To Gamify or Not? On Leaderboard Effects, Student Engagement and Learning Outcomes in a Cybersecurity Intervention

Mac Malone
tydeu@cs.unc.edu
UNC Chapel Hill

Yicheng Wang
yicheng@cs.unc.edu
UNC Chapel Hill

Kedrian James
kedjames@cs.unc.edu
UNC Chapel Hill

Murray Anderegg
anderegg@cs.unc.edu
UNC Chapel Hill

Jan Werner
jjwerner@cs.unc.edu
UNC Chapel Hill

Fabian Monrose
fabian@cs.unc.edu
UNC Chapel Hill

Abstract

We present a gamified learning experience for cybersecurity education that is designed to provide learners with an understanding of the knowledge and techniques needed to solve everyday problems while simultaneously immersing them in a competitive environment. We provide a framework for measuring skills demonstrated by students within an active learning setting where the primary focus is on practical expertise. We also examine several unique aspects of designing such a gamified framework (*e.g.,* the game itself must be insecure enough to be "hackable", but secure enough not to be abused), and discuss how the framework was used to expose students to various security concepts.

We found that our gamified experience heavily engaged students. We also encountered many pain points during our intervention and discovered a number of important aspects of gamified settings that must be carefully considered. For one, the goals of a semi-structured gamified exercise can sometimes lead to learners discovering solutions that do not meet the desired learning objectives. Furthermore, The exploratory nature of such exercises can also lead learners down a rabbit hole that, without a proper "safety net", they may not exit. Finally, complex tasks modeled after real world applications can leave little room for error, frustrating students and limiting instructors' ability to accurately assess different levels of skill. Based on our experience designing this intervention, we provide a number of transferable recommendations. The challenges we faced and the lessons we learned can be invaluable to those considering gamification as a cybersecurity education strategy.

CCS Concepts

• **Security and privacy** → *Cryptanalysis and other attacks*; • **Applied computing** → **Education**;

Keywords

Cybersecurity Education; Gamification; Active Learning

SIGCSE '21, March 13–20, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8062-1/21/03…$15.00
https://doi.org/10.1145/3408877.3432544

ACM Reference Format:
Mac Malone, Yicheng Wang, Kedrian James, Murray Anderegg, Jan Werner, and Fabian Monrose. 2021. To Gamify or Not? On Leaderboard Effects, Student Engagement and Learning Outcomes in a Cybersecurity Intervention. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21), March 13–20, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3408877.3432544

1 Gamification for Cybersecurity

Gamification is a natural fit for computer science education, and its use is particularly appealing in cybersecurity [35]. For instance, games spark interest in students who want to develop skills in computer science (*e.g.,* so they can create games like the ones they played when they were kids), and most games employ a wide range of cybersecurity techniques (*e.g.,* anti-cheat mechanisms) — making it easy to connect the gamified experiences to real-world applications. Furthermore, the adversarial nature of cybersecurity meshes well with the competitive nature of games.

That said, the introduction of gamified content into educational settings has been hotly debated [1, 7]. For the most part, supporters [5, 11, 21] of gamification encourage teachers to integrate these methods into their classrooms because game elements enhance learning by increasing engagement and motivation. Also, they facilitate social learning. Senko and Dawson [25] found that a "wanting to win" mindset improves the performance of participants especially when they are accompanied by strategies that support feelings of mastery. Detractors, on the other hand, argue that games prompt powerful emotional responses (*e.g.,* curiosity, satisfaction, and frustration) and by including game elements in educational settings we may be creating high levels of stereotype threat or detrimental upward social comparisons — both of which have been shown to negatively influence a students' academic performance [8, 9, 20].

Besides gamification, active learning has also been shown to be helpful in the quest towards mastery of cybersecurity topics. Unfortunately, educators must overcome numerous obstacles in adopting such practices (*e.g.,* the non-trivial investment in workhours, making exercises fun and customizable [18, 19, 27]). Thus, it is not surprising that frameworks for supporting active learning of cyber security concepts, especially in gamified settings, are few and far between. Our vision is to explore the use of gamified active learning exercises in ways that keep its benefits and minimize its negatives. Therefore, to better understand the intricacies of introducing gamified elements into security courses, we built a preliminary framework — coined Riposte — to test several hypotheses.

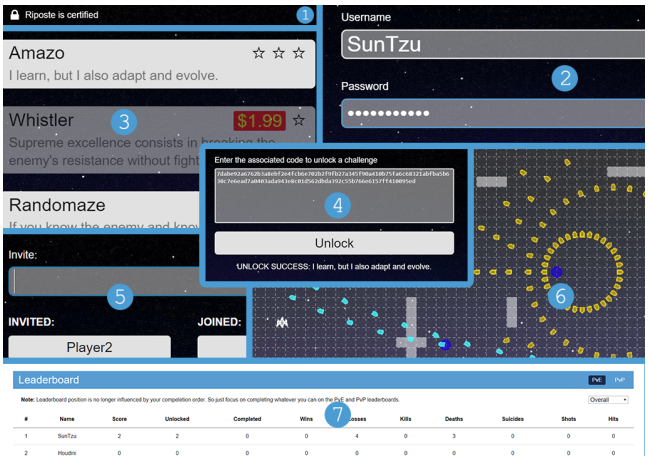

Figure 1: A collage of the components of Riposte: (1) Game certification, (2) login, (3) challenge selection, (4) unlocking, (5) lobby management, (6) gameplay, (7) leaderboard.

2 Riposte

The name, Riposte, comes from the fencing term meaning "a quick thrust given after parrying an opponent's lunge". The fencing dynamic accurately reflects the interplay between blackhats and security professionals (*e.g.,* a hacker exploits a vulnerability (lunges) then experts close the security hole (parry) and track down the blackhat (riposte)). The core of the learning experience is a browser-based game. A collage of the components of the game can be seen in Figure 1. Each student is provided with credentials that they use to log into the game (②). Players control a game unit (a spaceship) which they can move freely around a 2D board. They can fire projectiles, lay mines, and otherwise engage hostile enemy units. A demonstration of such gameplay can be seen in (⑥).

Riposte has two main styles of play: player versus environment (PvE) and player versus player (PvP). In PvE, players first select a challenge (③). Challenges have varying goals. Most involve defeating one or more AI game units by reducing their health to zero, but some involve solving a maze or surviving until a timer expires. The framework also supports certification of gameplay (*e.g.,* players need to upload a valid certificate to certify their victories (①)). By default, challenges can be locked and players have to unlock them via an *unlock code* (④). While some of these unlock codes are given to the students, others require the application of learned skills (*e.g.,* cryptanalysis or reverse engineering) to unlock the challenges. In PvP, players gather in a lobby and then join the game together (⑤). The players compete in a free-for-all where the last player standing wins. Results are displayed on a leaderboard (⑦). The incorporation of a leaderboard was done to motivate students via competition.

Assessing Learning Outcomes via a Skills Framework

In the design of our learning framework, we are exploring strategies for testing attained skills in the context of challenge-based learning. More specifically, since each exercise contains a practical test of knowledge — where the student has to apply the knowledge they learned to solve a problem — all except the lowest skill level involve some form of application.

As our goal is to measure students' mastery of concepts as a potential future practitioner, we place a lower emphasis on their ability to be up-to-date with recent developments. We associate a list of "action verbs" a la Parrish et al. [22] to help readers get a better sense of expectations when we say that a learner demonstrates a specific skill level. Our categorization of skill levels is:

❶ *Knowledge*: Acquired the knowledge taught and can manifest that by answering related questions. Most exams test at this level. **Verbs**: define, recognize, memorize, categorize.

❷ *Demonstration*: Can apply learned knowledge in a previously demonstrated way. Show-your-work exams and small assignments generally fall here. **Verbs**: replicate, reproduce, demonstrate, validate.

❸ *Adaption*: Can apply learned knowledge in contexts not seen before. Most traditional assignments tend to fall under this category. **Verbs**: adapt, paraphrase, expand, modify.

❹ *Familiarity*: Demonstrates an understanding of the problem domain and can independently close any knowledge gaps they may have when completing a task within it. Large assignments and small projects often fall here. **Verbs**: investigate, research, explore, analyze.

❺ *Cross-Domain Synthesis*: Demonstrates familiarity in several groups of concepts across different domains, and can draw on and synthesize their skills from these domains to complete a challenging task. Large projects fall in this category. **Verbs**: relate, compare, contrast, synthesize, survey.

❻ *Innovation*: Demonstrates deep understanding in a given domain to the point of inventing something new. Thesis-level work and particularly novel projects can demonstrate this level of skill in an area. **Verbs**: create, design, develop, hypothesize, theorize, invent.

❼ *Mastery*: Exhibits a deep understanding for a subject to the point of being able to direct, advise, and teach others. It is often hard to observe skill at this level in a conventional assignment, but successful group projects can sometimes reach this level. **Verbs**: teach, supervise, assess, advise, lead.

To assess these skills, we attempted to seamlessly incorporate five over-arching design principles. Achieving these goals simultaneously was non-trivial, especially given the nature of a cybersecurity course — **the game must be insecure enough to be hackable, but secure enough not to be abused.**

Usability: To support cross-platform portability, our learning experience is centered around a web-based game, wherein learners analyze aspects of the game's design using builtin tools (*e.g.,* Chrome's Developer Tools). We took special care to ensure that the interface is user-friendly and interacts with technologies (*e.g.,* JavaScript) that students are already familiar with.

Modularity: As this is primarily a framework for supporting education, challenges must be designed to teach some aspect of security (or computer science in general). At present, some challenges highlight simple client modifications, while others push the learner to gain a more detailed understanding of the game's logic. To achieve this, we required the ability to disallow certain kinds of modifications (*e.g.,* easier ones) on a *per-challenge* basis. Doing so mandated that challenges be modular and encapsulated such

that vulnerabilities of one are not necessarily shared across all. To support that capability, much of Riposte was parameterized by the active challenge and use those properties to determine what can and can not be done — both in terms of security and in gameplay.

Flexibility: One challenge with designing a highly modular game is determining the right place to put different aspects of game logic. Given that students need to demonstrate understanding of a topic (*e.g.,* reverse engineering) by *deliberately modifying the client*, we need to facilitate ample opportunity for doing so, while simultaneously protecting any logic we do not want learners to manipulate. To achieve this flexibility without sacrificing modularity, we designed the framework in a way that the client and server treat each other as black boxes, with only the communication protocol having defined semantics. The server handles messages produced by the client according to constraints of the current challenge (*e.g.,* if firing is disabled, the server ignores corresponding messages). The client provides a user interface that visualizes the information it receives and converts user interaction into messages sent to the server.

Stability: Our goal is to support hands-on exercises in both offensive and defensive security. Thus, students are expected to be able to modify the client, bombard the server, and, inevitably, break things — while not impeding the progress of other learners. To that end, our current design is distributed: each student is transparently provisioned a virtual machine (VM) wherein they conduct coursework. This VM hosts a container that launches a protected Riposte *satellite*. When a player logs into the the *master* Riposte (the one that hosts the game website), the master redirects the client to the respective satellite instance where the majority of gameplay logic takes place. To prevent students from leveraging any vulnerabilities in Riposte to (inadvertently or intentionally) attack the infrastructure, Riposte instances are sandboxed.

Transparency: To support a learning framework, one must be able to transparently monitor students' activity in order to accurately assess their performance. Thus, we record a wide variety of information about their progress. Some is made public (*e.g.,* via a leaderboard), but some is kept private.

3 Example Usage of the Learning Framework

To assess the versatility of the Riposte framework and to study the effects of gamified learning in a classroom setting, we developed and taught the following curriculum consisting of several exercises in the fall semester of 2019 to a class of 49 students (86% male).[1] Each exercise was designed to test an aspect of the student's security knowledge. Most exercises lasted for about a week, and students worked in teams of two. All the exercises were tied to the Riposte game, providing an immersive gamified experience throughout the semester. We note that the curriculum is not a recommendation for specific topics to be covered in a cybersecurity class, but rather an example of a gamified learning experience from which we observed what worked and what did not.

From our experience, the semi-structured nature of gamified exercises allowed students to use their creativity to go beyond what was asked of them. For brevity, we only discuss 3 (out of 9) exercises we created. The others exposed students to traffic inspection and manipulation, protocol reverse engineering (*e.g.,* API discovery), web security, binary reverse engineering, and using asymmetric cryptography. In what follows, we describe these three exercises, explain how they connect to the game, discuss what areas of computer security was being tested, and provide two metrics of difficulty: the *skill floor* (*i.e.,* the minimum skill level required for success), and the *skill ceiling* (*i.e.,* the maximum observable skill level). For grading purposes, demonstrating skill floor level proficiency is sufficient to earn a grade of C for the assignment.

Password Cracking: First, students performed offline password cracking of mnemonic [36] passwords. Learners were given files containing quotations and hashes of mnemonic passwords based on those quotations. They were tasked with writing a generator that produces mnemonic passwords using common word-to-character and character-to-character transformations. While ideas for basic transformations were provided, some complex transformations in the test set required independent research and out-of-the-box thinking to deduce. Second, the students performed online password cracking of specific game accounts. To mimic information available on social networks, learners were given a short bio for each target to inform their attacks. Once logged in as the target, students were tasked with beating them in PvP by joining a battle with their own account and then defeating the (motionless) victim. Successfully cracking all these accounts required applying the mnemonic password generator developed earlier along with new strategies (*e.g.,* dictionary attacks). The skill floor for that assignment is level ❹ for understanding password generation habits and general programming, while the skill ceiling is level ❻ as breaking some of the targets' passwords required ingenuity.

Client Side Modification: Students are introduced to the basic workflow of Riposte PvE: unlock a challenge using its code, accomplish the objective (*i.e.,* for the challenges in this assignment, defeat the bots), and move onto the next one. Most challenges contain bots that students cannot beat simply by playing normally. For example, one challenge, Blindspot, has bots positioned at 30° off of the cardinal directions from the player's avatar, and the player can only shoot at 45° offsets. To win, students need to inspect the network traffic to learn about the Riposte game API and then modify the client accordingly to defeat the bots. The skill floor for this exercise is level ❸ as it requires the student to apply the knowledge of client-side modification to different challenges, while the skill ceiling is level ❻ as the students can (and did) devise very creative client side hacks to win the challenges. While the students are shown the basic protocols used by the riposte game, they could also demonstrate innovation (level ❻) regarding traffic analysis by searching through the protocol space and utilizing hidden commands.

Malleability of Cipher Block Chaining Mode: Students are taught how challenges are locked using game codes. The 'correct' code used to unlock a challenge is a AES-CBC encrypted ciphertext based of the challenge description. Given access to a decoding oracle and a known ciphertext, the students are asked to write a program that would perform a CBC byte-flipping attack [23] to create the correct unlock code for any challenge. In class, students are shown the basics of the attack and use that knowledge to perform the attack on a 1-block message. They were also given a custom web UI within Riposte to visualize the attack as well as scripting tool for interacting with the decoding oracle. Learners must then extend

[1] All of the students had the same gamified learning experience.

the idea to generate a code (of several blocks) for a locked challenge. Students are also asked to research and find real world parallels to the attack on Riposte and explain how encryption was misused in each case. The skill floor is level ❹ (*familiarity*). Students can perform *Cross-Domain Synthesis* (level ❺) if they do their research and make connections to known vulnerabilities in existing software.

Assessing the Impact of Gamification

Given the debate over gamification in the classroom, one of our immediate questions regarding our exploratory format revolved around the use of the realtime leaderboard. Research has shown that individuals tend to make upward social comparisons, so it is likely that using leaderboards in academic environments would promote such comparisons. In small settings like ours (*e.g.,* less than 50-person classes), students are likely to know each other's identities (beyond a pseudonym on a screen) because they usually interact with each other, and when leaderboard position impacts one's grade, participants have incentives to stay on top of the leaderboard besides their interest in learning and playing the game. That is, rank in itself may be a motivating (or demotivating) force [16, 26]. Thus, we want to know *would students who ranked lower on the leaderboard be more likely to feel less interested and motivated by gamification?* (**Hypothesis 1a**) and *would the negative impacts of relative social comparisons on motivation and interest more heavily affects those that are doing worse?* (**Hypothesis 1b**).

We were also interested in knowing how gamification affects students' perception of their learning outcomes. Specifically, *would students who ranked lower on the leaderboard be more likely to have a lower perceived learning outcome, both in general and specific to the concepts tested?* (**Hypothesis 2a**), *would the the negative impacts of relative social comparisons on perceived learning outcomes more heavily affects those who are doing worse?* (**Hypothesis 2b**), and *would students who showed higher interest in gamification be more likely to have a higher perceived learning outcome, both in general and specific to the cybersecurity concepts tested?* (**Hypothesis 2c**).

Hypothesis 2c is based on the supposition that students who enjoy open ended, multi-solution challenges — a hallmark of gamified learning experiences — would be more likely to explore on their own and thus feel more learned as a result. Thus, an additional question was *would students interested in gamified learning be more likely to find the assignments intellectually challenging in a positive way?* (**Hypothesis 3**) We suspected that would be the case as early on in the integration of Riposte challenges into the course, and we noticed that certain students kept playing the game well after they attained near-perfect scores on the challenges. Playing for such extended periods may allow for more complete mastery of the taught material and better learning outcomes [15]. And so we wondered, *would students work on the assignment past what was needed to get full marks?*

To explore the questions, at the conclusion of the challenges, student were asked to complete questionnaires that included the following Likert-scale questions:

- The use of gamification improved my interest in the assignment. (*i.e., gamification interest*)
- The challenge helped me better understand ⟨topic in security⟩. (*i.e., specific learning outcome*)
- I learned a lot during the completion of the assignment. (*i.e., general learning outcome*)
- The assignment challenged me to think strategically ‖ was intellectually challenging ‖ challenged me to think outside of my comfort zone (*i.e., challenging & strategic thinking*)

We used that data to determine whether *correlations* exist between variables. The findings can not be used to indicate *causal* relationships as all data was collected after students finished the assignment and is thus uninformative about such relationships [17].

Summary of Findings: We performed a regression analysis where the independent variable was the student's leaderboard position and the dependent variable was their answer to the *gamification interest* question. We found no consistent evidence to support or reject **Hypothesis 1a**. We also performed two regression analyses using the same independent and dependent variables as those from Hypothesis 1a, but this time, looking at the correlation within students who scored above/below the median ("AM/BM") separately. While we were unable to find any statistically significant correlations, in the BM segments for the two challenges, we found weak negative correlation ($p < 0.15$) that warrant further study. This provides some evidence to support **Hypothesis 1b** that negative social comparisons more heavily affect students in the BM groups.

Regarding perceived learning outcomes, we performed a linear regression with leaderboard position as independent variables and *specific/general learning outcomes* as the dependent variables. We were unable to find statistically significant correlations in any of the assignments. Performing a similar analysis on the AM and BM segments separately, we were only able to find a weak ($\beta = -0.02$) albeit statistically significant ($p < 0.05$) negative correlation between leaderboard position and perceived general learning outcomes for **Hypothesis 2b**. In the case of **Hypothesis 2c** — with *gamification interest* as the independent variable and *specific/general learning outcomes* as the dependent variables — **we found consistent moderate positive correlations ($\beta = 0.45, p < 0.05$) between gamification interest and both types of perceived learning outcome across all but one of the challenges**.[2] And for perceived difficulty, **we found statistically significant ($p < 0.05$) positive correlations for Hypothesis 3** in two thirds of the applicable challenges.

Lastly, as a proxy to how much students enjoyed playing the game, we evaluated how much more learners played the game after achieving their first win. Two aspects are immediately apparent. First, some students seemed very engaged, playing more than 100 rounds for almost all of our challenges over 4-5 hours despite getting a win fairly early on. Second, students often kept playing the game even though doing so did not significantly improve the quality of their solution (three assignments had near zero medians). In fact, some ended up in a worse position than where they started (lower accuracy and more deaths). **These factors indicate that our students were heavily engaged and driven by the gamification aspect of the challenges instead of a rational approach to maximize performance.**

Overall, although the intervention could be deemed a success (*e.g.,* the course received some of the highest overall student evaluations in the department), there is still room for improvement.

[2] For the remaining one, we found positive correlation at $p < 0.10$.

Specifically, the results indicate that we were unable to sufficiently address the needs of those who were struggling. Reflecting upon this, we discuss some practical lessons and concerns below.

4 Challenges, Pitfalls, and Lessons Learned

Unlike other areas of computer science (*e.g.,* software development), where practitioners can leverage simplifying assumptions to quickly complete a task derived from an external need, cybersecurity practitioners need to be aware of, and repeatedly question, the validity of these simplifying assumptions to either prove the system's security or find exploitable weaknesses. As such, in this specific field, being able to find the right problems to solve is perhaps just as important as being able to solve them.

Our semi-structured gamified approach nicely mirrors this property — learners are given a distant goal and have to find out how to reach it. They explore the platform, discover its underlying assumptions, and develop ways to exploit them to achieve their goal. In most of our challenges, there are a host of assumptions the students can exploit, hidden at various depths within the system, and students are rewarded for capitalizing on these weaknesses.

To demonstrate this, consider the password security exercise. In part 1, students were tasked with coming up with various character-to-character and word-to-character transformations to generate a wide range of mnemonic passwords. While they were given some transformations in class and in the assignment text, relying on these alone was not sufficient to achieve a passing grade. The students therefore needed to explore and brainstorm in order to complete that part of the assignment. In part 2, the students had to guess the passwords of the targets from bios that mirrored information one might find on real social media. We provided a few examples of common password creation mistakes, but none of these could be applied as-is. Instead, good approaches required the student to get into the headspace of a victim and figure out how they would go about creating passwords [37]. Students needed to identify and properly understand the hints *and* adapt and develop strategies to generate the passwords within the family of passwords hinted at.

For some of the learners, this approached worked very well: they found the exploration aspect exhilarating and derived a high sense of accomplishment from finding the correct path. However, we identified three pedagogical drawbacks: imperfect correlation between learning objective and game goals, lack of a guardrail against depth-first thinking, and insufficient measurement of lower skill levels. We elaborate more on each of these points below.[3]

Correlation Between Learning Objective and Game Goals

By nature, there is no unique way of achieving the end goal of a semi-structured assignment. In an ideal semi-structured assignment, however, the various ways of achieving the end goal all require mastery over the intended learning objective. This, however, is tricky to get right, as the students do not necessarily know *a priori* what the learning objectives are. In the password cracking assignment, the learning objective is to understand the techniques people typically use to generate memorable yet relatively secure passwords. In part 1, this is tested via the student's ability to generate the common transformations needed to replicate a fixed set of mnemonic passwords we provide. In part 2, they are tasked with both identifying and exploiting information subtly hidden within the targets' social media presence, and then expanding upon their work in part 1 by developing guessing strategies beyond a single class of passwords (*i.e.,* mnemonics). However, not all students saw the correlation between the underlying learning objectives and the stated goals of the assignment. One student in particular perceived the first part as a mere password cracking exercise and employed a brute force approach to replicate our mnemonic set — without ever grasping why one would be doing this task. That is, the leaner knew that to replicate the answer set, one would sometimes need to transform an 'a' into a '4' or an 'e' to a '3', but never understood *why* someone would apply such a transformation when generating their password. As a result, the student got stuck on the latter part of the assignment because the learning objective of understanding people's password creation habits was never met.

Lack of Guardrail Against Depth-First Thinking

As stated before, in our intervention, students are encouraged and awarded for questioning assumptions and exploring seemingly far-fetched ideas. For the most part, we believe that this is a laudable as it models reality. However, being efficient during exploratory stages is a skill that is rarely explicitly covered in education, and our students were ill-prepared for our semi-structured assignment as a result. We did attempt to teach students what one should do to avoid going down unnecessary rabbit holes (*e.g.,* planning out an attack tree, trying low hanging fruits first, not being afraid to go back to brainstorming), but this was often insufficient. Many students were still thinking in a very "depth-first" fashion, insisting that they are on the right track despite the often extreme complexity of their supposed solutions. This led to a lot of frustration and often impeded their ability to master the intended learning objectives.

Insufficient Measurement of Lower Skill Levels

Another potential drawback of a complex semi-structured learning exercise is its high skill floor — a result of its requirement for innovation and lack of explicit instructions. For example, our password security assignment assesses students' ability to develop password guessing strategies instead of their ability to apply standard ones. This is also why we rated it with a skill floor of *familiarity* (level ❹). However, a side-effect of this is that we are unable to test students who have only mastered the material at lower skill levels. For the purpose of grading, there is no difference between someone who was able to efficiently (and exactly) apply the taught methods of guessing passwords versus someone who has no knowledge in the field. As such, this can be very discouraging for students who are at these lower levels, and can even give the perverse impression to students that their efforts are not valued and that it is, therefore, "not worth it" to try. This also has serious impacts on our ability to measure student growth, as it squashes a wide range of skills into a single point, making the desired type of "start-end" grade equivalency impossible to establish.

Based on these observations, we argue that a good semi-structured assignment must be placed within a structured framework that serves as a "safety-net" for students. This safety-net can take many forms, but at the least, it should (*i*) only assist the student when they have gone far off course — *i.e.,* to prevent circumventing the

[3]While we use the password cracking assignment as the running example to illustrate these drawbacks, they applied to our semi-structured approach in general.

exploratory nature of the assignment, and (*ii*) be accessible and clear enough for students to feel like their efforts are valued.

Looking back at our invention, our in-person office hours often served as this safety-net for our students: learners who were stuck would come to office hours, where we would provide more direct hints, explain the motivation behind parts of the assignments, pull students off of the wrong paths, and assess if someone is actively trying but just performing at a lower skill level. While this worked to a certain extent (evident by the fact that the feedback for the course was overwhelmingly positive, and a lot of people specifically praised how useful office hours were), it did not address any of the aforementioned problems completely. Alas, its success rests on time-consuming, close one-on-one interactions that are not scalable and ill-suited for teaching in the age of distance learning. Well-researched solutions to address these problems are warranted.

5 Toward a Winning Strategy

Moving forward, one way to address some of these concerns is to better elucidate the learning objectives of an assignment and make it clear what is required to achieve a *passing* grade. Having explicit sub-goals (*e.g.,* the next achievement in a series) would allow instructors to lower the measurable skill floor of an assignment, thereby allowing learners to demonstrate mastery of lower skill levels by completing the simpler sub-goals. Such rewards based on personal accomplishments might lessen frustration because learners would better perceive success during the early stages of the learning process [12]. Conceivably, achievements may even be used to unlock future challenges. Research shows that such learner-centered [34] approaches can help with both long-term goals (*e.g.,* being highest on the leaderboard) and short-term goals (*e.g.,* improving a self-determined number of places) [34]. If done correctly, we believe learners would find the experience more rewarding, especially when trying to solve semi-structured problems.

To provide virtual safety nets, one direction worth exploring is the use of open-sourced web platforms to better assess what the learners are doing in order to provide solution-specific feedback to struggling students. Such feedback may help upper bound the time and energy spent on incorrect approaches. Our plan is to integrate Jupyter notebooks into Riposte to monitor progress on coding tasks. We hope to report on our experience in future works.

As noted earlier, we choose to incorporate a very popular game element, namely a leaderboard,[4] as a prominent feature of the Riposte experience. Leaderboards are a popular gamification technique for enhancing engagement through social comparisons. They are also very popular in cybersecurity competitions. Our leaderboard was traditional in that the rankings of everyone's avatar was public. Our experience was consistent with that of Codish and Gilad [4], in that we found that leaderboards of this type motivate some, but was also be a demotivating factor for others. We recommend using more synergistic designs that promote the satisfaction of competence and autonomy [28] for most users. We are encouraged by the fact that autonomy-supportive leaderboards have been useful in setting behavioral goal changes in other contexts [10], and hope the same will be true in teaching cybersecurity concepts.

[4]Only a student's in-game ranking was displayed on the leaderboard, not the grade earned for the related assignment.

Another concern raised during our post-assessment was that although the framework was helpful in assessing students' proficiency, it fell short in assessing growth [2]. Conceptually, proficiency targets the minimum level of achievement that all students are expected to meet on their summative assessments — irrespective of incoming knowledge and experience. Consequently, proficiency-based mindsets can easily overlook student learning that did or did not occur as a result of a teacher's instruction [14]. Growth, on the other hand, compares the entering skill level of students to their final skill level. Arguably, growth potential as a cybersecurity practitioner might be even more important than proficiency on specific tasks. Thus, moving forward, we will explore how best to support growth mindsets.

6 Related Work

The "Principles of Computer Security Lab Manual" [31] provides exercises with instructions for educators. Unfortunately, while these exercises offer good introductory material, they only teach students how to use existing tools, without providing a good understanding of *what* the techniques employed by these tools are and *how* these techniques can be adapted for new scenarios. More widely adopted exercises are provided by the SEED labs [33] and the EDURange [32] projects. These standalone labs can be helpful, but we found that they fall short in their ability to engage students as they offer limited solution spaces for demonstrating levels of mastery.

Also germane are the growing number of courses that incorporate some form of "capture the flag" (CTF) activities [3, 6, 13, 29] or otherwise introduce gamified content [24] for enhancing cybersecurity skills. These classes can be widely popular, but the learning outcomes typically center on penetration testing on a particular threat or vulnerability. Švábenský et al. [30] take a twist on the prototypically CTF-based approach, providing an offering that focuses on assessing students' ability to use gamification techniques (*e.g.,* storyboarding, level design) to promote engagement on a specific topic in cybersecurity. Our desire to incorporate and evaluate gamification elements complement past efforts, but the ideas we explore to help students master various learning objectives in a fun and autonomy-supportive setting goes far beyond any of these efforts.

7 Conclusion

We reported on our experience designing, implementing, and evaluating a gamified learning strategy for teaching cybersecurity concepts. We outlined key considerations when designing cybersecurity exercises and discussed how we arrived at a tiered framework for evaluating a student's demonstrated skill in an active learning setting. We also reported on some of the pros and cons we observed when applying gamification in an educational setting. More research is needed to answer the question posed by the title, as it is currently unclear if the negative outcomes we experienced were due to our specific implementation of gamification or are something intrinsic to the teaching style. In short, while we support the use of gamification as a mechanism to promote hands-on learning in cybersecurity, additional work is needed on how best to minimize negative outcomes before gamification's full potential as a winning strategy for cybersecurity education [35] can be realized.

8 Acknowledgements

We thank Roman Rogowski, Deven Desai and Ryan Court for their contributions to early prototypes of the Riposte framework. We also express our gratitude to IEEE for an educational grant under the Try-CybSi initiative that supported the first author. We also thank Jonathan Dixon, Paul van Oorschat, Charles V. Wright and the anonymous reviewers for their insightful feedback.

References

[1] Shurui Bai, Khe Foon Hew, and Biyun Huang. 2020. Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review* 30 (2020).

[2] James S. Beri, Kenneth E. Berlinn, Mary E. Blackmon, Lynda S. Jackson, James Petrie, and Randy A. Leipa. [n.d.]. Measuring Student Growth: A Practical Guide to Educator Evaluation.

[3] Martin Carlisle, Michael Chiaramonte, and David Caswell. 2015. Using CTFs for an Undergraduate Cyber Education. In *Summit on Gaming, Games, and Gamification in Security Education*.

[4] David Codish and Ravid Gilad. 2014. Personality Based Gamification – Educational Gamification for Extroverts and Introverts. In *Conference for the Study of Innovation and Learning Technologies*.

[5] Thomas M. Connolly, Elizabeth A. Boyle, Ewan MacArthur, Thomas Hainey, and James M. Boyle. 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education* 59, 2 (2012), 661–686.

[6] Adrian Dabrowski, Markus Kammerstetter, Eduard Thamm, Edgar Weippl, and Wolfgang Kastner. 2015. Leveraging Competitive Gamification for Sustainable Fun and Profit in Security Education. In *Summit on Gaming, Games, and Gamification in Security Education*.

[7] Christo Dichev and Darina Dicheva. 2017. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education* 14 (Dec 2017).

[8] Pieternel Dijkstra, Hans Kuyper, Greetje van der Werf, Abraham P. Buunk, and Yvonne G. van der Zee. 2008. Social Comparison in the Classroom: A Review. *Review of Educational Research* 78, 4 (2008), 828–879.

[9] Adrian Dominguez, Joseba Saenz de Navarrete, Luis de Marcos, Luis Fernandez-Sanz, Carmen Pages, and Jose-Javier Martinez-Herraiz. 2013. Gamifying learning experiences: Practical implications and outcomes. *Computers & Education* 63 (2013), 380–392.

[10] Tracy Epton, Sinéad Currie, and Christopher J Armitage. 2017. Unique Effects of Setting Goals on Behavior Change: Systematic Review and Meta-Analysis. *Journal of Consulting and Clinical Psychology* 85 (2017), 1182–1198.

[11] J. Hamari, J. Koivisto, and H. Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *International Conference on System Sciences*. 3025–3034.

[12] Wang Hao and Sun Chuen-Tsai. 2011. Game reward systems: Gaming experiences and social meanings. In *DiGRA International Conference: Think Design Play*.

[13] C. Jordan, M. Knapp, D. Mitchell, M. Claypool, and K. Fisler. 2011. CounterMeasures: A game for teaching computer security. In *Annual Workshop on Network and Systems Support for Games*. 1–6.

[14] Lisa Lachlan-Hache and Marina Castro. 2015. Proficiency or Growth? An Exploration of Two Approaches for Writing Student Learning Targets. American Institutes for Research.

[15] Richard Landers and Amy Landers. 2015. An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance. *Simulation & Gaming* 45 (Apr 2015).

[16] Weiwen Leung. 2019. How Do One's Peers on a Leaderboard Affect Oneself?. In *CHI Conference on Human Factors in Computing Systems*. 1–11.

[17] Charles F. Manski. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60, 3 (1993), 531–542.

[18] Andrew McGettrick, Lillian N. Cassel, Melissa Dark, Elizabeth K. Hawthorne, and John Impagliazzo. 2014. Toward Curricular Guidelines for Cybersecurity. In *ACM Technical Symposium on Computer Science Education*. 81–82.

[19] Jelena Mirkovic, Melissa Dark, Wenliang Du, Giovanni Vigna, and Tamara Denning. 2015. Evaluating Cybersecurity Education Interventions: Three Case Studies. *IEEE Security and Privacy* 13, 3 (2015), 63–69.

[20] Dominique Muller and Marie-Pierre Fayant. 2010. On Being Exposed to Superior Others: Consequences of Self-Threatening Upward Social Comparisons. *Social and Personality Psychology Compass* 4, 8 (2010), 621–634.

[21] Cristina Muntean. 2011. Raising engagement in e-learning through gamification. *International Conference on Virtual Learning* (Jan 2011).

[22] Allen Parrish, John Impagliazzo, Rajendra K. Raj, Henrique Santos, Muhammad Rizwan Asghar, Audun Jazsang, Teresa Pereira, and Eliana Stavrou. 2018. Global perspectives on cybersecurity education for 2030: a case for a meta-discipline. *ACM Conference on Innovation and Technology in Computer Science Education* (2018).

[23] Mike Rosulek. 2020. *The Joy of Cryptography*.

[24] Z. Cliffe Schreuders, Thomas Shaw, Aimée Mac Muireadhaigh, and Paul Staniforth. 2018. Hackerbot: Attacker Chatbots for Randomised and Interactive Security Labs, Using SecGen and oVirt. In *USENIX Workshop on Advances in Security Education*.

[25] Corwin Senko and Blair Bryant Dawson. 2017. Performance-Approach Goal Effects Depend on How They Are Defined: Meta-Analytic Evidence From Multiple Educational Outcomes. *Journal of Educational Psychology* 109 (2017), 574–598.

[26] Emily Sun, Brooke Jones, Stefano Traca, and Maarten W. Bos. 2015. Leaderboard Position Psychology: Counterfactual Thinking. In *ACM Conference on Human Factors in Computing Systems*. 1217–1222.

[27] Christopher Theisen, Laurie Williams, Kevin Oliver, and Emerson Murphy-Hill. 2016. Software Security Education at Scale. In *International Conference on Software Engineering Companion*. 346–355.

[28] M. Vansteenkiste, J. Simons, W. Lens, K. Sheldon, and E. Deci. 2004. Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of personality and social psychology* 87, 2 (2004), 246–260.

[29] Giovanni Vigna, Kevin Borgolte, Jacopo Corbetta, Adam Doupé, Yanick Fratantonio, Luca Invernizzi, Dhilung Kirat, and Yan Shoshitaishvili. 2014. Ten Years of iCTF: The Good, The Bad, and The Ugly. In *Summit on Gaming, Games, and Gamification in Security Education*.

[30] Valdemar Švábenský, Jan Vykopal, Milan Cermak, and Martin Laštovička. 2018. Enhancing Cybersecurity Skills by Creating Serious Games. In *ACM Conference on Innovation and Technology in Computer Science Education*. 194–199.

[31] Richard Weiss, Jens Mache, and Erik Nilsen. 2013. Top 10 Hands-on Cybersecurity Exercises. *J. Comput. Sci. Coll.* 29, 1 (Oct 2013), 140–147.

[32] Richard S. Weiss, Stefan Boesen, James F. Sullivan, Michael E. Locasto, Jens Mache, and Erik Nilsen. 2015. Teaching Cybersecurity Analysis Skills in the Cloud. In *ACM Technical Symposium on Computer Science Education*. 332–337.

[33] Du Wenliang, Shang Mindgong, and Haizhi Xu. 2010. Enhancing Security Education with Hands-on Laboratory Exercises. In *Symposium on Information Assurance*.

[34] Kaitlyn M. Werner, Marina Milyavskaya, Emily Foxen-Craft, and Richard Koestner. 2016. Some goals just feel easier: Self-concordance leads to goal progress through subjective ease, not effort? *Personality and Individual Differences* 96 (2016), 237–242.

[35] Brad Wolfenden. 2019. Gamification as a winning cyber security strategy. *Computer Fraud & Security* 2019, 5 (2019), 9–12.

[36] Weining Yang, Ninghui Li, Omar Chowdhury, Aiping Xiong, and Robert W. Proctor. 2016. An Empirical Study of Mnemonic Sentence-Based Password Generation Strategies. In *Proceedings of the ACM Conference on Computer and Communications Security*. 1216–1229.

[37] L. Zhang-Kennedy, S. Chiasson, and R. Biddle. 2013. Password advice shouldn't be boring: Visualizing password guessing attacks. In *APWG eCrime Researchers Summit*. 1–11.