist others via knowledge sharing, but only if there is a reward for their help. *Consumers* resemble Achievers by completing challenges to learn new skills within the game; they are usually striving for achievement to earn prizes for their accomplishments. *Networkers* are Socializers motivated by finding useful contacts to increase their influence and chances of reward. Networkers care more about what they can use from others rather than just trying to establish meaningful connections. Lastly, the *Exploiter* aims for exploration, much like the Free Spirit. They seek loopholes and cheat codes to push the boundaries of the game. Exploiters will also build and create things in a game just to sell them for profit (Tondello et al., 2016).

Finally, the Disruptor user type aims to work against the system or game. Disruptors can act as four different subtypes: Griefer, Destroyer, Influencer, or Improver. *Griefers*, who want to hurt other users, either require a change of mindset or exemption from a gamified program.

Destroyers try to break the system through hacks or loopholes at the expense of others. Sometimes Destroyers do this because they have issues with a particular system, but often they find pleasure in breaking the game. Similarly, *Improvers* find hacks or loopholes in the system or game; their goal, however, is to point out things in a system that need improvement. *Influencers* try to change the way a game or system works by wielding power over other users. Influencers usually advocate on behalf of other users in order to increase the game experience. Griefers and Destroyers act as "Black Hats" or the negative perspective on game disruption whereas Influencers and Improvers represent the "White Hats" or the positive perspective that initiates change within a game (Tondello et al., 2016).

To continue this line of research, Tondello and colleagues conducted a study to determine the correlation of the Gamification User Types Hexad Scale with personality measures (i.e., the Big Five). The results of this study showed significant correlations between Hexad User Types and the Big Five personality traits: philanthropist/socializer with extraversion, achiever/player with conscientiousness, free spirit with openness, and philanthropist/socializer with agreeableness. Based on these user types, different elements of gamification increase engagement with a system or game. Some elements—including onboarding, feedback, themes, and narratives—work for any user type. These are generally the most basic elements of a game and appeal to the majority of users; however, these still require a valid reason for implementation. Other elements require a specific user type to be optimally effective (Table 1) (Tondello et al., 2016).

Table 1

Examples of Best Elements for Specific User Types (Tondello et al., 2016)

User Type	Gamification Elements
Philanthropist	Purpose, Collect & Trade, Gifting, Sharing Knowledge

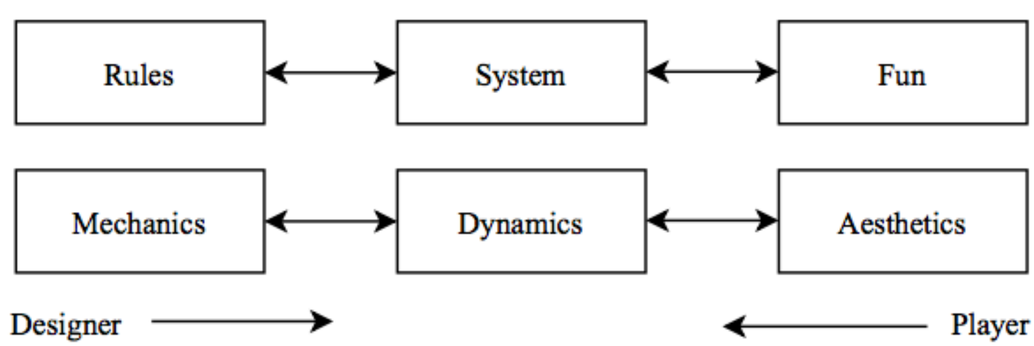
Achiever	Challenges, Certificates, Quests, Levels, Boss Battles
Socializer	Teams, Social Network, Social Status, Competition
Free Spirit	Exploration, Branching Choices, Easter Eggs, Unlockable Content
Player	Points, Experience Points, Leaderboards, Badges, Virtual Economy
Disruptor	Innovation Platform, Voting, Anonymity

Elements of Gamification

The term "elements" often operates as a blanket statement when referring to parts of a game. Currently, there is no consensus on the organization of gamification elements (Hunicke et al., 2004; Tondello et al., 2016; Marczewski, 2018; Zichermann & Cunningham, 2011; Jackson, 2016). Hunicke and colleagues (2004) developed a framework to help designers dissect and analyze the varying parts of a game. The Mechanics, Dynamics, and Aesthetics (MDA) framework distinguishes the basic elements used in game design in order for non-game contexts to utilize these discrete aspects of games (Figure 5).

Figure 5

MDA Framework (Hunicke et al., 2004)



Looking at the left side of Figure 5, the designer thinks in terms of mechanics and rules. *Mechanics* represent the rules and aspects of the game at the algorithmic level (Hunicke et al., 2004). Points, levels, badges, challenges, and many other common gamification "elements" are actually considered game mechanics (Zichermann & Cunningham, 2011). *Dynamics* comprise the users' interactions with those mechanics within the system and aim to create aesthetic experiences (Zichermann & Cunningham, 2011; Hunicke et al., 2004). *Aesthetics* describe the emotions aroused by the game based on the users' interactions, which includes perceptions of "fun" (Zichermann & Cunningham, 2011). From a game designer perspective, the mechanics influence the dynamics, which lead to an aesthetic interpretation by the user. In contrast, users initially perceive the game's aesthetics, which eventually lead to the dynamic interactions, and ultimately see the mechanics that influence these (Zichermann & Cunningham, 2011).

The research included a recommended organizational structure for gamification elements. Tondello and colleagues (2016) sort gamification elements by user type (Table 1). Zichermann and Cunningham (2011) classify elements based on categories and subcategories, such as how points are further broken down into cash scores, experience points, and composite metrics. Jackson (2016) developed categories specifically for the gamification of learning. These categories are Achievement (Progression), Rewards, Story, Time, Personalization, and Microinteractions. Achievements consist of Points, Badges, Leveling, Leaderboards, Progression Bars, and Certificates. These elements give users a sense of accomplishment and feedback on skill improvement. According to Jackson (2016), Achievements also motivate users to continue participating in the activity when they feel they are making progress. Rewards include Tools, Collectives, Bonuses, and Power-Ups. This category is similar to Achievement but can utilize fixed and variable reward schedules to encourage users with additional extrinsic motivation to continue. Stories can be narratives or quests within the system that create a more compelling theme for the activity. Time elements include countdowns, which provide a sense of urgency for the user, and schedules, which create a more structured environment to guide the user through steps. Personalization is a common way to make the program more interactive. This is often implemented through avatar customization, character naming, and interactive conversations within the program. Finally, Microinteractions are used to provide additional interactivity or excitement within the design. These include sound effects, Easter eggs, novel transition screens, and hover/toggle animations (Jackson 2016).

The lack of agreed upon organization structure for gamification elements has not stopped industry from implementing the elements in software. A multiplicity of domains—including commerce, education, health, marketing, and sustainable consumption—use gamification elements in various manners (Hamari et al., 2014). For example, "Duolingo"—a company that helps users learn new languages—awards badges for completing activities and places them on a leaderboard based on their knowledge acquisition. A fitness application called "Zombies, Run!" provides a narrative of the user being chased by zombies. As the story progresses, it prompts the user to run when zombies are near or walk when they are at a safe distance. Table 2 contains additional examples of gamification across different domains. When companies implement gamification into their products or systems, they often lack, or fail to broadcast, their rationale for the specific elements chosen. Future research must delve into the motivation behind using certain elements over others.

Table 2

Examples of Gamification Elements in Different Domains

Domain	Gamification Elements and Source of Implementation
Commerce	Social Network (Venmo), Customization (PayPal)
Education	Badges/Leaderboard (Duolingo), Unlockables/Achievements (Forest)
Health	Levels (C25K), Social Challenges (Apple Activity), Narrative (Zombies, Run!)
Marketing	Prizes (McDonald's), Rewards (Starbucks), Gifting (LinkedIn)
Sustainable Consumption	Points/Competition (SAP), Rewards/Points (RecycleBank), Leaderboard (WeSpire)

Efficacy of Gamification

Gamification may give companies a sense of improvement, but it does not necessarily solve the elemental problems of that business (Zichermann & Cunningham, 2011). Hamari and colleagues (2014) conducted a literature review to determine the application and efficacy of gamification in different contexts and included 24 studies. The most common gamification elements used included points, leaderboards, badges, levels, stories, clear goals, feedback, rewards, progress, and challenges. They found a largely positive support for gamification, but the results depended highly on the context of implementation (e.g., education, commerce, health). The area of application most related to this dissertation is education. The nine studies conducted in the educational domain had positive learning outcomes in terms of motivation and engagement; however, one particular study identified challenges with implementing competition (e.g., leaderboards) in educational contexts (Hakulinen et al., 2013). These challenges may indicate a need for user type analysis (e.g., Tondello et al.'s Hexad scale) in order to determine if competitive gamification will benefit users.

A follow-up review of the Hamari et al. (2014) study summarized the effects of gamification specifically in the realm of education (Majuri, Koivisto, & Hamari, 2018). The researchers collected a total of 128 empirical studies and found similar elements of gamification as utilized in other non-educational domains: Achievement (e.g., challenges, badges, leaderboards), Social (e.g., teams, social networking), Immersion (e.g., avatars, narrative, in-game rewards), Non-digital elements (e.g., check-ins, financial reward, motion tracking), and Miscellaneous (e.g., virtual helpers, health points, adaptive difficulty). The review also discussed the psychological and behavioral outcomes related to the implementation of game elements (Majuri et al. 2018). Psychological outcomes included perceived enjoyment, engagement, affect, flow, effort, workload, frustration, perceived competition, motivation, self-efficacy, and personality. Behavioral outcomes included assignment grade, academic performance, speed, time, experience points, accuracy, leaderboard position, participation activity, knowledge transfer, stress levels, psychophysiological measures, and cooperation. Since this literature review was education-focused, many of the behavioral outcomes (e.g., academic performance, grades) differed from those measured in other domains (e.g., time spent, calories burned).

The most commonly implemented gamification elements in the education research included points, leaderboards, badges, challenges, levels, teams, progress, social networking, performance feedback, timer, narrative, and avatars (Majuri, Koivisto, & Hamari, 2018). Although not a meta-analysis, the authors aggregated the positive (gamification yielded higher scores than comparison group), null (no difference between gamification and comparison group), or negative (gamification yielded lower scores than comparison group) results. Points (52 studies) had the most positive results by a vast majority, but also the most null results. Similarly, leaderboards (43 studies) and badges (39 studies) yielded a high number of positive results along with more than ten null and several negative findings each. Narratives (11 studies) had all but one positive result among the group of studies that utilized this element. Based on the successful findings from the studies that utilized Narratives, Majuri and colleagues (2018) recommended that future researchers implement more immersion-based elements in the realm of education. One striking limitation to the majority of these studies involved the use of overlapping gamification elements. Using gamification elements in conjunction with each other limits what researchers can glean from each individual element. In response to this shortcoming, the authors recommended that future studies isolate the elements in an educational setting in order to control for overlapping elements (Majuri, Koivisto, & Hamari, 2018).

Another literature review in the educational domain focused on gamification in e-learning (Rozman & Donath, 2019). The idea that gamification can be easily implemented into e-learning served as the motivation for conducting this review. The researchers examined nineteen prior literature reviews (consisting of 2631 studies) and found increasing numbers of articles regarding gamification and e-learning as technology improves. They also discovered that the majority of these empirical studies found primarily positive results for motivation, performance (measured as a learning outcome), attitudes, engagement, and social interactions. These affective measures were largely similar to the outcomes measured in other domains (e.g., motivation and engagement). The most commonly implemented gamification elements included points, badges, leaderboards, and levels. These elements are often used due to their design simplicity and relatively simple implementation in comparison to other gamification elements (Rozman & Donath, 2019).

Based on the evidence seen in these summative works, some of the most popular and successful gamification elements include points, badges, leaderboards, and narratives (Hamari et al., 2014; Majuri, Koivisto, & Hamari, 2018; Rozman & Donath, 2019). The subsequent sections will highlight instances in which the implementation of these elements benefit users.

Points and Badges

Points, one of the most commonly used elements of gamification, gives value to users for completing or achieving tasks (Zichermann & Cunningham, 2011). Points exist in sporting events, video games, test scores, shopping rewards, and many other aspects of life. The most popular ways to implement points include experience points (XP), redeemable points, skill points, karma points, and reputation points. Experience points reflect the rank, proficiency, or expertise of a user. Points provide a form of immediate feedback, which allows the user to see their progress along the way (Kapp, 2012). These points cannot act as currency within the game or system, but they continue to increase as the user puts in more time. Redeemable points form a virtual economy within a game or system that allows users to exchange points for items or perks. Skill points, often related to experience and redeemable points, reward users for completing certain activities or goals. Karma points, in comparison, allow users to give points to other users without any direct benefit to the former. Finally, users can earn reputation points for positive interactions with other users to incentivize trustworthy behavior. Karma points and reputation points exemplify how points can be used in a social context. Regardless of the implementation method, it is important to assign worth to particular tasks or interactions to value activities differently. For example, mini-quizzes could count for 10 points whereas tests count for 50 points. This allows users to understand the significance of some activities over others within the system (Zichermann & Cunningham, 2011).

While points incentivize core activities within the game or system, badges provide visual rewards for completing tangential activities (Hamari, 2017). From a design perspective, a badge must consist of the graphical/ textual elements, the reward of the earned badge, and the conditions in which the badge is earned. Badges represent one of the most popular gamification elements, particularly in the educational domain.

Leaderboards

Leaderboards differ from other elements of gamification in that they evoke a sense of competition in users (Koivisto & Hamari, 2014). Users enjoy interacting with leaderboards to top the scoreboard, receive attention, and gains status over other users (Zichermann & Cunningham, 2011). Leaderboards provide competitive feedback on overall performance—as opposed to more immediate feedback from points—in order to help users learn from mistakes and improve future performance (Athanasopoulos & Hyndman, 2011). While there are numerous ways to structure a leaderboard, two distinct designs exist: no-disincentive and infinite (Jia et al., 2017). No-disincentive leaderboards place users in the middle of the scoreboard by showing scores immediately above and below that user. This allows users to view the immediate better scores they need to beat to improve rather than just viewing the top scores. Infinite leaderboards recognize that new scores will eventually succeed old scores, but it is not feasible to display every score. These leaderboards often employ tiered scoreboards (e.g., local leaderboards or experience-separated leaderboards) where users are only shown a select number of other users' scores.

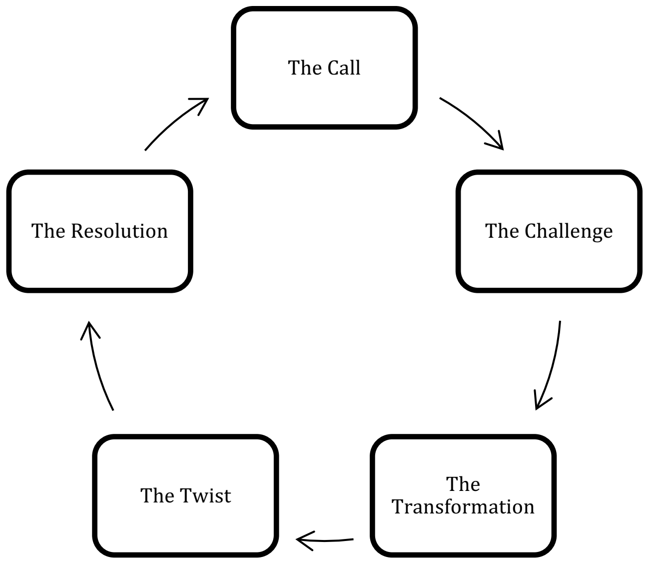
Mixed evidence exists as to whether leaderboards effectively improved measured outcomes, but some studies suggest that individual differences impact these outcomes (Höllig et al., 2018). Trait competitiveness— "the enjoyment of interpersonal competition and the desire to win and be better than others"—represents one characteristic that can dictate interactions with leaderboards (Spence & Helmreich, 1983; Höllig et al., 2018). More specifically, interpersonal success (one underlying construct of trait competitiveness) guides users to aim for personal development rather than simply winning (Höllig et al., 2018). A study by Höllig and colleagues (2018) found that leaderboards serve as an indicator of user progress in comparison to that of other users, which relates to the personal development aspect of trait competitiveness. Similarly, a study that supported the effectiveness of leaderboards in a classroom setting theorized that leaderboards work well in America since western culture supports individualism and competition (Paisley, 2013). Another study, conducted in the domain of education, found that leaderboards evoked higher interest and enjoyment in male students over female students (Gurjanow & Ludwig, 2017). Some users report negative affects toward elements of gamification that encourage competition (Hamari et al., 2014). According to a survey conducted by Jia and colleagues (2017), this negative perception of leaderboards particularly relates to those who are ranked lower on the leaderboard. On the other hand, users that score higher on extraversion (as evaluated by the Big Five personality assessment) often report a more positive experience regardless of their position on the leaderboard (Jia et al., 2017). The survey also noted that the context of implementation impacted the perceptions of leaderboards: users seem to prefer leaderboards in fitness apps over social networking environments (Jia et al., 2017). These findings indicate that leaderboards are preferred in naturally competitive settings (e.g., athletic activities) and when the users are inherently competitive. A recent study conducted by Johnson and colleagues found that the combination of leaderboards and performance gauges did not yield a significant result over the control group in the context of Naval Submarine Officer Basic Course students; however, this study did not assess leaderboards on their own to determine effectiveness (Johnson, Bailey, Mercado; 2020). Based on this evidence, a gap still exists as to whether leaderboards impact motivation or engagement when implemented in a non-game context.

Narratives

Narratives represent an element of gamification that, in the context of training, immerse the learner in a compelling story (Jackson, 2016). These stories allow us to shape and recall information within a particular context (Pujolà & Argüello, 2019). A narrative can provide motivation for learning by creating a game element for the class structure (e.g., a geography class could create teams based on countries around the world) (Witte et al., 2017). In order to create an effective narrative, certain elements must exist in the story: a hook/elevator pitch, central conflict, mystery, and character/backstory details (Bell, 2018). Palomino and colleagues stated that "the Gamification narrative element can be understood as the process in which the user builds his own experience through a given context, exercising their freedom of choice in a given space or period of time, bounded by the system's logic" (Palomino et al., 2019). Marczewski (2019) recommends following a "Soap Hero's Journey" to guide the story (Figure 6). This narrative formula is comprised of a call to action, a challenge, a transformation, a twist, and a resolution.

Figure 6

Soap Hero's Journey Diagram.



One recent study examined how narratives affected learning perceptions and knowledge acquisition (Armstrong & Landers, 2017). Armstrong and Landers (2017) found that perceptions of learning were significantly higher for the narrative condition (compared to a control group); however, knowledge scores were also lower than those in the control group. This demonstrates how narratives can increase engagement with training, but there may be other consequences to its implementation. Another study assessed how a full story differed from a detached scenario (Pujolà & Argüello, 2019). They found that most instructors preferred using scenarios, but the stories allowed for better integration into the teaching tasks. Miranda (2015) implemented a simulated inspection task to engage participants with a vigilance task. Interestingly, storytelling was most successful at encouraging less motivated participants, but not as effective overall. In sum, the efficacy of Narratives still remains unknown. Further research must isolate narratives as a variable and utilize a full story plot (rather than just a scenario) in order to move closer to a more concrete finding.

Summary

As outlined, gamification has potential to increase motivation, engagement, and learning outcomes based on the elements used and environment. Prior research indicates that implementing game elements seems to have a positive impact particularly in the educational domain (Hamari et al., 2014; Majuri, Koivisto, & Hamari, 2018; Rozman & Donath, 2019). Points, leaderboards, narratives, and badges are some of the most effective and widely used elements in previous literature. In order to use gamification as a training tool, it is important to first identify the steps required to develop an effective training program.

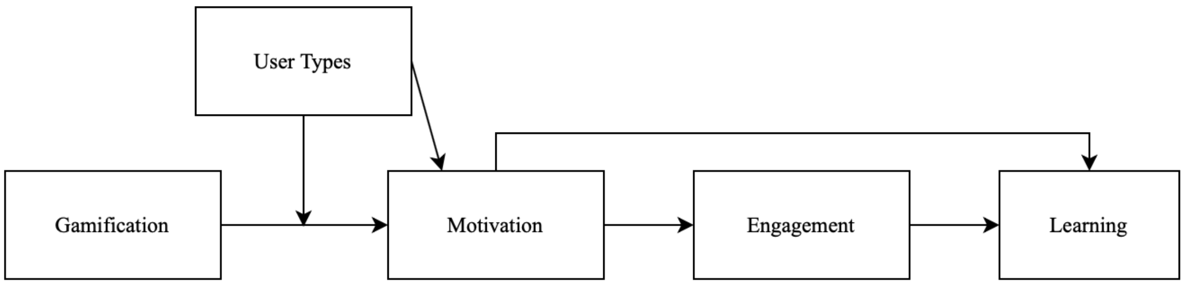
Theoretical Model

A conceptual model of the relationship between *gamification, user types, motivation, engagement,* and *learning* is shown in Figure 7. Beginning at the left-hand side of the model, *game mechanics* represent the different elements of a game that can be implemented in a non-gaming context (Deterding et al., 2011). Next we have *motivation*, "the inclination, energy, emotion, and drive relevant to learning, working effectively, and achieving" (Martin et al., 2017). *Game mechanics* influence how motivated a person is; however, this relationship may be modified by their specific *user type*. Previous literature suggests that personality (i.e., *user type*) impacts the motivational effects of the type of games or gamification elements (Jia et al., 2016; Nacke, Bateman, & Mandryk, 2014; Johnson et al., 2012). This relationship between game mechanics, user types, and motivation also impacts user *engagement*. *Engagement* represents "the behaviors that reflect [the] inclination, energy, emotion, and drive" behind motivation (Martin et al., 2017).

Ultimately, the affects (*motivation*) and behaviors (*engagement*) of the user impact their *learning* capability.

Figure 7

Conceptual Model of the Impact of Gamification on Learning



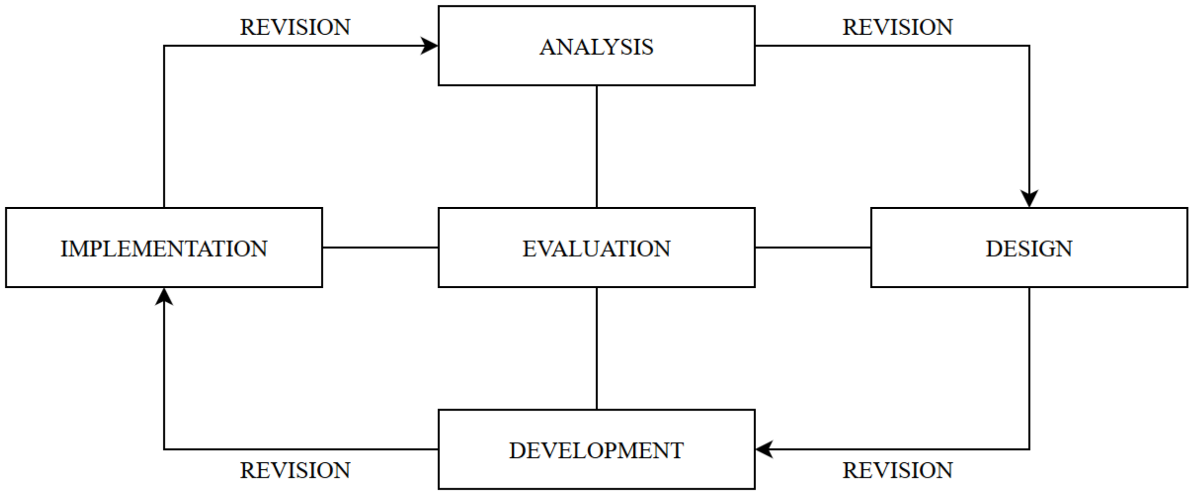
Training

Overview

Training is the "systematic acquisition of knowledge, skills, and attitudes [KSAs] that together lead to improved performance in a particular environment" (Salas et al., 2006). Simply put, trainers implement some sort of intervention in order to evoke a cognitive or behavioral change. According to Salas and colleagues (2006), effective training should let trainees learn the desired KSAs, have opportunities to practice these KSAs, and receive prompt feedback on their performance. A popular framework used in performance-based training design is the "Analysis, Design, Development, Implementation, and Evaluation" (ADDIE) Model of Instructional Design (Figure 8) (Branch, 2009). The ADDIE framework aims to create new and effective approaches to facilitating learning.

Figure 8

ADDIE Model of Instructional Design



Training Analysis

The first step in performing a training analysis is the need analysis (Goldstein & Ford, 2002). Needs analyses allow trainers to determine where training is necessary, what KSAs need training, and who specifically needs training (Goldstein, 1993). Three main components comprise a standard needs analysis: organizational, task, and person analysis. The organizational analysis identifies aspects of the organization that influence delivery strategies of training materials. This includes company goals, resources, limitations, and organizational support for training transfer (Goldstein 1993). Such restraints could pose a threat to proper training design or transfer (Salas, 2001). Next, the task analysis homes in on the specific job of interest. The task analysis phase is essential for categorizing details of the job in order to create useful learning objectives and identify needed KSAs (Salas et al., 2006). Particularly, it is difficult to observe knowledge and attitudes, so it is important to pinpoint these within the task analysis. Lastly, a person analysis determines exactly who needs to be training and what KSAs they are lacking. Person analyses often reveal that different people within the organization (e.g., different departments or job profiles) require different training. This type of analysis is important for identifying trainees' motivation to begin training and competencies to acquire the KSAs (Salas et al., 2006).

According to the ADDIE model, the "Analysis" phase requires developers to find current gaps in performance within the organization (Branch, 2009). It is also essential to identify the instructional goals of the organization, the intended trainees, and the necessary resources needed to complete the training. Additionally, the training developers must determine prospective delivery methods and create a general plan for the program (Branch, 2009). In sum, a training needs analysis identifies deficits within the organization, task, and individuals in order to isolate specific shortfalls (Salas et al., 2006). The "Analysis" phase sets a precedent for the "Design" and "Development" phases of the training program.

Training Design and Development

During the design stage, the training specialist selects the instructional strategy and plans how to implement that strategy (Allen, 2006). This includes reviewing previous training literature, assessing media options, and determining the instructional methods. Trainers must also determine if the training benefits more from face-to-face instruction or online methods. The design stage utilizes planning tools, such as storyboards, to map out the potential training sections. During the design stage, trainers must also consider how to provide useful feedback to trainees throughout the program. Additionally, the design stage dictates which gamification elements best suit the training.

In turn, the development stage involves the actual generation and construction of the materials. Training developers also create course parameters and lesson plans during this phase (Noe, 2006). The course parameters provide an overview of elements such as the purpose, lesson length, target audience, prerequisites, course format, room arrangement, and necessary equipment. The detailed lesson plan expands on the course overview in order to create a guide for trainers to aid in information delivery. It breaks down the entire lesson into sections in order to detail the topic, instructor's job, trainee's activity, and the estimated time that section will last.

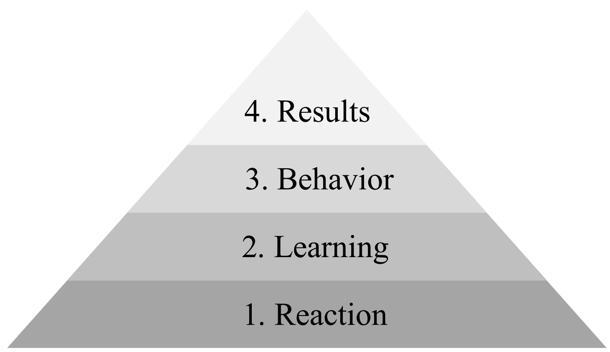
Training Evaluation

Training evaluation is an essential step in ensuring the training program yields effective outcomes (Bates, 2004). Outcomes refer to the measures used by the trainer to determine if the training accomplished what it set out to do (Noe, 2005). One popular training evaluation perspective emphasizes the significance of a construct-oriented approach (Kraiger, Ford, and Salas, 1993). The most commonly measured training outcomes include cognitive outcomes, skill-based outcomes, affective outcomes, results, and return on investment (Kraiger, Ford, and Salas, 1993; Noe, 2006). Cognitive outcomes assess the level of familiarity that trainees have with knowledge-based information. This includes verbal knowledge, organizational knowledge, and various cognitive strategies (Kraiger, Ford, and Salas, 1993). Skill-based outcomes measure certain behaviors or motor skills necessary for the job. Affective outcomes include trainee reactions to the training or attitudes that may have changed due to the training, which includes motivation, self-efficacy, and goal setting (Kraiger, Ford, and Salas, 1993). Self-efficacy refers to one's belief in their performance that influences the ability to control events in their lives (Bandura, 2010). Next, results examine how the training impacted the company in regard to equipment downtime, employee turnover, accidents, and customer service (Noe, 2006). Finally, the return on investment determines whether the training's benefits outweigh the costs.

One of the most popular frameworks in training evaluation is the Kirkpatrick Model (Kirkpatrick, 1976). It addresses the necessity for systematically approaching and understanding training evaluation (Shelton & Alliger, 1993). This framework consists of four levels of analysis: reaction, learning, behavior, and results. *Reactions* include trainee responses to the training based on how engaging or relevant they find the information. *Learning* measures how well the trainees acquired the intended KSAs from the training program. *Behavior* refers to the trainees' application of the KSAs in their actual job. Finally, *results* highlight the degree to which the training brought about the intended outcomes. As illustrated in Figure 9, the Kirkpatrick Model requires lower levels of criteria must yield positive outcomes before progressing to measure higher levels. For example, if participants' initial reactions to the training appear negative, the researchers may consider making the training more enjoyable for the learners before assessing any learning outcomes.

Figure 9

Kirkpatrick Model of Training Evaluation



Critics of the Kirkpatrick Model argue that the framework is incomplete (Bates, 2004). Namely, the framework lacks acknowledgement of the organizational, individual, and training design factors (Cannon-Bowers, Salas, Tannenbaum, & Mathieu, 1995). These factors can have a significant impact on the effectiveness of training before, during, and after the intervention. Another critique of the Kirkpatrick Model states that the framework assumes a causal relationship between reactions and learning (Bates, 2004). Several meta-analyses pertaining to the Kirkpatrick Model found very little evidence that causality between levels of the model exists (Alliger & Janak, 1989; Alliger, Tannenbaum, Benett, Traver, & Shotland, 1997).

Training evaluation often contains two parts: formative evaluation and summative evaluation. *Formative evaluation* ensures good organization of the training as well as trainee satisfaction. *Summative evaluation* occurs at the end of the training and determines if trainees' KSAs changed due to the intervention. It is also essential to choose an evaluation design that best fits the needs of the training and organization (Table 3) (Noe, 2005). For example, a "pretest/posttest" design--which tests trainees before and after the training--is low cost, low time commitment, and still a fairly strong design. A "posttest with comparison group" design does not implement a pretest and instead uses a comparison group to determine what outcomes developed in the training condition. This design is slightly higher cost and time than the "pretest/posttest" design and still only yields moderate reward. A "pretest/posttest with comparison group" design combines aspects from the other two designed mentioned. It utilizes a pre- and post-test as well as a comparison group in order to assess training outcomes. This design is about the same cost and time as the "posttest only with comparison group" and is even stronger in regard to limiting threats to validity (Noe, 2006).

Table 3

Comparison of Evaluation Designs (Noe, 2005)

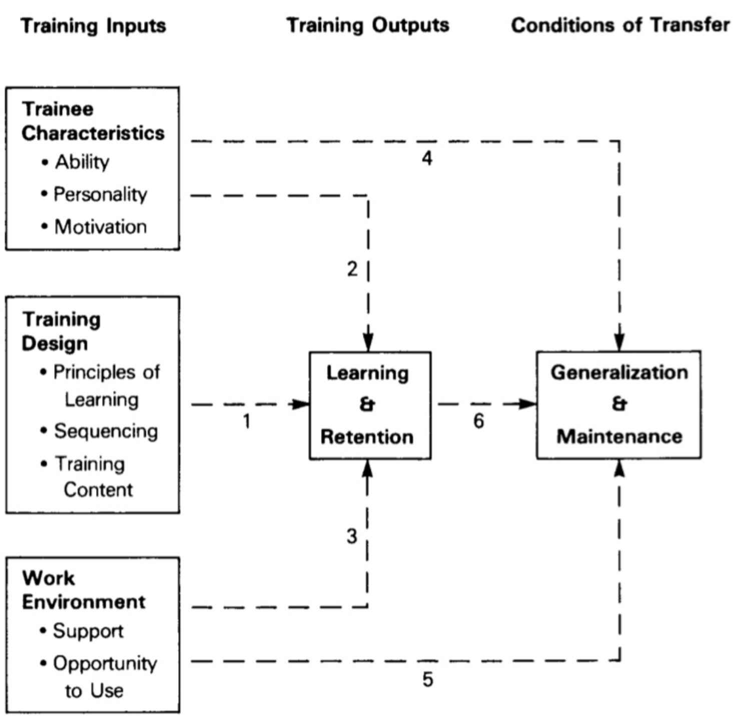
		Measures				
Design	Groups	Pretraining	Posttraining	Cost	Time	Strength
Pretest/ Posttest	Trainees	Yes	Yes	Low	Low	Medium
Posttest Only with Comparison Group	Trainees and Comparison	No	Yes	Medium	Medium	Medium
Pretest/ Posttest with Comparison Group	Trainees and Comparison	Yes	Yes	Medium	Medium	High

Training Transfer

Transfer of training represents another important step in the training process (Noe, 2005). This refers to the utilization of KSAs developed during the training over an extended period of time. A popular model of the transfer process (Figure 10) takes the trainee characteristics, training design, and work environment into account for successful transfer (Baldwin & Ford, 1988).

Figure 10

Model of Training Transfer Process (Baldwin & Ford, 1988)



Trainee characteristics include trainees' motivation to perform the training as well as their basic cognitive or physical ability to complete the training (Noe, 2005). For example, the trainees must be able to read the training materials in order to successfully complete a written training program. Furthermore, the design of the training program must be conducive to learning by providing specific learning objectives, appropriate feedback, and opportunities to practice (Noe, 2005). Additionally, the work environment must accommodate the training through opportunities to use the acquired KSAs and support from company managers, coworkers, and technology.

Trainee characteristics, training design, and the organization’s environment all impact the initial learning and retention of the material. They also affect the generalization and maintenance of the transferred KSAs (Baldwin & Ford, 1988). Generalization refers to the ability to approach new job-specific problems with the KSAs obtained during training. In other words, being able to generalize the trained capabilities in a situation similar to that of the training. Maintenance is the continuous practice and application of the new KSAs over time (Noe, 2006). Maintenance is essential in order to reduce the risk of transience and aid retention of the newly acquired capabilities.

The foundational element of training transfer lies in the type of transfer: near, far, or both (Noe, 2006). Near transfer utilizes training strategies that are identical to the situations the trainee will encounter on the job. Equipment usage represents a skill that benefits from near transfer (Noe, 2006). The theory of identical elements ensures that near transfer occurs (Noe, 2006). The theory of identical elements dictates that the training should mirror the task in regard to equipment, work environment, and any other factors (Thorndike & Woodworth, 1901). This type of training promotes near transfer by structuring the training as similarly to the task at hand as possible (van der Locht, van Dam, & Chiaburu, 2013). On the other hand, far transfer refers to the application of training to a more generalized work environment (Noe, 2006). This occurs when the training environment—including the equipment, tools, and tasks—is different from what trainees will encounter in the real scenario. The stimulus generalization approach supports the concept of far transfer by emphasizing broad principles rather than specific procedures (Noe, 2006). Far transfer is essential when the training cannot directly reflect on-the-job work, such as with interpersonal skills. Additionally, the cognitive theory of transfer proposes the utilization of both near and far transfer. It emphasizes both specific, meaningful material as well as schemas that encourage effective storage of the general content. This theory also states that providing trainees with application assignments can increase the chances of long-term recall (Noe, 2006). Appropriate conditions for the application of the cognitive theory include nearly all types of training and environments.

Summary

In short, training consists of some intervention that evokes a cognitive or behavioral change within the trainees. The training design process guides designers in effective needs assessment, training intervention development, and evaluation of the program (Branch, 2009). Proper needs assessment allows the researchers to determine which knowledge, skills, and attitudes require improvement through the training process. Once the needs assessment identifies gaps in the KSAs, the design stage permits the training specialists to select an instructional strategy and plan how to implement that strategy (Allen, 2006). Following the design and development of the training program, training evaluation determines if the training accomplished what it set out to do (Noe, 2005). Measurable outcomes (e.g., cognitive outcomes, skill-based outcomes, affective outcomes, results, and return on investment) allow researchers to gauge the success of the training (Kraiger, Ford, and Salas, 1993; Noe, 2006). Proper guidance in training development is essential in a number of high-risk domains (i.e., military, medical, and aviation). Training program developers must determine which KSAs, implementation strategies, and evaluative measures best fit a particular domain. Aviation instruction, particularly on topics related to weather, represents a field that may benefit from reformed training.

Aviation Weather Domain

Aviation Weather-Related Accidents

Weather-related accidents have a high probability of fatalities in General Aviation (GA) (Air Safety Institute, 2019). General aviation largely refers to aviation operations that do not fall under military or commercial classifications. Many of these accidents occurred during GA personal operations during the day (FAA, 2010; Fultz & Ashley, 2016; Air Safety Institute, 2008). Often, these accidents occur when pilots fly Visual Flight Rules (VFR) in Instrument Meteorological Conditions (IMC) (Air Safety Institute, 2019). This phenomenon refers to pilots attempting to look out the cockpit window to navigate (as opposed to relying on navigational instruments) despite reduced visibility conditions. Additionally, hazardous weather phenomena—including wind, low visibility, high density altitudes, carburetor icing, downdrafts, heavy precipitation, turbulence, icing, thermal lift, extreme temperatures, and lightning—may also significantly affect the safety of GA flights (FAA, 2010; Fultz & Ashley, 2016). Fultz and Ashley (2016) cited certain areas of concern for GA pilots in regard to aviation weather phenomena. They found that temperature, humidity, and pressure led to 20% of weather-related accidents with carburetor icing and high density altitude as the most prevalent causes.

Aviation Weather Training

In order to reduce these accidents, it is prudent to review current aviation weather instructional techniques and consider tools that may be useful for aviation training. Currently pilots receive aviation weather training as part of obtaining flight certification(s) or rating(s) (FAA, 2020). However, there are no mandates for GA pilots to obtain training on aviation weather as part of recurrent training (i.e., training to remain current after completing their certificate(s) or rating(s)). This makes their initial aviation weather training vital to their long-term success of avoiding weather-related accidents. Relevant weather information needed for aviation weather training can be found in Advisory Circular (FAA, 2020). The Advisory Circular, released by the Federal Aviation Administration (FAA), describes weather in terms of weather products. Weather products include charts, graphs, and codes used to convey weather information to pilots (e.g., Satellite, Radar, METAR, Winds Aloft, and numerous others). Although the basic required aviation weather knowledge comes from the Advisory Circular, the method of training this information varies from program to program regardless of whether the pilot attends a Part 61 or Part 141 program. According to Guinn and Rader (2012), sizeable gaps exist in the required meteorology coursework within college aviation degree programs. One program Guinn and Rader examined required a minimum of one meteorology course, and this course did not have to be aviation-specific meteorology. Further, it was unclear whether or not the aviation weather courses also provided a theoretical background to the products. Accompanying weather theory with product interpretation may improve pilots' ability to understand weather product readings (Guinn & Rader, 2012). With the high level of variability in training and a high fatality rate amongst weather-related GA accidents, it is essential to develop best practices that emphasize the significance of weather knowledge and find ways to engage pilots-in-training.

Aviation Weather Product Overview

When considering aviation weather knowledge, this information generally falls into one of three categories: weather phenomena, weather hazard products, and weather hazard product sources and their applications (Lanicci et al., 2020). Table 4 further breaks down this classification of aviation weather information.

Table 4

Weather products and their classifications according to Lanicci et al. (2020)

Category	Description	Examples
Weather Phenomena	Basic concepts and meteorological theory	Clouds, fronts, cyclones, jet streams, atmospheric phenomenon
Weather Hazard Products	Text-based and graphical information displays generated by FAA-approved sources	Meteorological Reports (METARs), Significant Meteorological Information (SIGMETs)
Weather Hazard Product Sources/ Applications	Classification of official product sources according to the current Advisory Circular (AC)/ how pilots use the information they view	Non-standardized graphical METARs, radar charts with different symbology or color schemes

Weather hazard products play an essential role in assisting pilots in preflight and inflight operations (Parson et al., 2005). Within this category of weather knowledge, weather products can be further divided into Observation, Analysis, and Forecast products. *Observation* products retrieve raw weather data from airborne or surface level sensors (FAA, 2016). For example, METARs and Radar provide meteorological information regarding the current weather phenomena. METARs report current weather conditions at a specific airport. Radar reports the presence or absence of precipitation in a particular area. *Analysis* products generate representations of the observed data in order to relay the information to users (FAA, 2016). Surface analysis and Ceiling Visibility Analysis (CVA) represent two graphical Analysis products that convey current weather conditions. Surface analyses report pressure systems and fronts across an area. CVAs relay information related to ceiling and visibility in an area. Lastly, *forecast* products predict potential weather phenomena based on meteorological observations and modeling (FAA, 2016). Convective/Non-convective SIGMETs and Graphical Airmets (G-Airmets) both calculate when certain hazardous weather may occur. SIGMETs illustrate predictions for severe turbulence, thunderstorms, icing, and volcanic ash. G-Airmets report forecasted ceiling, visibility, and freezing levels.

The Federal Aviation Administration (FAA) recommends a 3-P model to guide GA pilots' preflight weather planning and in-flight decision making: *perceive, process,* and *perform* (FAA Safety, 2006). *Perceive* includes looking for hazards that could harm the flight in a variety of weather products. *Process* consists of interpreting the information received by the weather products to form a big picture of the potential weather. *Perform* requires pilots to act in a way that avoids or diminishes potentially dangerous elements. When examining the weather products, if pilots are not able to develop an accurate understanding of the weather (i.e., "perceive and process") and/or plan weather appropriate flight routes (i.e., "perform"), the door may be open for those pilots to encounter unexpected hazards.

Current Issues with Radar

Despite the valuable information provided by weather products, pilots are still struggling with products such as radar (Blickensderfer et al., 2018). Current literature indicates a gap exists between what the weather products aim to broadcast and what pilots glean from it. One area of research interest has been the effectiveness of Next Generation Weather Radar (NEXRAD) to communicate aviation weather hazards to GA pilots. NEXRAD "makes conventional reflectivity observations and also uses the 'Doppler effect' to measure motion of clear air and atmospheric phenomena within storms" (NOAA, 2019). Beginning nearly two decades ago, research has found indication of pilots misinterpreting convective weather scenarios. First, Latorella and Chamberlain

(2002) assessed Graphical Weather Information Systems (GWISs) in order to determine potential usability issues in these systems. They discovered that pilots were using GWISs to fly closer to hazardous weather.

In a follow-up study, Latorella and Chamberlain (2004) found that higher resolution NEXRAD displays prompted pilots to fly closer to areas of bad weather in order to "tactically" avoid it. Additionally, Beringer and Ball (2004) assessed the effects of NEXRAD display resolution on pilots' decision to fly into weather. Similar to Latorella and Chamberlain (2002; 2004), they found that higher resolution displays prompted pilots to fly closer to severe weather due to higher confidence in being able to narrowly avoid the weather. The pilots did not factor in system latency and short-notice weather changes in their judgments of weather severity. A decade later, Knecht (2016) also reported deficiencies in the capability of NEXRAD displays to communicate information to GA pilots. This work emphasized the importance of implementing future weather predictions and a range ring in order to improve judgments of closest point approach. From a training perspective, research has demonstrated gaps in pilots' understanding of NEXRAD, but that pilots did benefit from short training courses (Cobbett et al., 2014; Blickensderfer et al., 2016).

Potential for Gamification in Aviation Weather Training

As theory points to gamification as a provider of intrinsic motivation, gamification has potential to increase pilot's level of engagement if used in weather training. In the realm of aviation weather training, it is essential to keep in mind what motivates pilots through their training. According to Chidester and colleagues (1991), pilots-in-training are highly motivated to perform their best. Training related directly to flying is intrinsically motivating (Chidester et al., 1991), but motivation to complete weather training may be more extrinsic. That is, GA pilots in training tend to view weather as a necessary but not that interesting aspect of flight training (Robert Thomas, personal communication). Hence, pilots may require some external encouragement to appreciate the significance since weather-related coursework. For example, to fulfill the requirements of the biannual flight reviews, pilots have the option to complete an hour of ground school or take a Wings-accredited course. While they are only required to take one Wings course, they receive a commemorative pin denoting their level of course completion. This reward provides extrinsic motivation. Perhaps a similar type of external motivation (such as gamification elements) could be implemented during initial weather training. If the allure of a pin motivates pilots to complete training, adding carefully considered gamification elements may achieve similar results.

Research indicates that pilots are a highly competitive subset of the population. The "pilot persona" categorizes most pilots as competitive, adrenaline-fueled, and goal-oriented (Weiss, 2016). Considering these descriptions of pilots in conjunction with Tondello et al. (2016), it is likely that GA pilots would fall in the "Achiever" category. Specifically, the drive to master a skill motivates Achievers, and furthermore, achievers aspire to be the best at a given skill (Tondello et al., 2016). Thus, the implication for aviation weather training is that implementing a competitive element of gamification may resonate with GA pilots' need for mastery and relatedness. Competition is not a tool that works in any training domain; the group must work well under competitive pressure. While this generalization is not enough to encompass the entire GA pilot population, it could provide a starting point for identifying the elements best implemented for each user type.

In considering the competitive/goal-oriented nature of pilots, gamification elements such as Achievements and Stories could best motivate pilots-in-training through elevated engagement and motivation. Specifically, implementing a narrative could provide context for this type of training and give pilots a sense of purpose for completing the training (Miranda, 2015). A narrative or story can increase engagement in the training, which may eliminate the need for other external reward (Miranda, 2015). Based on what is known about pilot personas (i.e., their highly competitive, goal-oriented attitudes), a leaderboard could also provide encouragement to perform better based on their competitive nature.

Further research is required to determine if interactions between gamification elements (e.g., narrative, leaderboard, or both) generate stronger training effects, both for knowledge acquisition and retention. One method includes investigating gamification element interactions in a basic psychological study (e.g., a serial recall task) to assess effects on engagement and performance. Additionally, implementing gamification element interactions in a fully developed aviation weather training program could provide ground to assess long-term effects of gamification. In the realm of a gamified aviation weather training, it will be essential to include a robust set of training effectiveness measures including assessments of knowledge and skill gain as well as key attitudes (e.g., engagement in the aviation weather-specific task, motivation to complete such a training program, and self-efficacy) (Kraiger, Ford, & Salas, 1993).

Literature Review Summary

1. Due to the high popularity of gaming, researchers have begun to implement aspects of these games into real life (known as gamification). Gamification is commonly defined as "the use of game design elements in a non-game context" (Deterding et al., 2011).

2. Motivation and engagement may be influenced by certain game mechanics and user types, which then could also impact learning (Deterding et al., 2011; Martin et al., 2017; Jia et al., 2016; Nacke, Bateman, & Mandryk, 2014; Johnson et al., 2012).

3. Prior research indicates that implementing game elements seems to have a positive impact particularly in the educational domain (Hamari et al., 2014; Majuri, Koivisto, & Hamari, 2018; Rozman & Donath, 2019).

4. The most commonly implemented gamification elements in the education research included points, leaderboards, badges, challenges, levels, teams, progress, social networking, performance feedback, timer, narrative, and avatars (Majuri, Koivisto, & Hamari, 2018). Several studies reveal that narratives and leaderboards may have an effect on engagement and motivation (Armstrong & Landers, 2017; Majuri, Koivisto, & Hamari, 2018); however, a gap still exists as to whether leaderboards and narratives impact motivation or engagement when implemented in a non-game context (i.e., an online training program).

5. Based on the successful findings from the studies that utilized Narratives, Majuri and colleagues (2018) recommended that future researchers implement more immersion-based elements in the realm of education.

6. Using gamification elements in conjunction with each other limits what researchers can glean from each individual element. The review conducted by Majuri and colleagues (2018), recommends that future studies isolate the elements in an educational setting in order to control for overlapping elements. Further research must also isolate narratives as a variable and utilize a full story plot (rather than just a scenario) in order to move closer to a more concrete finding.

7. Training evaluation is an essential step in ensuring the training program yields effective outcomes (Bates, 2004). Training is the "systematic acquisition of knowledge, skills, and

attitudes [KSAs] that together lead to improved performance in a particular environment" (Salas et al., 2006). Effective training should let trainees learn the desired KSAs, have opportunities to practice these KSAs, and receive prompt feedback on their performance. A "pretest/posttest with comparison group" design combines aspects from the other two designed mentioned. It utilizes a pre- and post-test as well as a comparison group in order to assess training outcomes.

8. Studies indicate that aviation weather training for GA pilots is underdeveloped, particularly for NEXRAD (Cobbett et al., 2014; Blickensderfer et al., 2016). From a training perspective, research has demonstrated gaps in pilots' understanding of NEXRAD, but that pilots did benefit from short training courses (Cobbett et al., 2014; Blickensderfer et al., 2016). Further research is needed to bridge the gaps in this realm of training in order to improve pilots' understanding of the material and overall safety of flight.

9. Research indicates that pilots are a highly competitive subset of the population (oriented (Weiss, 2016).

10. The "pilot persona" categorizes most pilots as competitive, adrenaline-fueled, and goal-oriented (Weiss, 2016). Considering these descriptions of pilots in conjunction with Tondello et al. (2016), it is likely that GA pilots would fall in the "Achiever" category type.

11. In considering the competitive/goal-oriented nature of pilots, gamification elements such as Achievements and Stories could best motivate pilots-in-training through elevated engagement and motivation. Specifically, implementing a narrative could provide context for this type of training and give pilots a sense of purpose for completing the training

(Miranda, 2015). Leaderboards could drive pilots towards success through their sense of competition and achievement (Höllig et al., 2018).

Purpose

After reviewing the literature, gamification has the potential to increase pilots' level of motivation, engagement, and learning in the domain of aviation weather training. The purpose of the current study is to determine effectiveness of two types of gamification mechanics—narrative and leaderboard—on motivation, engagement, and knowledge acquisition in a Next Generation Radar (NEXRAD) online training program.

Hypotheses

IV1: Narrative

H1: There will be a significant effect of Narrative such that those in the Narrative Present (NP) conditions will produce higher scores on the multivariate dependent variable (DV) than the Narrative Absent (NA) conditions.

H_0: $\underline{x}_{NP} = \underline{x}_{NA}$

H_A: $\underline{x}_{NP} > \underline{x}_{NA}$

H1a: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *intrinsic motivation score* than the Narrative Absent (NA) conditions.

H1b: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *engagement score* than the Narrative Absent (NA) conditions.

H1c: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *knowledge acquisition score* than the Narrative Absent (NA) conditions.

IV2: Leaderboard

H2: There will be a significant effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce higher scores on the multivariate DV than the Leaderboard Absent (LA) conditions.

H_0: $\underline{x}_{NP} = \underline{x}_{NA}$

H_A: $\underline{x}_{NP} > \underline{x}_{NA}$

H2a: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *intrinsic motivation score* than the Leaderboard Absent (LA) conditions.

H2b: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *engagement score* than the Leaderboard Absent (LA) conditions.

H2c: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *knowledge acquisition score* than the Leaderboard Absent (LA) conditions.

Interactions

The interaction between Narrative and Leaderboard will be significant such that Conditions 1 and 5 will produce higher scores on the multivariate DV than the other conditions.

The interaction between Leaderboard and User Type will be significant such that Conditions 1 and 2 will produce higher scores on the multivariate DV than the other conditions.

Additional Analyses

H3a: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *intrinsic motivation score* than the Other (O) conditions.

H3b: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *engagement score* than the Other (O) conditions.

H3c: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *knowledge acquisition score* than the Other (O) conditions.

H4a: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *trainee reactions* than the Narrative Absent (NA) conditions.

H4b: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *trainee reactions* than the Leaderboard Absent (LA) conditions.

H4c: There will be a significant effect of User Type such that those in the Achiever User Type (A) conditions will produce greater *trainee reactions* than the Other User Type (O) conditions.

H5a: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *change in self-efficacy scores* than the Narrative Absent (NA) conditions.

H5b: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *change in self-efficacy scores* than the Leaderboard Absent (LA) conditions.

H5c: There will be a significant effect of User Type such that those in the Achiever User Type (A) conditions will produce greater *change in self-efficacy scores* than the Other User Type (O) conditions.

Chapter 3

Methods

Introduction

This chapter outlines the method used in this dissertation. First, this chapter provides the overarching research design of the study as well as a detailed description of the training intervention and gamification manipulation. After which, it discusses the measures used to assess differences in conditions. Also included is a discussion of the population of participants—including the eligibility, safety, and ethical considerations—as well as the quantity governed by the power analysis. This chapter also documents the procedure and stimuli used in the study.

Research Design

The purpose of this study was to assess the effects of gamification in an online NEXRAD training program. In order to test this intervention, the proposed study used a 2x2 between-subjects quasi-experimental design. The independent variables are narrative (presence vs. absence) and leaderboard (presence vs. absence). As shown in Figure 11, the three IVs combine to generate four study conditions: (1) narrative present/ leaderboard present, (2) narrative present/ leaderboard absent, (3) narrative absent/ leaderboard present, and (4) narrative absent/ leaderboard absent.

Figure 11

Study Conditions

		Leaderboard	
		Present	Absent
Narrative	Present	*Condition 1*	*Condition 3*
	Absent	*Condition 2*	*Condition 4*

The dependent variables assessed in this study were motivation, engagement, trainee reactions, self-efficacy, and NEXRAD knowledge.

Participants

This study was approved by Embry-Riddle Aeronautical University's Institutional Review Board to ensure participant protection and safety. Participants were all eighteen years of age or older and were at least in training to obtain a pilot's license. The participants were recruited through Embry-Riddle Aeronautical University's SONA system (i.e., a software platform that allows researchers to provide students extra credit/monetary compensation in exchange for study participation) and classes within the aviation and meteorology departments and elsewhere around campus. Participants signed up for the study of their own free will. Once they agreed to participate, they were given a detailed Informed Consent form to read and sign electronically. This form outlined all risks, benefits, and expectations of their participation. No one, other than the researchers, had access to any of the responses. If the participant wished to drop out during the study, their information was not used in any capacity.

Forty-one (6 female) Embry-Riddle Aeronautical University students participated in the study. The average age of the participants was 21 (SD = 3.49). Participants had a median of 130 flight hours (M = 157.61, SD = 103.97) and a median of 2 years flying (M = 2.41, SD = 1.40). The pilot status breakdown is as follows: 4 student pilots, 21 Private pilots, and 16 Commercial pilots. Twenty-four of these participants were Instrument-rated, and 4 were Certificated Flight Instructors.

Experimental Manipulation/Independent Variable

The experimental manipulation is gamification type: Narrative or Leaderboard. The gamification techniques were applied to the NEXRAD Training course.

NEXRAD Training course

While the realm of aviation weather training overall needs reform, the focus of this study was on assessing the efficacy of a radar training program. As described in the literature review, prior studies indicate that aviation weather training for GA pilots is underdeveloped, particularly for NEXRAD (Cobbett et al., 2014; Blickensderfer et al., 2016). For the current study an existing, online course on NEXRAD was selected (Thomas et al., 2015). A fully online instructional method was chosen to reach the greatest number of trainees during the pandemic. The course focuses on knowledge of Radar Basics, Reflectivity, Radar Beams, Radar Modes, and Radar Interface. The specific learning objectives are as follows (see Table 5):

Table 5

Learning Objectives for the Training Course

Content Topic	Learning Objective
Radar Basics	LO 1: Describe the basic principles of radar
Reflectivity	LO 2: State the basics of reflectivity
	LO 3: Distinguish between reflectivity intensity terms
Radar Beams	LO 4: Describe how radar beams work
	LO 5: Compare the three types of refraction
	LO 6: Recognize how to overcome refraction
Radar Modes	LO 7: Compare VCP Precipitation and Clean Air Mode
Radar Interface	LO 8: Compare base and composite reflectivity products
	LO 9: Describe how a mosaic is made
	LO 10: Recognize radar limitations due to data latency

To prepare the course for the current study, several steps were taken. First, the number of topics for radar training were condensed by a subject matter expert in order to accommodate potential time constraints of the study. Additionally, the training program was converted to Qualtrics to efficiently distribute the course virtually. The training program slides, designed and narrated in PowerPoint, were converted to short videos embedded in the Qualtrics survey. These videos provide both auditory and visual information in order to better engage users. The course parameters provide an overview of elements such as the purpose, lesson length, target audience, prerequisites, course format, room arrangement, and necessary equipment. The detailed lesson plan expands on the course overview in order to create a guide for trainers to aid in information delivery.

Gamification Type 1: Narrative

The narrative is intended to present a meaningful story to the participants, provide rationale for the assignment, and make a “less exciting” topic more enjoyable (Pandey, 2018). This brief narrative used in the study was created based on a “Soap Hero’s Journey”: The Call → The Challenge → The Transformation → The Twist → The Resolution (Marczewski, 2019). Prior to the training, the narrative told to trainees is as follows:

The Call: Imagine… You are just starting your ground school and flight training. You are eager to begin flying, but your instructor explains that it is essential to first understand the fundamentals of aviation weather.

The Challenge: In order to prepare for a quiz next class, he asks you to review an online course on radar basics. It is important that you pay attention to the information on these slides in order to impress your instructor and reinforce your understanding of radar.

The Transformation: You decided to take this quiz seriously in order to improve your skills.

The Twist: As an added bonus, your flight instructor proposes a challenge: the top students in the class will win a free hour of flight training! It is important that you pay attention to the information on these slides in order to impress your instructor and reinforce your understanding of radar.

After the training but before the post-test, trainees see the last portion of the narrative:

The Resolution: You have now completed the training, and it's time for the quiz. Don't forget that the top students in the class will win a free hour of flight training! Please try your best to complete the following quiz.

Gamification Type 2: Leaderboard

The leaderboard was intended to evoke a sense of competition in participants who identify as an Achiever user type (Marczewski, 2018). Immediately following the pretest, a leaderboard presents "Today's High Scores" to participants with their pretest scores coded in. These scores represented the participant's real score and fictitious scores for the other four spots (i.e., the leaderboard showed the same fake scores to every participant with their actual score). As an example, participants who scored 95% would see their score in the 3rd place spot and the fictitious scores would appear above and below their score. They were then instructed to pay attention to the training material and reminded of the leaderboard at end of the training in order to motivate them to pay attention to the material. After the posttest, they were again shown the leaderboard with their posttest placement on the board.

Figure 12

Leaderboard with Participant's Real Score and Fictitious Other Scores

Leaderboard

1. YOUR SCORE	100
2. P291	97
3. P177	91
4. P322	81
5. P119	67

After a pilot study using the aforementioned parameters, the gamification elements' salience and frequency were increased. To make the narrative more captivating, this element was converted into a video with a man discussing how the FAA needed accurate results to increase safety around pilots' use of NEXRAD. Subject matter experts agreed that this would provide a more compelling reason for pilots to engage in the training. Additionally, the fictitious leaderboard scores were lowered (from 100, 97, 91, 81, and 67 to 91, 84, 63, and 47) to reflect a more realistic range (as demonstrated by the pilot participant data). The training was also split into two sections with a short break in the middle to reintroduce the gamification element(s) depending on the condition. Provisions were also implemented to better encourage participants to watch the training video.

Measures

All measures were distributed to participants via Qualtrics survey software.

Demographics

The demographic questionnaire (Appendix A) included items to identify participant's age (interval), sex (nominal), education (ordinal), and pilot certificate (ordinal). This data was used for exploratory analyses and future research questions.

Weather Self-Efficacy

In order to assess participants' confidence in their NEXRAD capabilities, a ten item self-efficacy scale was implemented ($\propto = 0.868$) (Appendix B). The items are measured on a 1 (Strongly Disagree) to 7 (Strongly Agree) Likert scale. This scale was originally developed and administered in a previously validated NEXRAD training course (Cobbett, Blickensderfer, & Lanicci, 2014). An overall self-efficacy score was computed for each participant by averaging the 10 items together.

Trainee Reactions

Six reaction items, created by Long and colleagues (2008), measured trainee reactions to the training ($\propto = 0.935$) (Appendix D). This survey assessed reactions on three dimensions: technology satisfaction, enjoyment, and relevance of course content (Long et al., 2008) (note that the Long et al. items were originally adapted from Kettanurak et al. (2001) and Morgan & Casper (2000)). A five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) measured these items. Each participant had an overall score (an average of their responses to each of the six items) as well as scores on each of the three sub-dimensions. This data was used for exploratory analyses as well as training validation.

Engagement

The User Engagement Scale (UEC) (Appendix C) was developed and validated by O'Brien and colleagues ($\propto$ = 0.836) (2018). The UEC measures engagement as self-reported by the user. There are a total of 30 items on the UEC, each scored on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The UEC measures overall engagement and four sub-dimensions: perceived usability (eight items), aesthetic appeal (five items), focused attention (seven items) and reward factor (ten items). An overall engagement score was computed for each participant by averaging the 30 items together. Each participant also had an average score for each of the four sub-dimensions. The UEC was administered after the experimental trials.

Table 6

Cronbach's $\propto$ Values for Each Subscale of the UEC

Scale Dimension	$\propto$
Perceived Usability	0.71
Aesthetic Appeal	0.88
Focused Attention	0.89
Reward Factor	0.85
Total	**0.84**

Motivation

Motivation was measured in this study using a 37-item variation of the Intrinsic Motivation Inventory (IMI) ($\propto$ = 0.935) (Ryan, 1982). The IMI measures motivation across six dimensions: interest/enjoyment (seven items), perceived competence (six items), effort/importance (five items), pressure/tension (five items), perceived choice (seven items), and value/ usefulness (seven items). In this questionnaire, participants reported perceived motivation on a 1 (Not true at all) to 7 (Very true) Likert scale. Appendix E includes the 37 items both in numerical order and broken down by dimension. Considering reverse coded items, each participant had an overall Intrinsic

Motivation Score (computed for each participant by averaging the 37 items together) as well as an average score for each of the six dimensions.

Table 7

Cronbach's ∝ Values for Each Subscale of the IMI

Scale Dimension	**∝**
Interest/Enjoyment	0.90
Perceived Competence	0.77
Effort/Importance	0.84
Pressure/Tension	0.77
Perceived Choice	0.72
Value/Usefulness	0.85
Total	**0.94**

NEXRAD Knowledge Assessment

A multi-disciplinary team of meteorologists, pilots, and human factors psychologists developed two parallel forms of a weather knowledge test (Appendix F). Some questions stemmed from a previous aviation weather product interpretation test (Blickensderfer et al., 2017). A human factors specialist and a meteorologist reviewed these questions in relation to the training program and revised/added additional questions to ensure the questions fit the learning objectives of the revised NEXRAD course. Each of the parallel forms contain 20 multiple-choice questions with two questions assessing each learning objective. The reason for the parallel forms is to allow for unbiased assessment of NEXRAD knowledge pre-intervention ($\propto = 0.560$) and post-intervention ($\propto = 0.703$). For each participant, a score out of 100% (score/20*100) was calculated for each test. Once both a pre- and post-test score were obtained, change scores were calculated based on the difference in score (posttest percentage minus pretest percentage).

Weather Training Questionnaire

The Weather Training Questionnaire consisted of eight 5-point Likert scale items to determine how much prior experience each participant had ($\propto = 0.764$). Each question asked participants, "For the following, please rate how much training you have had for…" and proceeded to name 8 topics related to Radar and aviation weather training: basic weather theory, single site Radar, National Mosaic Radar, Radar Coded Message (RCM), 1-800-wxbrief (call-in), 1800wxbrief.com (Leidos), Foreflight, and aviationweather.gov (Aviation Weather Center website).

User Type

The HEXAD Gamification User Types questionnaire consists of 27 7-point Likert scale items (Tondello et al., 2016) (Appendix G). Users receive a score for each of the six user types: Philanthropist, Socializer, Free Spirit, Achiever, Disruptor, or Player. Each user type score was calculated by summing all responses for that category. User type for each participant is determined based on which score(s) are the highest. Once user types were calculated, participants were either categorized as "Achiever" or "Other" (i.e., any other user type). Table 8 shows the internal consistency / reliability for each user type.

Table 8

Cronbach's ∝ Values for Each Subscale of the HEXAD User Type Questionnaire

Scale Dimension	∝
Philanthropist	0.86
Socializer	0.90
Free Spirit	0.37
Achiever	0.64
Disruptor	0.78
Player	0.86

Procedure

Once participants received the email invitation to participate and clicked on the link, they were directed to the Qualtrics survey. All subsequent actions from the participant were completed in Qualtrics. Qualtrics randomly assigned participants to one of four conditions: (1) narrative present/ leaderboard present, (2) narrative present/ leaderboard absent, (3) narrative absent/ leaderboard present, and (4) narrative absent/ leaderboard absent.

First, participants gave their consent to participate in the training and then proceeded to fill out the demographic questionnaire. Next, participants filled out the weather training questionnaire and self-efficacy questionnaire. Participants then completed the NEXRAD knowledge pre-test. After completion of the pre-test, participants completed the online training program tailored to their experimental condition.

In condition 1, participants received the narrative, saw the leaderboard, went through the training, finished the narrative, then saw the leaderboard once more. In condition 2, participants received the narrative, went through the training, and then finished the narrative. In condition 3, participants saw the leaderboard, went through the training, and saw the leaderboard once more. Finally, participants in condition 4 only went through the training.

Immediately following the training, participants completed the NEXRAD knowledge posttest. Participants then filled out the User Engagement Survey and Intrinsic Motivation Inventory. Lastly, participants were debriefed on the study, thanked for their involvement, and compensated $15.

Power Analysis

A statistical power analysis was conducted via a priori sample size determination in order to make conclusions with higher confidence. G*Power version 3.1.9.6 was used to compute this sample size using the "MANOVA: Global effects" setting. The following variables were used in the analysis:

Effect Size (η^2) = 0.25

Power (1- β) = 0.95

Type I Error Probability (α) = 0.05

Number of groups = 4

Response variables (DVs) = 3

Based on the input variables, a sample size of 36 was recommended. This represents the minimum number of participants needed to detect an effect. After discussion with the dissertation committee, a sample size of 40 (10 participants per condition) was determined as adequate.

Chapter 4

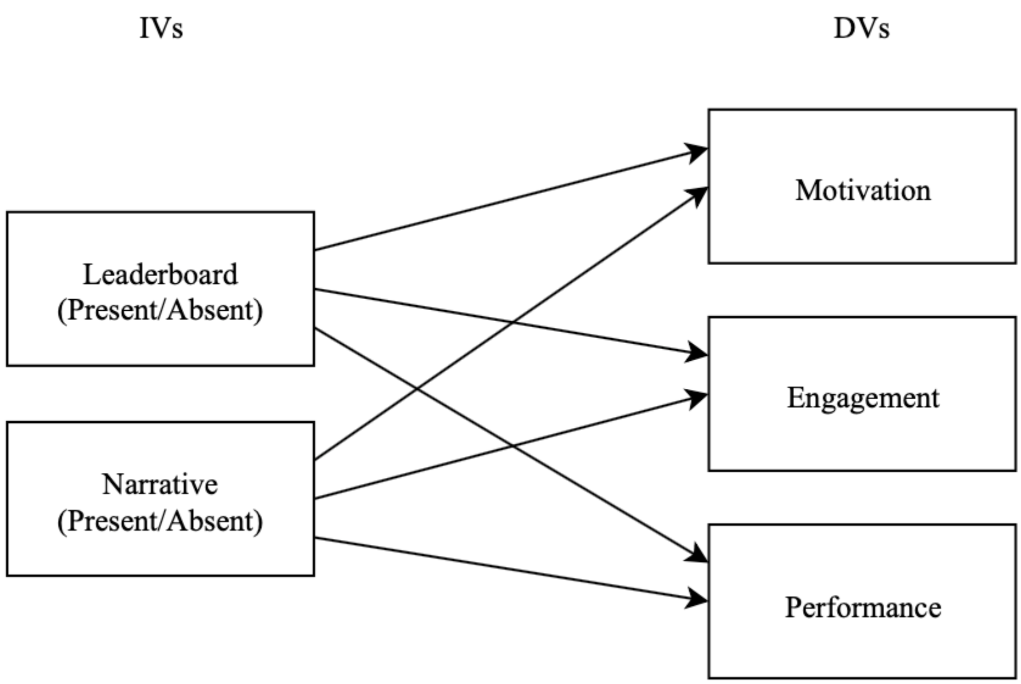
Results

Proposed Analyses

Data was exported from Qualtrics and organized in Microsoft Excel. All analyses were conducted using IBM's Statistical Package for Social Sciences (SPSS) software version 22. A between-subjects multivariate analysis of variance (MANOVA) was proposed to assess the effects of Narrative and Leaderboard on Engagement, Motivation, and Knowledge acquisition. If the analysis was significant, main effects and Bonferroni post hoc analyses of Engagement, Motivation, and Knowledge acquisition were to be assessed.

Figure 13

Statistical Model for Two-Way MANOVA



Descriptive Statistics

Main Dependent Variables

The descriptive statistics for the study variables are shown in Table 9 and 10. An intercorrelation matrix of the variables is shown in Table 11.

Table 9

Descriptive Statistics for the DVs per Condition

Measure	Narrative + Leaderboard		Narrative		Leaderboard		No Gamification		Total	
	n = 12		*n* = 9		*n* = 11		*n* = 9		***n* = 41**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	***M***	***SD***
Knowledge acquisition	14.58	16.02	12.78	22.93	14.09	19.21	14.44	13.56	**14.02**	**17.47**
Engagement	3.70	0.37	3.71	0.28	3.72	0.45	3.45	0.40	**3.65**	**0.38**
Motivation	5.52	0.52	5.23	0.47	5.67	0.47	5.48	0.77	**5.49**	**0.56**
Trainee Reactions	4.51	0.45	4.39	0.60	4.52	0.44	4.35	0.58	**4.48**	**0.48**

Table 10

Descriptive Statistics for the DVs per Main Effect

Measure	Narrative Present		Narrative Absent		Leaderboard Present		Leaderboard Absent		Total	
	n = 21		*n* = 20		*n* = 23		*n* = 18		***n* = 41**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	***M***	***SD***
Knowledge acquisition	13.81	18.77	14.25	16.49	14.35	17.21	13.61	18.30	**14.02**	**17.47**
Engagement	3.71	0.32	3.60	0.44	3.71	0.40	3.58	0.36	**3.65**	**0.38**
Motivation	5.39	0.51	5.59	0.61	5.59	0.49	5.36	0.63	**5.49**	**0.56**
Trainee Reactions	4.58	0.43	4.37	0.52	4.49	0.47	4.46	0.50	**4.48**	**0.48**

Table 11

Intercorrelation Matrix for Main DVs

Measure	1	2	3	4
1. Knowledge acquisition	--			
2. Engagement	-0.28	--		
3. Motivation	0.17	-0.17	--	
4. Trainee Reactions	-0.01	.40**	-0.15	--

**. Correlation is significant at the 0.01 level (2-tailed).

User Type

User Type was calculated such that participants were categorized as an Achiever, Player, Socializer, Philanthropist, Disruptor, or Free Spirit. Table 12 shows breakdown of participants by User Type and condition. If a user identified as more than one user type (i.e., received the same highest score for multiple types), then all identified user types were counted. Therefore, the frequency total for each condition will be higher than the actual number of participants in the study. Overall, 10 participants identified as an Achiever user type for one of their main user types.

Table 12

Table of User Types by Condition and Frequency

User Type	**Narrative + Leaderboard**	**Narrative**	**Leaderboard**	**No Gamification**	**Total**
Philanthropist	8	5	5	6	**24**
Socializer	4	3	1	5	**13**
Free Spirit	1	2	2	2	**7**
Achiever	5	2	1	2	**10**
Disruptor	0	0	0	0	**0**
Player	2	3	5	2	**12**

Assumptions Testing

The MANOVA assumptions must be checked before conducting hypothesis testing.

Assumption 1. The three dependent variables (Knowledge acquisition, Engagement, and Motivation) measured are all analyzed as continuous variables.

Assumption 2. The two independent variables (Narrative Presence and Leaderboard Presence) are categorical, independent groups.

Assumption 3. Different participants were used for each separate condition (1, 2, 3, or 4).

Assumption 4. Sample size was adequate based on the power analysis performed prior to the study.

Assumption 5. Boxplots and Mahalanobis Distances found no significant univariate or multivariate outliers in the data set. Therefore, no outliers were removed for analysis.

Assumption 6. A Shapiro-Wilk's test for normality indicated all dependent variables (Knowledge acquisition, Engagement, and Motivation) were normally distributed for both independent variables (Narrative Presence and Leaderboard Presence) ($p > 0.05$).

Assumption 7. There is not a linear relationship between any pairs of dependent variables except Engagement and Trainee Reactions.

Assumption 8. A Box's test of equality of covariance matrices was conducted to assess homogeneity of variance-covariances matrices. The assumption of homogeneity of variance-covariances was not violated ($p = 0.13$).

Assumption 9. To meet assumptions for a MANOVA, dependent variables should be moderately correlated with each other. There is no multicollinearity.

Since the assumption of multicollinearity was not met, hypothesis testing will be run as separate ANOVAs rather than one MANOVA for the model.

Hypothesis Testing

Narrative

Hypothesis 1a proposed that there will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *intrinsic motivation score* than the Narrative Absent (NA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Narrative was not significant $F(1,37) = 1.33$, $p = 0.26$, $\eta p^2 = 0.04$.

Hypothesis 1b proposed that there will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *engagement score* than the Narrative Absent (NA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Narrative was not significant $F(1,37) = 0.99$, $p = 0.33$, $\eta p^2 = 0.03$.

Hypothesis 1c proposed that there will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce a higher *knowledge acquisition score* than the Narrative Absent (NA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Narrative was not significant $F(1,37) = 0.01$, $p = 0.92$, $\eta p^2 = 0.00$.

Leaderboard

Hypothesis 2a proposed that there will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *intrinsic motivation score* than the Leaderboard Absent (LA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Leaderboard was not significant $F(1,37) = 1.82$, $p = 0.19$, $\eta p^2 = 0.05$.

Hypothesis 2b proposed that there will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *engagement score* than the Leaderboard Absent (LA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Leaderboard was not significant $F(1,37) = 1.18, p = 0.28, \eta p^2 = 0.03$.

Hypothesis 2c proposed that there will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce a higher *knowledge acquisition score* than the Leaderboard Absent (LA) conditions. A two-way ANOVA was conducted to assess differences in Narrative and Leaderboard presence. The main effect for Leaderboard was not significant $F(1,37) = 0.02, p = 0.90, \eta p^2 = 0.00$.

Interactions

A two-way, between subjects ANOVA was conducted to assess the effects of Leaderboard presence/absence and Narrative presence/absence on engagement. The interaction was not significant $F(1,37) = 1.26, p = 0.27, \eta p^2 = 0.03$.

A two-way, between subjects ANOVA was also conducted to assess the interaction effects of Leaderboard presence/absence and Narrative presence/absence on motivation. The interaction was not significant $F(1,37) = 0.08, p = 0.78, \eta p^2 = 0.002$.

Lastly, a two-way, between subjects ANOVA was conducted to assess the interaction effects of Leaderboard presence/absence and Narrative presence/absence on knowledge acquisition. The interaction was not significant $F(1,37) = 0.04, p = 0.85, \eta p^2 = 0.001$.

User Type

Hypothesis 3a proposed that there will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *intrinsic motivation score* than the Other (O) conditions. The effect was not significant $F(1,39) = 1.13$, $p = 0.30$, $\eta p^2 = 0.03$.

Hypothesis 3b proposed that there will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *engagement score* than the Other (O) conditions. The effect was not significant $F(1,39) = 0.03$, $p = 0.86$, $\eta p^2 = 0.001$.

Hypothesis 3c proposed that there will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher *knowledge acquisition score* than the Other (O) conditions. The effect was not significant $F(1,39) = 0.53$, $p = 0.47$, $\eta p^2 = 0.01$.

Trainee Reactions

Hypothesis 4a proposed that there will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *trainee reactions* than the Narrative Absent (NA) conditions. Hypothesis 4b proposed that there will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *trainee reactions* than the Leaderboard Absent (LA) conditions. A two-way analysis of variance (ANOVA) was conducted to assess differences in trainee reactions between Narrative (present/absent) and Leaderboard (present/absent). The interaction results were not significant $F(1,37) = 0.01$, $p = 0.91$, $\eta p^2 = 0.001$. The main effect for Narrative Presence was not significant $F(1,37) = 1.82$, $p = 0.19$, $\eta p^2 = 0.05$. The main effect for Leaderboard Presence was not significant.

Hypothesis 4c proposed that there will be a significant effect of User Type such that those in the Achiever User Type (A) conditions will produce greater *trainee reactions* than the Other User Type (O) conditions. A one-way analysis of variance (ANOVA) was conducted to assess differences in trainee reactions between user type (Achiever/Other). The results were not significant $F(1,39) = 0.05, p = 0.83, \eta p^2 = 0.001$.

Table 13

Descriptive Statistics by User Type (Achiever or Other)

Measure	Other		Achiever		Total	
	M	*SD*	*M*	*SD*	***M***	***SD***
Knowledge acquisition	15.16	17.68	10.50	17.23	**14.02**	**17.47**
Engagement	3.66	0.39	3.63	0.37	**3.65**	**0.38**
Motivation	5.43	0.60	5.65	0.42	**5.49**	**0.56**
Trainee Reactions	4.49	0.49	4.45	0.46	**4.48**	**0.48**

Exploratory Analyses

Change in Knowledge acquisition

Though it was not part of the original hypotheses, the researchers examined change in knowledge acquisition to determine if learning occurred across conditions. A paired samples t-test was conducted to determine whether there was a statistically significant mean difference between the pretest and posttest knowledge acquisition scores, regardless of condition. Participants scored higher on the posttest (M = 62.68, SD = 18.74) than the pretest (M = 48.66, SD = 15.93). This difference in score was a statistically significant increase $t(40) = 5.14$, $p < .001$.

Table 14

Descriptive Statistics for Pretest and Posttest Knowledge acquisition Scores

Test	Mean	N	SD	SE
Posttest	62.68	41	18.74	2.93
Pretest	48.66	41	15.93	2.49

Player User Type

After reviewing the results for User Type, we determined that the gamification did not have a significant effect on Achiever user types. Further analyses were proposed in order to determine if recategorizing the user types would have a greater effect. Instead of dividing participants into Achiever user type and Other (all other user types grouped together), the groups were redistributed into Player user type and Other. Player user types rely on rewards within the game for motivation, so they may be more susceptible to the gamification elements. Three one-way ANOVAs were conducted to assess the effects of User Type on Engagement, Motivation, and Knowledge acquisition. The effect of User type on Engagement was not significant $F(1,39) = 2.25, p = 0.14$, $\eta p^2 = 0.05$. The effect of User type on Motivation was not significant $F(1,39) = 3.82, p = 0.06$, $\eta p^2 = 0.09$. The effect of User type on Knowledge acquisition was not significant $F(1,39) = 0.05$, $p = 0.82$, $\eta p^2 = 0.001$.

Perceived Competence

Since posttest self-efficacy was not collected, researchers examined the Intrinsic Motivation Inventory subscale “perceived competence” instead to determine group differences. A two-way analysis of variance (ANOVA) was conducted to assess differences in perceived competence between Narrative (present/absent) and Leaderboard (present/absent). The interaction results were not significant $F(1,37) = 0.001$, $p = 0.97$, $\eta p^2 = 0.001$. The main effect for Narrative Presence was not significant $F(1,37) = 0.02$, $p = 0.88$, $\eta p^2 = 0.001$. The main effect for Leaderboard Presence was not significant $F(1,37) = 0.001$, $p = 0.98$, $\eta p^2 = 0.001$.

Table 15

Descriptive Statistics for Perceived Competence by Condition

Condition	M	SD
Narrative + Leaderboard	5.06	1.05
Narrative	5.04	0.76
Leaderboard	5.09	1.16
No Gamification	5.09	0.68
Total	**5.07**	**0.92**

Chapter 5

Discussion

Summary of Results

Before delving further into the findings, table 16 contains an abridgement of the results presented in the prior section of this dissertation.

Table 16

Summary of Results

Hypothesis	Analysis Performed	Result	Outcome
H1: There will be a significant effect of Narrative such that those in the Narrative Present (NP) conditions will produce higher scores on the multivariate DV than the Narrative Absent (NA) conditions.	Between groups, two-way ANOVA	No significant difference detected	Hypothesis not supported
H2: There will be a significant effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce higher scores on the multivariate DV than the Leaderboard Absent (LA) conditions.	Between groups, two-way ANOVA	No significant difference detected	Hypothesis not supported
H3a: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher intrinsic motivation score than the Other (O) conditions.	Between groups, two-way ANOVA	No significant difference detected	Hypothesis not supported
H3b: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher engagement score than the Other (O) conditions.	Between groups, two-way ANOVA	No significant difference detected	Hypothesis not supported
H3c: There will be a significant main effect of User Type such that those in the Achiever (A) conditions will produce a higher knowledge acquisition score than the Other (O) conditions.	Between groups, two-way ANOVA	No significant difference detected	Hypothesis not supported
H4a: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *trainee reactions* than the Narrative Absent (NA) conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported

H4b: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *trainee reactions* than the Leaderboard Absent (LA) conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported
H4c: There will be a significant effect of User Type such that those in the Achiever User Type (A) conditions will produce greater *trainee reactions* than the Other User Type (O) conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported
H5a: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *self-efficacy* than the Narrative Absent (NA) conditions.	Data Unavailable	Data Unavailable	Hypothesis could not be tested
H5b: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *self-efficacy* than the Leaderboard Absent (LA) conditions.	Data Unavailable	Data Unavailable	Hypothesis could not be tested
H5c: There will be a significant effect of User Type such that those in the Achiever User Type (A) conditions will produce greater *self-efficacy* than the Other User Type (O) conditions.	Data Unavailable	Data Unavailable	Hypothesis could not be tested
Exploratory Analysis: There will be a significant difference between pretest and posttest knowledge acquisition scores across all conditions.	Paired samples t-test	Significant difference detected	Hypothesis supported
Exploratory Analysis: There will be a significant effect of User Type such that those in the Player User Type conditions will produce higher scores on the multivariate DV than the Other User Type conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported
Exploratory Analysis: There will be a significant main effect of Narrative such that those in the Narrative Present (NP) conditions will produce greater *perceived competence* than the Narrative Absent (NA) conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported
Exploratory Analysis: There will be a significant main effect of Leaderboard such that those in the Leaderboard Present (LP) conditions will produce greater *perceived competence* than the Leaderboard Absent (LA) conditions.	Between groups, one-way ANOVA	No significant difference detected	Hypothesis not supported

Hypothesized Results

Narrative

It was hypothesized that there would be a significant effect of Narrative such that those in the Narrative Present conditions will produce higher scores on the multivariate DV (engagement, motivation, knowledge gain) than the Narrative Absent conditions. The findings from this study did not support this hypothesis; Narrative had no effect on engagement, motivation, or knowledge acquisition. With a lack of effect on any of these variables, the strength of the independent variable is suspect. That is, one reason that no effect was detected could be due to the lack of real consequence in the narrative's message since participants volunteered to take part in the training. The researcher attempted to strengthen the narrative after the pilot study, but it still did not have any increased impact on knowledge acquisition, engagement, or motivation in comparison to conditions without a narrative.

Another potential explanation from the lack of findings is the mixed evidence on the effectiveness of narratives in training. As mentioned previously in this dissertation, a study conducted by Armstrong and Landers (2017) found significantly higher perceptions of learning (i.e., trainee reactions) but lower knowledge scores for the narrative condition (compared to a control group). This demonstrates how narratives may not always have an impact on actual learning or knowledge acquisition. Similarly, Miranda's (2015) study did not find an overall impact of narrative on engagement in a vigilance task. The efficacy of Narratives remains unknown, particularly in an online training capacity. It is likely that gamification elements like Narratives must have increased salience in an online setting to evoke changed outcomes. A future study might utilize Narratives in a required setting (such as a real classroom) in order to encourage a sense of urgency to accompany the compelling story supplied by the training program.

Leaderboard

It was also hypothesized that there would be a significant effect of Leaderboard such that those in the Leaderboard Present conditions will produce higher scores on the multivariate DV than the Leaderboard Absent condition. Leaderboard had no effect on engagement, motivation, or knowledge acquisition, and thus, the findings from this study did not support this hypothesis. One explanation for the lack of impact is the user types of the research participants. That is, Leaderboards are thought to be most effective with "Achiever" or "Player" user types. There was a prior assumption that pilots would most likely be "Achiever" or "Player" user types in parallel with their competitive dispositions. However, based on the responses on the user type questionnaire, the majority of participants (n = 24) identified as *Philanthropist* user types for at least one of their top user types (i.e., user type scores tied for first). Philanthropists are Intrinsic user types, meaning they are driven by an inward desire to perform the activity as dictated by the RAMP framework (Tondello et al., 2016). Specifically, Philanthropists seek a sense of purpose through helping others and expecting nothing in return. Since participants opted into the study of their own volition and compensation was only $15, it is possible that only "helpful" people signed up to participant in this research. Thus, it is possible the Leaderboard presence did not have a strong enough impact to influence Philanthropists who were the majority of participants; the Philanthropists may simply have not cared about their position on the leaderboard.

Additionally, it was presumed that pilots would have high trait competitiveness, or "the enjoyment of interpersonal competition and the desire to win and be better than others" (Spence & Helmreich, 1983; Höllig et al., 2018). Only ten participants identified as Achievers user types, who typically strive to be the best (Tondello et al., 2016). This trait competitiveness may not have been strong enough to counteract possible pervasive pilots' views that weather is a less exciting part of aviation training. In other words, there may be other flight tasks where the competitive spirit of pilots would have an effect. Another issue is the placement of individuals on the leaderboard. One survey conducted by Jia and colleagues (2017) found that negative perceptions of leaderboards are particularly present in those who are ranked lower on the leaderboard. While the researcher attempted to create a reasonable spread of fictitious scores on the leaderboard (based on real participant scores in the pilot study) to keep from discouraging participants, future iterations could lower the fictitious leaderboard scores even more to create more optimism going into the training.

User Type

It was hypothesized that there would be a significant effect of User Type such that those in the Achiever User Type conditions will produce higher scores on the multivariate DV than the Other User Type conditions. Intrinsic user types—like Achievers and Philanthropists—are driven by an inward desire to perform the activity (Tondello et al., 2016). Achievers have a need for mastery within a game and constantly try to improve themselves without needing others' validation. The findings from this study did not support this hypothesis. An exploratory analysis looked to see if there was a significant effect of User Type when operationally defined as Player (rather than Achiever) versus all other categories. This analysis was also not significant. In terms of possible reasons why an effect was not found, as previously stated, most participants identified as Philanthropist user types for at least one of their top user types (i.e., user type scores tied for first). The small sample of Achiever user types likely made it more difficult to find an effect. In terms of the Player user types, these users are not motivated intrinsically and rely on rewards within the game. The "rewards" of gamification were not salient enough to find any effect. Finally, the gamification elements were not chosen for Philanthropist user types; therefore, the Leaderboard may not be suitable for this sample. Philanthropists typically prefer gamification elements that allow them to interact in helpful ways with their fellow users (Tondello et al., 2016). For example, Philanthropists may enjoy having a rationale behind a given task that allows them to feel like they are part of some greater purpose. Alternatively, Philanthropist could use gifting/sharing items or knowledge to show their altruism within the system or training. If a population seems to identify as Philanthropists, these gamification elements could benefit future training. For future studies, the user type questionnaire can be utilized as a pre-screener in order to obtain a sample of Achiever user types large enough to find an effect.

Trainee Reactions

It was hypothesized that there would be a significant main effect of Narrative such that those in the Narrative Present conditions will produce greater trainee reactions than the Narrative Absent conditions. It was also hypothesized that there would be a significant main effect of Leaderboard such that those in the Leaderboard Present conditions will produce greater trainee reactions than the Leaderboard Absent conditions. The findings from this study did not support this hypothesis. Trainee reactions across groups were high ($M = 4.48$, $SD = 0.48$), but not different enough to find any between groups differences. A larger scale item for trainee reactions could rule out potential ceiling effects; however, their trainee reactions were significantly positively correlated with engagement scores ($r(39) = 0.40$, $p = 0.01$). While this finding is positive for the training program content, it does not provide any additional insight into how gamification elements impact trainee reactions.

Self-Efficacy and Perceived Competence

It was also hypothesized that there would be significant main effects of Narrative and Leaderboard on self-efficacy. It was predicted that those in the Narrative Present conditions will produce greater change in self-efficacy scores than the Narrative Absent conditions. Additionally, it was those in the Leaderboard Present conditions will produce greater change in self-efficacy scores than the Leaderboard Absent conditions. The posttest self-efficacy data was not collected, so these hypotheses could not be tested. Instead, researchers examined the Intrinsic Motivation Inventory subscale "perceived competence" to assess affect after the training. The two measures are compared in table 17. A two-way ANOVA was conducted to assess differences in perceived competence between Narrative (present/absent) and Leaderboard (present/absent). A significant difference was not detected between the four conditions. While pretest self-efficacy and posttest perceived competence are not directly comparable (though both are measured on a 7-point Likert scale), (posttest) perceived competence scores were notably higher than (pretest) self-efficacy scores. This *could* indicate an overall increase in participants' confidence in their abilities due to the training, but not the gamification elements.

Table 17

Comparison of Self-Efficacy and Perceived Competence Measures

Self-Efficacy Measures	Perceived Competence Measures
I am confident in my ability to use NEXRAD data and products.	I am satisfied with my performance at this task.
There are some tasks in using NEXRAD that I cannot do well.	I think I am pretty good at this activity.
When my performance is poor, it is due to my lack of ability to work with NEXRAD.	This was an activity that I couldn't do very well.
I doubt my ability to use NEXRAD.	I was pretty skilled at this activity.
I have all the skills needed to use NEXRAD very effectively.	After working at this activity for a while, I felt pretty competent.
Most people in my industry can use NEXRAD better than I can.	I think I did pretty well at this activity, compared to other students.
I am an expert at using NEXRAD.	

My future in this industry is limited because of my lack of skills with NEXRAD.
I am very proud of my skills and abilities in using the NEXRAD.
I feel nervous when others watch me use NEXRAD.

Training Implications

To understand the impact of this type of training, we must revisit the Kirkpatrick Model of Training Evaluation (Kirkpatrick, 1976). As previously noted, this framework consists of four levels of analysis: reaction, learning, behavior, and results. For the purpose of this dissertation, only the first two levels were able to be measured. Additionally, the Kirkpatrick Model requires positive outcomes of the lower levels before progressing to measure higher levels.

Reactions

Reactions include trainee responses to the training based on how engaging or relevant they find the information (Kirkpatrick, 1976). While scores on the trainee reactions questionnaire did not differ across groups, reactions collapsed across conditions were overwhelmingly positive. Similarly, engagement did not significantly differ between groups, but overall engagement was above average (M = 3.65, SD = 0.38). We also saw a significant, positive correlation between trainee reactions and engagement, $r(39) = 0.40$, $p = 0.01$. These scores reveal that participants reacted well to the training and emerged with a perception of valuable knowledge gained; however, this outcome was not influenced by the gamification elements incorporated into the training. Since positive findings occurred for Reactions, the next level of the Kirkpatrick Model—Learning—can be examined.

Learning

Learning measures how well the trainees acquired the intended KSAs from the training program (Kirkpatrick, 1976). This study measured learning through the knowledge acquisition scores on the pretest and posttest. Again, there were no significant differences in change in knowledge acquisition between conditions; however, a significant change between pretest and posttest was present overall. While this could indicate a testing effect (and in turn a threat to internal validity), it is likely that the parallel forms are distinct enough to overcome this threat. This finding suggests that successful learning occurred, but not due to the presence of gamification. Based on the positive learning outcomes, normal progression of training assessment would continue on to the behavioral level of the Kirkpatrick Model. This particular training study did not examine long-term learning outcomes such as behavior changes (e.g., pilots applying knowledge of how to overcome refraction) and organizational results (e.g., fewer weather-related aviation accidents).

Limitations

A number of limitations exist for this study. First, consider logistical limitations. One logistical limitation was the inability to observe participants while completing the study. While on one hand, this allowed a certain accuracy of how online training works (the trainers really never know what the learners are actually doing), it also opens the door for lack of experimental control (see the interval validity discussion below). The researchers put in safeguards (i.e., only compensating for 100% completion and asking participants to promise to watch the entire training video) to encourage participants to behave as honestly as possible. This may improve some of the threats to internal validity and increase honesty within the training but cannot account for everything.

Another logistical gap was the omission of the self-efficacy posttest, which prevented the researchers from directly identifying changes in self-efficacy, both between groups and within participants. Future efforts should include posttest self-efficacy, as it has years of research supporting its use as an effective attitudinal variable in training effectiveness studies.

Validity

Internal Validity

The study was designed to keep as many variables as possible constant; however, the study was administered remotely and asynchronously. This allows for variability in study environment among participants. For example, participants may have been multitasking while completing the training (e.g., eating dinner, talking to friends, walking to class, watching television) and did not give their full attention to the study at hand. To combat this, the researchers put in safeguards to encourage participants to pay attention to the training despite its unmoderated presentation. The study is also unable to take into account the different conditions (e.g., day of the week, time of day, number of ambient distractions, type of device) that participants experience while undergoing the training. Further, only about half of the participants who began the training continued all the way to completion. Despite these threats to internal validity, participants were randomly assigned to conditions prior to beginning the study; this may compensate for any unaccounted variables introduced during the study.

External Validity

In many cases, a tradeoff exists between internal and external validity. While this study may have lower internal validity, it could represent a realistic perspective on how pilots conduct online training. Participants were able to complete the training on their own time using whatever device (e.g., laptop, desktop computer, smartphone) that they would normally use to carry out online training. Additionally, the researchers were not next to the participants while they completed the training; this takes pressure off the participants that could lead to observer effects. Since this study did not take place in a laboratory setting, it may accurately reflect how an aviation student would perform on an online training course in their line of study, particularly if the training was optional (or an extra credit opportunity).

Construct Validity

The main dependent variables in this study were Engagement, Motivation, and Knowledge acquisition. In order to properly assess Engagement and Motivation, the researchers used previously validated measures after the training. For Knowledge acquisition, a team of experts altered a previously validated knowledge exam to create two parallel forms of radar questions. These knowledge tests were also reexamined during the first pilot study (Appendix H). Based on these precautions, the researchers determined that the effects of the study are due to the experimental manipulation rather than misinterpreted measures.

Further, the two independent variables were Narrative (present/absent) and Leaderboard (present/absent). Multiple pilot studies were conducted to ensure that the treatment was salient enough for participants (Appendix H). When the stimulus did not seem to be salient enough after the first pilot study, changes were made to increase frequency of gamification appearance. Further, the narrative was transformed from a lighter text-based story to a more serious video to accompany the storyline. The second pilot study employed manipulation checks to ensure that participants saw the gamification elements throughout the training. Based on these precautions, the researchers concluded that the gamification elements were properly applied to the study.

Statistical Validity

To determine if a statistically valid conclusion has occurred, it is essential to assess the sample size and statistical analyses employed in the study. A power analysis was conducted to determine the appropriate sample size needed to find an effect. A sample of 36 (9 per condition) was proposed by the power analysis. This study obtained a sample of 41 (approximately 10 per condition), which met the needed number of participants. Further, a statistical model that used a multivariate analysis of variance was initially proposed for this study. After data collection, the researchers determined that the assumption of multicollinearity between the dependent variables was not met. Therefore, separate analyses of variance were conducted to increase potential for finding effects. Most notably, all dependent variables in the study (as measured by previously validated questionnaires) demonstrated strong internal consistency. Based on these measures, the researchers concluded that this study had adequate statistical validity.

Future Directions

Below is a summary of recommendations for future iterations of this research or related research:

- A future iteration may be needed to replicate the study in-person with participants to determine if online training is a less effective tool.
- Future iterations could lower the fictitious leaderboard scores even more to create more optimism going into the training.
- Future iterations should utilize both a pretest and posttest self-efficacy scale to measure affective changes.

- A future study might utilize Narratives in a required setting (such as a real classroom) in order to encourage a sense of urgency to accompany the compelling story supplied by the training program.
- For future studies, the user type questionnaire can be utilized as a pre-screener in order to obtain a sample of Achiever user types large enough to find an effect.

Conclusion

It was hypothesized that gamification—particularly gamification elements targeted at pilots—would increase engagement, motivation, and knowledge acquisition in an online NEXRAD training program. Based on the results of this dissertation, the narrative and leaderboard treatments do not appear to have any effect on the dependent measures. Next steps may include conducting this research in a laboratory setting to increase internal validity. Overall, the online training program proved to be a useful tool in increasing radar-specific knowledge, but it could not be proven to be a result of the gamification elements themselves.

References

Air Safety Institute. (2008). Pilots Association Air Safety Foundation (2008). Nall report: Accident trends and factors for 2007. Frederick, MD: AOPA Air Safety Foundation.

Air Safety Institute. (2019). 28th Joseph T. Nall Report. Retrieved from https://www.aopa.org/training-and-safety/air-safety-in-stitute/accident-analysis/joseph-t-nall-report/non-com-mercial-fixed-wing

Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association: JMLA*, *103*(3), 152.

Allen, W. C. (2006). Overview and evolution of the ADDIE training system. *Advances in Developing Human Resources*, *8*(4), 430-441.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, 42, 331–342.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341–358.

Athanasopoulos, G., & Hyndman, R. J. (2011). The value of feedback in forecasting competitions. *International Journal of Forecasting*, 27(3), 845-849.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel psychology*, 41(1), 63-105.

Bandura, A. (2010). Self-efficacy. *The Corsini encyclopedia of psychology*, 1-3.

Bates, H. (2018, May). *5 Businesses using green gamification to save the world*. Retrieved from https://www.givingforce.com/5-businesses-using-green-gamification-to-save-the-world/

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and program planning*, *27*(3), 341-347.

Blickensderfer, B., Lanicci, J., Guinn, T., Thomas, R., Thropp, J., King, J., Ortiz, Y., Cruit, J., DeFillipis, N., Berendschot, K., McSorley, J., Kleber, J. (2019, January). *Combined report: aviation weather knowledge assessment & general aviation pilots' interpretation of weather products. (FAA Grant #14-G-010).* Unpublished project report. Embry-Riddle Aeronautical University, Daytona Beach, FL.

Bell, K. (2018). *Game on!: Gamification, gameful design, and the rise of the gamer educator*. JHU Press.

Branch, R. M. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer Science & Business Media.

Bureau of Labor Statistics. (2015, May). *American Time Use Survey: Leisure time on an average day.* Retrieved from https://www.bls.gov/tus/charts/chart9.txt

Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. Military Psychology, 7, 141 – 164.

Carroll, J. *The Adventure of Getting to Know a Computer.* Computer 15, 11 (1982), 49-58.

Chidester, T. R., Helmreich, R. L., Gregorich, S. E., & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. *The International Journal of Aviation Psychology*, *1*(1), 25-44.

Cobbett, E. A., Blickensderfer, E. L., & Lanicci, J. (2014). Evaluating an education/training module to foster knowledge of cockpit weather technology. *Aviation, space, and environmental medicine*, *85*(10), 1019-1025.

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2014). Flow. In *Flow and the foundations of positive psychology* (pp. 136-137, 232). Springer, Dordrecht.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134.

Deci, E., & Ryan, R. (1990). A Motivational Approach to Self: Integration in Personality. Nebraska Symposium on Motivation. *Perspectives on Motivation.* 38. 237-88.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of Self-Determination Research*. The University of Rochester Press: Rochester USA.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). *From game design elements to gamefulness: defining gamification.* In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15). ACM.

FAA. (2016). ADVISORY CIRCULAR. Retrieved from: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information /documentID/1030235

FAA Safety. (2006). *General Aviation Pilot's Guide to ... - faasafety.gov*. GA Pilots Guide to Preflight Weather Planning. Retrieved April 21, 2020, from https://www.faasafety.gov/files/gslac/courses/content/33/346/GA%20Weather%20Decision-Making%20Aug06.pdf

Federal Aviation Administration. (2010). Weather-related aviation accident study 2003–2007. Retrieved from http://www.asias.faa.gov/pls/apex/f?p=100:8:0::NO::P8_STDY_VAR:2

Fultz, A. J., & Ashley, W. S. (2016). Fatal weather-related general aviation accidents in the United States. *Physical Geography*. Taylor & Francis. Retrieved from https://doi.org/10.1080/02723646.2016.1211854

Goldstein, I. 1993. *Training in Organizations: Needs Assessment, Development and Evaluation.* Monterey, CA: Brooks/Cole. 3rd ed.

Goldstein, I., & Ford, J. K. (2002), *Training in Organizations: Needs Assessment, Development, and Evaluation*, 4th ed., Wadsworth, Belmont, CA.

Groh, F. (2012). Gamification: State of the art definition and utilization. Institute of Media Informatics Ulm University, 39, 31.

Guinn, T. A., & Rader, K. M. (2012). Disparities in weather education across professional flight baccalaureate degree programs. *The Collegiate Aviation Review International*, *30*(2).

Gustafson, K., & Branch, R. M. (2002). What is instructional design. *Trends and issues in instructional design and technology*, 16-25.

Hakulinen, L., Auvinen, T., & Korhonen, A. (2013, March). Empirical study on the effect of achievement badges in TRAKLA2 online learning environment. In 2013 *Learning and teaching in computing and engineering* (pp. 47-54). IEEE.

Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in human behavior*, *71*, 469-478.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). *Does Gamification Work? A Literature Review of Empirical Studies on Gamification.* In HICSS (Vol. 14, No. 2014, pp. 3025-3034).

Höllig, C. E., Tumasjan, A., & Welpe, I. M. (2018, January). The interaction of trait competitiveness and leaderboard design-an experimental analysis of effects on perceptions and usage intention. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Huotari, K. & Hamari, J. *Defining gamification: a service marketing perspective*. In Proceedings of the 16th International Academic MindTrek Conference, October 3-5, 2012, Tampere, Finland, ACM, pp. 17-22.

Jackson, M. (2016). *Gamification elements to use for learning* [PowerPoint slides]. Enspire. https://trainingindustry.com/content/uploads/2017/07/enspire_cs_ gamification_2016.pdf

Jia, Y., Liu, Y., Yu, X., and Voida, S. Designing Leaderboards for Gamification: Perceived Differences Based on User Ranking, Application Domain, and Personality Traits. *CHI 2017*, (2017), 1949–1960.

Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016, May). Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2001-2013).

Johnson, Cheryl & Bailey, Shannon & Mercado, Alyssa. (2020). Does Gamification Work? Analyzing Effects of Game Features on Learning in an Adaptive Scenario-Based Trainer. 10.1007/978-3-030-50788-6_36.

Johnson, D., Wyeth, P., Sweetser, P., & Gardner, J. (2012, September). Personality, genre and videogame play experience. In *Proceedings of the 4th International Conference on Fun and Games* (pp. 117-120).

Kapp, K. M. (2012). Games, gamification, and the quest for learner engagement. T and D, 66 (6), 64-68.

Kettanurak, V., Rammurthy, K. and Haseman, W.D. (2001), "User attitude as a mediator of learning performance improvement in an interactive multimedia environment: an empirical investigation of the degree of interactivity and learning styles", *International Journal of Human-Computer Studies*, Vol. 54 No. 1, pp. 541-83.

Koivisto, J. and Hamari, J. Demographic differences in perceived benefits from gamification. *Computers in Human Behavior 35*, (2014), 179–188.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology, 78(2), 311–328. https://doi.org/10.1037/0021-9010.78.2.311

Kroemer, K., Kroemer, H., & Kroemer-Elbert, K. (2001). *Ergonomics: How to design for ease and efficiency* (2nd ed.). Prentice Hall.

Lanicci, J. M., Guinn, T. A., King, J. M., Blickensderfer, B., Thomas, R., & Ortiz, Y. (2020). A Proposed Taxonomy for General Aviation Pilot Weather Education and Training. *Journal of Aviation/Aerospace Education & Research*, 29(1), 65-87.

Long, L. K., DuBois, C. Z., & Faley, R. H. (2008). Online training: The value of capturing trainee reactions. *Journal of Workplace Learning*.

Majuri, J., Koivisto, J., & Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd International GamiFIN Conference, GamiFIN 2018*. CEUR-WS.

Marczewski, A. (2013). The Intrinsic Motivation RAMP. Retrieved from https://www.gamified.uk/gamification-framework/the-intrinsic-motivation-ramp/.

Marczewski, A. (2015). User Types. In *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design* (1st ed., pp. 65-80). CreateSpace Independent Publishing Platform.

Marczewski, A. (2018). 52 Gamification Mechanics and Elements. Retrieved January 13, 2020, from https://www.gamified.uk/user-types/gamification-mechanics-elements/.

Maslow, A. (1943). A theory of human motivation. *Psychological review*, *50*(4), 370.

Michael, D., & Chen, S. L. (2005). Serious games: Games that educate, train, and inform. Muska & Lipman/Premier-Trade.

Miranda, A. T. (2015). *Gamified low-prevalence vigils* (Doctoral dissertation, Wichita State University).

Morgan, R.B. and Casper, W.J. (2000), “Examining the factor structure of participant reactions to training: a multidimensional approach”, *Human Resource Development Quarterly*, Vol. 11 No. 3, pp. 301-17.

Nacke, L. E., Bateman, C., & Mandryk, R. L. (2014). BrainHex: A neurobiological gamer typology survey. *Entertainment computing*, 5(1), 55-62.

O’Brien, H., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies,* 112(28-39).

Paharia, R. Who coined the term “gamification”? Quora, 2010. http://goo.gl/CvcMs.

Palomino, P. T., Toda, A. M., Oliveira, W., Cristea, A. I., & Isotani, S. (2019, July). Narrative for gamification in education: why should you care?. In 2019 IEEE 19th International *Conference on Advanced Learning Technologies (ICALT)* (Vol. 2161, pp. 97-99). IEEE.

Pandey, A. (2018, July 20). How You Can Use Storytelling In Corporate Training: Featuring 5 Effective Examples. Retrieved from https://elearningindustry.com/storytelling-in-corporate-training-featuring-effective-examples-use

Parson, S., Beringer, D., Knecht, W., Crognale, M. A., Wiegmann, D., & Beard, B. L. (2005). General aviation pilot’s guide to preflight weather planning, weather self-briefings, and weather decision making.

Pujolà, J. T., & Argüello, A. (2019). Stories or scenarios: implementing narratives in gamified language teaching. *GamiLearn.*

Rigby, S., & Ryan, R. (2011). *Glued to games: How video games draw us in and hold us spellbound: How video games draw us in and hold us spellbound*. AbC-CLIo.

Ritterfeld, U., Cody, M., & Vorderer, P. Serious Games: Mechanisms and Effects. Routledge, London, 2009.

Rozman, T., & Donath, L. (2019). The Current State of the Gamification in E-Learning: A literature Review of Literature Reviews. *Journal of Innovative Business and Management*, *11*(3).

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology, 43(3), 450.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual review of psychology*, *52*(1), 471-499.

Salas, E., Wilson, K. A., Priest, H. A., & Guthrie, J. W. (2006). Design, delivery, and evaluation of training systems. *Handbook of human factors and ergonomics*, 3, 472-512.

Shelton, S., & Alliger, G. (1993). Who's afraid of level 4 evaluation? A practical approach. *Training & Development*, *47*(6), 43-47.

Siang, A. C., & Rao, R. K. (2003, December). Theories of learning: a computer game perspective. In *Fifth International Symposium on Multimedia Software Engineering, 2003. Proceedings.* (pp. 239-245). IEEE.

Spence, J. and Helmreich, R. Achievement-related motives and behavior. In *Achievement and achievement motives: Psychological and sociological approaches*. Freeman, San Francisco (CA), 1983, 10–74.

Thomas, R., Cobbett, E., Lanicci, J., Cruit, J., & Blickensderfer, B. (2015). Using NEXRAD products during convective weather situations. *Embry-Riddle Aeronautical University: online course.* Retrieved from www.FAAsafety.gov

Thorndike, E.L. and Woodworth, R.S. (1901), "The influence of improvement in one mental function upon the efficiency of other functions", *Psychological Review*, Vol. 8 No. 4, pp. 247-261

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016, October). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 229-243). ACM.

van der Locht, M., van Dam, K., & Chiaburu, D. S. (2013). Getting the most of management training: the role of identical elements for training transfer. Personnel Review.

Weiss, V. (2016, March). Is there such thing as a pilot personality? Retrieved from: https://disciplesofflight.com/pilot-personality-traits/

Witte, J., Westbrook, R., Witte, M., & Cobanoglu, C. (2017, May). Gamification and training. In *Proc. of Global Conference on Education and Research* (GLOCER 2017).

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. "O'Reilly Media, Inc.".

Appendices

Appendix A

Demographic Questionnaire

1. Participant ID: ____________
2. Age: __________
3. Sex:
    - o Male
    - o Female
    - o Other
    - o Prefer not to answer
4. Highest Level of Education:
    - o Some high school
    - o High school diploma or equivalent
    - o Some college
    - o College degree
    - o Graduate degree
5. Degree or current major/minor: ____________________
6. Pilot Status
    - o None (not training to become a pilot)
    - o Student (in training to become a pilot)
    - o Private
    - o Commercial
    - o ATP
    - o CFI/CFII
7. What part did you receive the majority of your licenses under (over 50%)?
    - o Part 141-Flight Schools
    - o Part 142-Training Centers (ERAU)
    - o Part 61-Certification (Fixed Based Operator-FBO, etc)

Appendix B

Self-Efficacy Survey

I am confident in my ability to use NEXRAD data and products.

There are some tasks in using NEXRAD that I cannot do well.

When my performance is poor, it is due to my lack of ability to work with NEXRAD.

I doubt my ability to use NEXRAD.

I have all the skills needed to use NEXRAD very effectively.

Most people in my industry can use NEXRAD better than I can.

I am an expert at using NEXRAD.

My future in this industry is limited because of my lack of skills with NEXRAD.

I am very proud of my skills and abilities in using the NEXRAD.

I feel nervous when others watch me use NEXRAD.

Appendix C

User Engagement Scale (1-5 Likert Scale)

FA.1 I lost myself in this experience.
FA.2 I was so involved in this experience that I lost track of time.
FA.3 I blocked out things around me when I was using the training program.
FA.4 When I was using the training program, I lost track of the world around me.
FA.5 The time I spent using the training program just slipped away.
FA.6 I was absorbed in this experience.
FA.7 During this experience I let myself go.
PU.1 I felt frustrated while using this the training program.
PU.2 I found this the training program confusing to use.
PU.3 I felt annoyed while using the training program.
PU.4 I felt discouraged while using this the training program.
PU.5 Using this training program was taxing.
PU.6 This experience was demanding.
PU.7 I felt in control while using this training program.
PU.8 I could not do some of the things I needed to do while using the training program.
AE.1 This training program was attractive.
AE.2 This training program was aesthetically appealing.
AE.3 I liked the graphics and images of the training program.
AE.4 The training program appealed to the visual senses.
AE.5 The screen layout of the training program was visually pleasing.
RW.1 Using the training program was worthwhile.
RW.2 I consider my experience a success.
RW.3 This experience did not work out the way I had planned.
RW.4 My experience was rewarding.
RW.5 I would recommend the training program to my family and friends.
RW.6 I continued to use the training program out of curiosity.
RW.7 The content of the training program incited my curiosity.
RW.8 I was really drawn into this experience.
RW.9 I felt involved in this experience.
RW.10 This experience was fun.

Appendix D

Trainee Reactions

The course content was clear.

I could easily understand course content.

I was able to navigate through course content.

I found the course easy to use.

I was satisfied with presentation of course content.

I had a positive learning experience.

Appendix E

Intrinsic Motivation Inventory (IMI)

1. I didn't really have a choice about doing this task.
2. I did this activity because I wanted to.
3. I think doing this activity could help me to _____.
4. I did not feel nervous at all while doing this.
5. I was very relaxed in doing these.
6. I am satisfied with my performance at this task.
7. I enjoyed doing this activity very much.
8. I think I am pretty good at this activity.
9. I tried very hard on this activity.
10. I thought this activity was quite enjoyable.
11. I think this is important to do because it can _____.
12. This was an activity that I couldn't do very well.
13. While I was doing this activity, I was thinking about how much I enjoyed it.
14. I was anxious while working on this task.
15. I was pretty skilled at this activity.
16. This activity did not hold my attention at all.
17. I felt very tense while doing this activity.
18. I felt like it was not my own choice to do this task.
19. I felt like I had to do this.
20. I believe doing this activity could be beneficial to me.
21. I believe this activity could be of some value to me.
22. This activity was fun to do.
23. I think this is an important activity.
24. I would be willing to do this again because it has some value to me.
25. I felt pressured while doing these.
26. I did this activity because I had to.
27. I would describe this activity as very interesting.
28. It was important to me to do well at this task.
29. I did this activity because I had no choice.
30. I didn't put much energy into this.
31. I believe I had some choice about doing this activity.
32. After working at this activity for a while, I felt pretty competent.
33. I put a lot of effort into this.
34. I didn't try very hard to do well at this activity.
35. I think that doing this activity is useful for _____.
36. I think I did pretty well at this activity, compared to other students.
37. I thought this was a boring activity.

<Scoring>
Intrinsic Motivation Dimensions

Interest/ Enjoyment: 7, 10, 13, 16, 22, 27, 37
7. I enjoyed doing this activity very much.
10. I thought this activity was quite enjoyable.
13. While I was doing this activity, I was thinking about how much I enjoyed it.
16. This activity did not hold my attention at all. *
22. This activity was fun to do.
27. I would describe this activity as very interesting.
37. I thought this was a boring activity. *

Perceived Competence: 6, 8, 12, 15, 32, 36
6. I am satisfied with my performance at this task.
8. I think I am pretty good at this activity.
12. This was an activity that I couldn't do very well. *
15. I was pretty skilled at this activity.
32. After working at this activity for a while, I felt pretty competent.
36. I think I did pretty well at this activity, compared to other students.

Effort/ Importance: 9, 28, 30, 33, 34
9. I tried very hard on this activity.
28. It was important to me to do well at this task.
30. I didn't put much energy into this. *
33. I put a lot of effort into this.
34. I didn't try very hard to do well at this activity. *

Pressure/ Tension: 4, 5, 14, 17, 25
4. I did not feel nervous at all while doing this. *
5. I was very relaxed in doing these. *
14. I was anxious while working on this task.
17. I felt very tense while doing this activity.
25. I felt pressured while doing these.

Perceived Choice: 1, 2, 18, 19, 26, 29, 31
1. I didn't really have a choice about doing this task. *
2. I did this activity because I wanted to.
18. I felt like it was not my own choice to do this task. *
19. I felt like I had to do this. *
26. I did this activity because I had to. *

29. I did this activity because I had no choice. *
31. I believe I had some choice about doing this activity.

Value/ Usefulness: 3, 11, 20, 21, 23, 24, 35

3. I think doing this activity could help me to _____.
11. I think this is important to do because it can _____.
20. I believe doing this activity could be beneficial to me.
21. I believe this activity could be of some value to me.
23. I think this is an important activity.
24. I would be willing to do this again because it has some value to me.
35. I think that doing this activity is useful for _____.

* Reversed Items: 1, 4, 5, 12, 16, 18, 19, 26, 29, 30, 34, 37

Appendix F

Knowledge Tests

Pretest

1. LO 1: Objects along the radar path will scatter the energy back to the receiver. MORE objects of the same size in a given volume return ____ energy signals.

    a. stronger

    b. weaker

    c. bent

    d. scattered

2. LO1: The radar emits a pulse of _______________ energy.

    a. electromagnetic

    b. radioactive

    c. kinetic

    d. gravitational

3. LO 2: When a radar unit emits energy, the energy returned back from rain and hail is called _________.

    a. precipitation

    b. reflectivity

    c. ground clutter

    d. cone of silence

4. LO 2: Going from 10 dBZ to 40 dBZ implies Z (reflectivity) is _______ times greater.

    a. 10

    b. 100

    c. 1000

d. 10,000

5. LO 3: Fog, when observed on a radar, would most likely result in ______ dBZ of reflectivity.

    a. 75

    b. 50

    c. -25

    d. 0

6. LO 3: What weather phenomenon is associated with very high numbers (~ 75) on the reflectivity scale?

    a. Icing

    b. Fog

    c. Hail

    d. Snow

7. LO 4: The Cone of Silence refers to __________.

    a. the area above a radar unit that does not sense data

    b. a clear area near heavy precipitation

    c. a region with little to no cloud coverage

    d. All of the above

8. LO 4: Which of the following is NOT true about beam spreading?

    a. It produces a decrease in resolution with distance from the radar.

    b. It can make two identical storms at different distances from the radar appear different.

    c. It makes the same object appear larger close to the radar and smaller farther away from the radar.

    d. It can show a line of storms breaking up as it nears the radar.

9. LO 5: When ________ occurs, the radar will indicate precipitation at locations where satellite imagery is showing free of clouds

a. anomalous propagation

b. cone of silence

c. latency

d. base reflectivity

10. LO 5: Superrefraction occurs when the beam shoots _______________ than normal.

a. higher

b. lower

c. more clearly

d. more spread out

11. LO 6: What condition does NOT impact radar beam refraction?

a. Humidity

b. Temperature

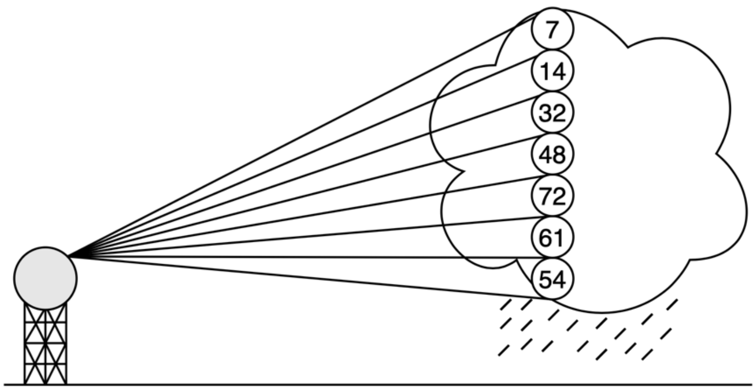
c. Cloud coverage

d. Air density

12. LO 7: All of the following are true about NEXRAD Clear Air Mode except it ________.
a. automatically starts up in Clear Air Mode
b. may detect dust, pollen, birds, or bugs
c. rotates faster so images update quicker
d. frequently shows ground clutter

13. LO 7: NEXRAD Precipitation Mode __________.
a. is used when there is no significant rain detected
b. is more sensitive (rain is harder to detect)
c. rotates slowest
d. can see much higher in the atmosphere

14. LO 8: The below picture shows NEXRAD radar reflectivity values at different altitudes within the center of the storm resulting from different radar elevation scans. Values shown are decibels of reflectivity. A base reflectivity image would display which value at the location of the storm?



a. 32

b. 7

c. 72

d. 54

15. LO 8: This radar product displays the echo reflectivity resulting from an individual radar's lowest elevation scan.
    a. Base reflectivity
    b. Composite reflectivity
    c. Radar mosaic
    d. Radar storm cell reflectivity

16. LO 8: Composite reflectivity displays the location of _______________ reflectivity value in the volume scan regardless of altitude.

a. minimum

b. maximum

c. average

d. estimated

17. LO 9: If multiple radars' reflectivities intersect when a mosaic is created, the _________ reflectivity is displayed.
    a. lowest

b. highest

c. base

d. middle

18. LO 9: The combination of multiple radar's reflectivity data onto single image is called ____.

a. composite reflectivity

b. a mosaic

c. precipitation mode

d. refraction

19. LO 10: Radar images are typically _______ old.

a. a few seconds

b. a few minutes

c. about half an hour

d. a few hours

20. LO 10: What causes radar data latency?

a. Data from multiple radar sites being processed then uploaded

b. Increased cloud coverage in the area slows down transmission

c. Differences in atmospheric density between the current area and data source

d. Poor inflight data connection

Posttest

1. LO 1: Objects along the radar path will scatter the energy back to the receiver. LARGER objects return ____ energy signals.

a. stronger

b. weaker
c. bent
d. scattered

2. LO 1: RADAR stands for ______________ Detection and Ranging.
   a. Radio
   b. Radiation
   c. Ratio
   d. Rain

3. LO 2: In what unit is reflectivity reported?
   a. kt
   b. SM
   c. dBZ
   d. kW

4. LO 2: If reflectivity is 10^{-3}, then reflectivity converted to dBZ is ______.
   a. -30
   b. .33
   c. 30
   d. 1000

5. LO 3: A reflectivity value of 35 dBZ is considered __________.
   a. Light
   b. Moderate
   c. Heavy
   d. Extreme

6. LO 3: A reflectivity value of 55 dBZ is considered __________.
    a. Light
    b. Moderate
    c. Heavy
    d. Extreme

7. LO 4: Radar beams travel _______.
    a. slower than the speed of sound
    b. at the speed of sound
    c. slower than the speed of light
    d. at the speed of light

8. LO 4: A radar beam bends due to ___________.
    a. differences in atmospheric density
    b. curvature of the earth
    c. cloud coverage
    d. heavy precipitation

9. LO 5: Ducting is a type of _________.
    a. subrefraction
    b. superrefraction
    c. anomalous propagation
    d. false echo

10. LO 5: When a radar beam overshoots objects that would usually be detected, _____ has occurred.

a. **subrefraction**
b. superrefraction
c. anomalous propagation
d. false echo

11. LO 6: How does one mitigate the effects of refraction?
    a. Check multiple radar sites
    b. Check satellite imagery
    c. Understand that knowing exactly how the current conditions will impact a radar beam is difficult
    d. **All of the above**

12. LO 7: The two modes of radar are _________.
    a. category 1 and 2.
    b. **clear air and precipitation.**
    c. heavy and light.
    d. the radar only has one mode.

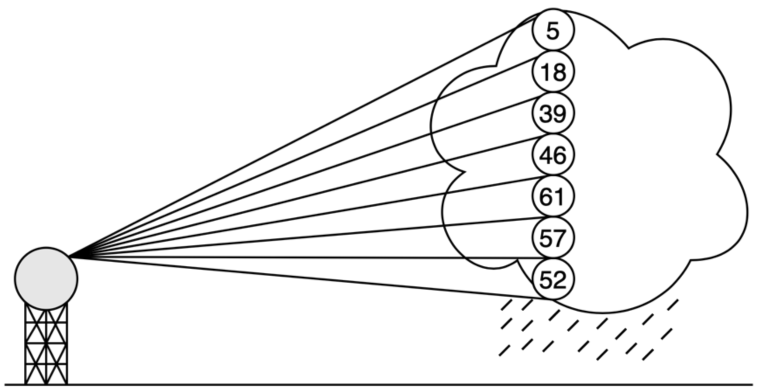
13. LO 7: Which of the following is NOT a Precipitation Mode VCP?
    a. **31**
    b. 21
    c. 11
    d. 121

14. LO 8: ________ reflectivity displays information from the lowest elevation angle scan.
    a. **Base**
    b. Composite
    c. Precipitation
    d. Clear Air

15. LO 8: This radar product displays the maximum reflectivity within the volume scan, regardless of altitude.

a. Base reflectivity
b. Composite reflectivity
c. Radar mosaic
d. Radar storm cell reflectivity

16. LO 8: The below picture shows NEXRAD radar reflectivity values at different altitudes within the center of the storm resulting from different radar elevation scans. Values shown are decibels of reflectivity. A composite reflectivity image would display which value at the location of the storm?



a. 5

b. 39

c. 61

d. 52

17. LO 9: The data used to create mosaics in the cockpit come from _______.

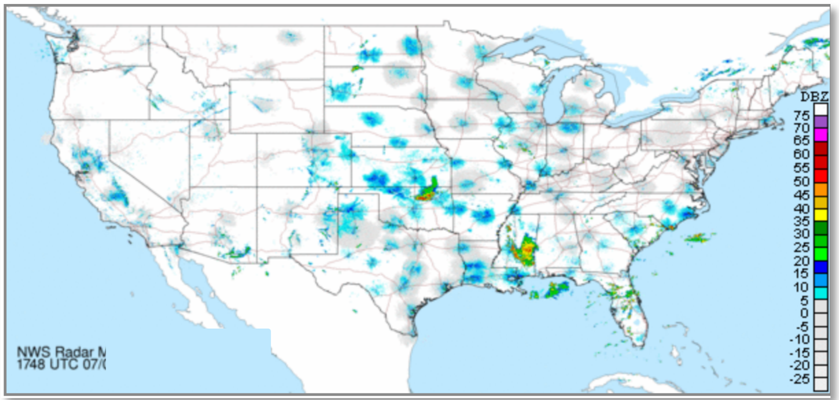
a. Vendors like WSI

b. Composite Reflectivity

c. Echo Tops

d. Radar Assembled Message

18. LO 9: What type of radar image is shown below?



a. Base Reflectivity

b. Clear Air Reflectivity

c. Mosaic

d. None of the Above

19. LO 10: Higher resolution radar images __________.

a. Allow pilots to maneuver closer to storms

b. Are uploaded faster than low resolution images

c. Are pulled from different radar collection sites

d. Have just as much data latency as low resolution images

20. LO 10: Ground-based radar uplink information can sometimes be as much as ________ old.

a. One hour

b. 15-20 minutes

c. 4-5 minutes

d. 30-40 seconds

Appendix G

User Type Questionnaire

The HEXAD Gamification User Types Questionnaire: Background and Development Process - Appendix

HEXAD Gamification User Types Questionnaire

Recommended scale: 7-point Likert scale from 1 = "strongly disagree" to 7 = "strongly agree".

To calculate how representative each user type is for a user, the user type scores for all items relating to the same type are to be added up with the maximum score per type being 35 (100%).

Nr.	User type	Item
1	Socializer	Interacting with others is important to me.
2	Philanthropist	It makes me happy if I am able to help others.
3	Free Spirit	It is important to me to follow my own path.
4	Socializer	I like being part of a team.
5	Disruptor	I like to provoke.
6	Achiever	I am very ambitious.
7	Player	I like competitions where a prize can be won.
8	Socializer	It is important to me to feel like I am part of a community.
9	Free Spirit	I often let my curiosity guide me.
10	Philanthropist	I feel good taking on the role of a mentor.
11	Disruptor	I like to question the status quo.
12	Socializer	It is more fun to be with others than by myself.
13	Player	Rewards are a great way to motivate me.
14	Free Spirit	I like to try new things.
15	Achiever	I like defeating obstacles.
16	Player	I look out for my own interests.
17	Philanthropist	I like helping others to orient themselves in new situations.
18	Disruptor	I see myself as a rebel.
19	Socializer	I enjoy group activities.
20	Achiever	It is important to me to always carry out my tasks completely.
21	Free Spirit	I prefer setting my own goals.
22	Disruptor	I dislike following rules.
23	Philanthropist	I like sharing my knowledge.
24	Achiever	It is difficult for me to let go of a problem before I have found a solution.
25	Player	Return of investment is important to me.
26	Free Spirit	Being independent is important to me.
27	Achiever	I like mastering difficult tasks.

Appendix H

Weather Training Questionnaire

For the following, please rate how much training you have had for each:

	None at all	A little	A moderate amount	A lot	A great deal
Basic weather theory	○	○	○	○	○
Single Site RADAR	○	○	○	○	○
National Mosaic RADAR	○	○	○	○	○
RADAR Coded Message (RCM)	○	○	○	○	○
1800-WXBrief (Call-in)	○	○	○	○	○
1800wxbrief.com (Leidos)	○	○	○	○	○

Appendix I

Pilot Studies

An original set of twenty-four participants were assessed to determine initial group differences and provide study design recommendations. Eight participants, based on time taken to complete the study, did not thoroughly watch the training video and were eliminated from analyses, which left a total of 16. A one-way MANOVA was conducted to assess group differences based on motivation score and performance score. Table 6 shows the descriptive statistics for the pilot study data. The results were not significant (Wilk's $\lambda = .797$, $F(6,22)=.439$, $p = 0.845$, $\eta p^2=0.107$).

DV	Condition	Mean	Std. Deviation	N
Motivation Score	Narrative + Leaderboard	4.99	0.54	4
	Narrative	5.19	0.40	5
	Leaderboard	4.58	1.14	3
	No Gamification	5.38	0.60	4
	Total	**5.08**	**0.65**	**16**
Performance Score	Narrative + Leaderboard	13.75	16.52	4
	Narrative	14.00	27.02	5
	Leaderboard	5.00	17.32	3
	No Gamification	22.50	17.08	4
	Total	**14.38**	**19.57**	**16**

Upon further investigation of the Motivation subscale *Effort/Importance*, the control condition (Condition 4) scored the highest and had the lowest standard deviation. This discrepancy in effort between groups could account for the lack of effect found between conditions. Further inspection of the knowledge test pre-test and post-test revealed low average scores on both, regardless of condition (see Table 7).

		Leaderboard	
		Present	Absent
Narrative	Present	Pretest: *M* = *43.8* ± 8.5 Posttest: *M* = *57.5* ± *15*.5	Pretest: *M* = 38.3 ± 15.3 Posttest: *M* = 43.3 ± 20.8
	Absent	Pretest: *M* = *49.0* ± *6*.5 Posttest: *M* = 63.0 ± 26.6	Pretest: *M* = 50.0 ± 17.3 Posttest: *M* = 72.5 ± 5.0

A second pilot study was briefly conducted to confirm that changes to the gamification elements did increase its salience in the training study. A sample of 6 participants were asked after they completed the study to relay their leaderboard scores, explain the story, or both depending on their experimental condition. They were told that this would have no impact on their compensation. All 6 participants correctly answered the questions, so the researchers continued with the training study as is.