All Work and no Play – Game-Based vs. Text-Based Learning in IT-Security*

Marcus Gelderie[1] and Tamara Wanner[1]

[1] Aalen University of Applied Sciences
marcus.gelderie@hs-aalen.de, tamara.wanner@hs-aalen.de

Abstract. Game-based learning is a popular tool in IT-security. We study the effects of game-based learning in the context of phishing-emails in comparison with text-based learning by asking participants to classify previously unseen emails after completing their learning exercise. We correlated participants scores with their professional background of learners. Our results indicate that professional background is an important indicator of the success of game-based learning and suggest that learners with a technical background might benefit from text-based learning to a larger degree than from game-based learning.

Keywords: Game-based-learning; Text-based-learning; Gamification; Phishing; IT-Security

1 Introduction

The security of computer systems is not a purely technical artifact. It depends, to a large degree, on the behavior of humans interacting with that system. Phishing, the act of misleading a victim to perform an unwanted and harmful act (such as visiting a specific website or downloading a file), is well-known and statistically relevant example [1]. The mitigation of threats and attacks by training users is a topic of active research (e.g. [2, 3, 4, 5]). A significant subset of research studies the effects of gamification and game-based learning in this context (e.g. [3, 5]). The approaches used vary widely (see also related work below). Some approaches use truly game-based designs, where users play an actual game (e.g. [3, 5, 4]). Other approaches augment classical training approaches with gamified elements (e.g. [2]). Most of these approaches address different training needs (phishing URLs in [3, 5, 4] vs. security-awareness in requirements-engineering in [2]). Oftentimes, a benefit from gamification over classical approaches is reported. Gamification is, by and large, shown to be an effective tool. Moreover, gamification is typically reported to be more engaging.

The existence of different types of learners (such as auditory or visual) is a fact well-known to teachers across disciplines. Moreover, depending on their background, certain individuals will have different ways of reasoning about technology and its underlying,

* This work has been developed within the project ‘BAK Game’ which is funded by the German Federal Ministry for Economic Affairs and Climate Action.

“invisible” behavior: When a web-page fails to load, a computer scientist might think of specific problems, such as errors in DNS resolution, but less familiar users might only identify some unspecified “network problem”. This kind of background knowledge likely factors into such individual’s ability to benefit from different learning approaches. It might also factor into their opinion of the learning approach: whether it is perceived as refreshing or tedious, for example.

In this paper, we present a study of the success rate of game-based learning in comparison with more traditional text-based, when correlated with the professional background of the target audience. The learning goal in this paper is the ability to detect malicious mails (such as phishing). Specifically, we developed and tested a game-based learning approach to teach individuals about risks of malicious emails and the signs by which to recognize possibly malicious mails. The game appears to show an email program and learners “read” incoming emails, very similar to real-world scenarios. We also prepared traditional text-based learning material that teaches reader about the same signs of malicious mails (such as suspicious URLs or attempts at pressuring readers into downloading something). We then set out to answer four questions: is the game-based learning approach more effective at raising the detection rate of previously unseen mails, can that difference, if any, be attributed to a subject’s professional background, which approach, if any, leads to a better learning satisfaction with the method and with the outcome, and, lastly, do learners feel more motivated by one approach over the other? Our results indicate that text-based learning outperformed game-based learning and that this effect was most pronounced when learners come from a technical background. Nevertheless, learners are more satisfied with and motivated by our game-based approach in comparison with the text-based approach.

A related study of [6] compared multiple learning methods (training via general video/quiz, simulated phishing emails and a leaderboard). They conducted an experiment with a role play as an executive assistant in a normal working day where the participants had to answer work-related emails and doing work-related tasks while reporting phishing emails. Also, [7] used a role-playing quiz application but with a focus on password security. In our work the participants received no emails for a specific role.

Malicious URLs are studied in [5, 3, 8] using a game-based approach that teaches users to distinguish malicious URLs from legitimate ones. Users are presented links and must decide whether they are malicious or not. Correct answers are rewarded and wrong answers are punished and eventually lead to a game-over scenario. The study in [8] took place in education for students in school and compares different learning methods and finds that all participants in all learning-tasks have become better at correctly assessing phishing emails. The instructors-learning-methods had the best results.

Perrault et al. [9] conducted a study with an interactive online quiz (classifying of ten screenshots of email). A focus is on rapid feedback that can be provided by interactive tools. The author concludes that an interactive phishing quiz can impact collage students' awareness and behavioral intentions about phishing. While [9] focuses on the effectiveness, in particular, self-effectiveness, to phishing attempts, our paper measures effectiveness using a sample set of unseen mails and correlates the result with subjects' professional background.

Schreuders et al. [10] focus on students' assessed learning activities in higher education in teaching and learning by using a gamified module for computer security with open source software and virtual environment. In our work we focus on phishing emails and compare the gamified-task with text-based learning.

In [11] subjects are exposed to phishing emails during their normal work routine. This computer-based, but not gamified approach, is shown to be more effective than other learning methods.

2 Goals and Methodology

In the second chapter we present four research questions, explain our study design and compare the teaching approaches.

2.1. Research Goals and Study Design.

In the following, we outline a study designed to answer the ensuing research questions:

- Q1: Is game-based learning more effective than classical text-based learning?
- Q2: To what extent is a person's professional background relevant to the success?
- Q3: Do consumers of the game-based material report a higher satisfaction?
- Q4: Do consumers of the game-based material feel motivated to learn?

Our study was designed for a German speaking group of participants. Therefore, all questions and all training material is available in German only. Below, we give English translations of the questions relevant to this paper. Moreover, we chose to conduct our study at a college providing higher education. Again, this choice is reflected in the questions shown below. Our study is structured into four parts:

1. An initial questionnaire to assess participants background, learning preferences, questions about home office and phishing.
2. A learning task, which as either gamified or text-based.
3. A post-learning assessment of effectiveness by classifying four previously unseen emails as malicious or not.
4. A final questionnaire to measure participant's motivation and satisfaction with their learning method (text-based or game-based).

Participants are randomly put in one of two groups: The first group is provided gamification-based learning material (Group G) and the other group is provided text-based learning material (Group T). Participants in both groups complete all four parts of the study, where the second and fourth part differ between groups T and G (see below).

In the first part, participants have to complete an initial questionnaire. Question at this stage include age, the job category (administration, finance, human resources, education and research, facility management, workshop/repair and stock, student office), professional discipline (economics, business informatics, industrial engineering, mechanical engineering, electrical engineering, computer science, psychology, chemistry/materials, design, medical / pharmaceutical, media/communications, optics/acoustics, management, other [free text]). There are additional questions about how relevant subjects perceive phishing to be and how well they think they would handle phishing, as well as questions about learning preferences. In the second part, participants are directed to either the game-based learning game, or text-based learning material. We describe the material in more detail below. They have to complete working through this material, before being redirected to part three. In the third part, participants have to classify four previously unseen mails as "phishing" or "not phishing". Afterwards, in the fourth and final part, the participants are to complete the final questionnaire. In this last questionnaire, we ask questions about the learning method that they have just used. We ask about their motivation ("Gamified/text-based learning motivates me"), whether they think that they can apply what was learned in the future ("I will be able to use what I have just learned in the future"), and whether they would like to use the specific teaching method ("I would use text-based/game-based learning to educate myself in the area of IT-security").

2.2. Teaching Approaches in Comparison.

Participant were randomly assigned to one of two groups: One used game-based learning material (Group G) and the other used text-based learning material (Group T).

Group G. First the game-based task had a tutorial level, where the players got to know how the game-based learning task works. There were two emails, one said "That is a phishing email, please click the button for phishing". The other was "This is not a phishing email, please click the button for no phishing". After that they had to classify six emails whether it is phishing or not with a click on the buttons "no phishing" and "phishing". Also, there is a button that shows hints for each email. The rapid feedback is shown when the player had a wrong answer. For each email they can collect points for the right classification of phishing (wrong answers cost points, as does using the "hints" feature). At the end of the game there is an overview of how many points the player collected and what emails they identified correct or not. Also, additional information is given about how to recognize phishing.

In Figure 1 is an example of one of the phishing emails in the game-based-learning task. This figure is a screenshot of the game after the player classified a phishing email wrongly as no phishing. The red text is the rapid feedback/error analysis directly after classifying an email wrongly. There the player can see what signs in this particular mail indicate that an email is a phishing email by hovering her mouse over the red areas.

Group T. In the text-based-learning part the participants had to read a text on how about recognizing phishing emails. In total there were nine sections with a short explanation of what one has to be aware in each section (one to approximate six sentences). The sections were "Fake sender address", "Receiver Address Field", "Suspicious subject", "Generalized salutations", "Request for reply", "Attachments", "Spelling errors and grammar", "Psychological pressure" and "Hyperlinks".

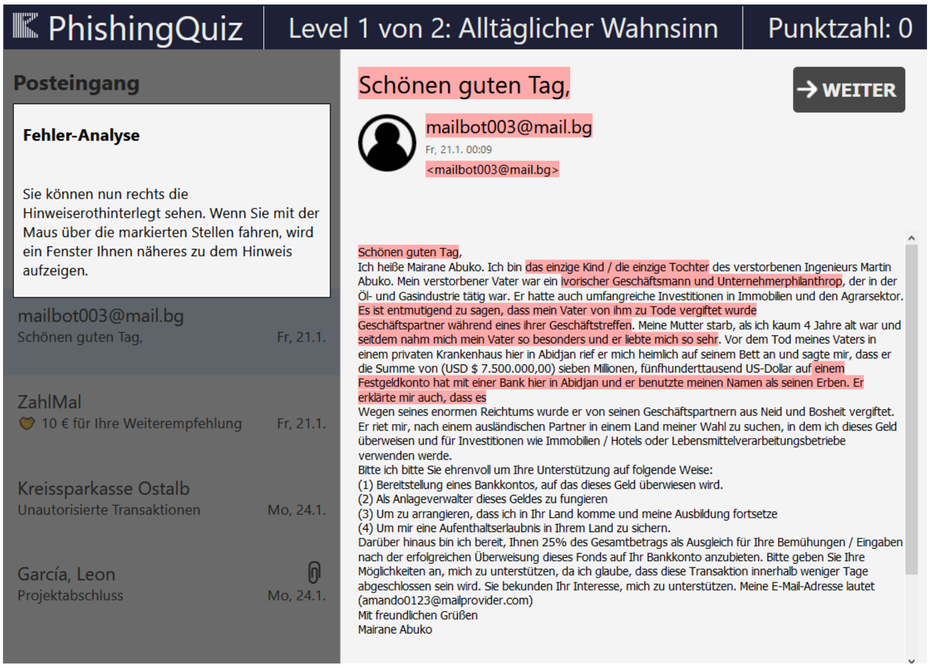

Fig. 1. Screenshot of email in the game-based learning task after false classification.

3 Outcomes

In this chapter we analyze the participants' preferences and self-assessment, compare group G and group T and technical and non-technical scientific disciplines.

3.1 Participants.

The sample comprised 90 participants, 45 participants in group G and 45 participants in group T. But in group G only 36 persons completed all tasks. We made no distinction between female and male participants. The participants did not have to answers all questions, therefore n<90 is possible. With a range of age from 20 to 71 and a mean age of 39.9 years, 95.6% of the participants were employees (response rate 19.1%) and 4.4 % were students (response rate 0,07%) of an institution in higher education. Of the sample 68.2% are in the discipline "teaching and research", the other participants with respectively 1-3% in other disciplines as Management Administration, Media designer, Human Resources, PR, Project Management, Secretary's office, Student department, Program Management, Student, Research Assistant, Third-party project, Finance Department, Graduate Campus, Information Technology, Central Student Service, IT and International Office[1]. As a result, too few samples exist within individual

---

[1] Recall that participants could choose to give a description of their field or area as free text, if none of the given options seemed suitable. This is why some answers were no listed as options in the above description of the questionnaire.

disciplines to draw statistically sound conclusions. To be able to work with sufficiently large sample sizes, we divide the participants into two groups: technical and non-technical disciplines. The technical group (n=48) includes participants from electrical engineering, computer science, Mechanical Engineering, Materials science, Mechatronics, Industrial Engineering. The non-technical group (n=42) includes participants from International Office, Business/Management, Chemistry/Materials, Design, E-Learning, History of art and culture, Media/Communication, Medicine/Health/Pharmacy, Optics/Acoustics, Psychology, Administration and Professional development.

3.2 Analysis of Participant Preferences and Self-assessment.

In part one of the questionnaire the participants (n=90) had to rate their knowledge in recognizing phishing mails (scale "very easy", "easy", "neutral", "difficult", "very difficult"). For 22% of the participants the self-assessment of recognizing phishing is "very easy", for 50% of the participants it is "easy". 24% assess their knowledge of recognizing phishing emails as "neutral" and only 3% as "difficult".
The participants (n=89) were asked what they like more: learning via gamified learning or traditional learn material (reading texts). 42 % of the participants prefer gamified learning, 13.5% prefer learning while reading texts (traditional learn material).

3.3 Comparison Group G and Group T.

Participant motivation. One of the questions for group G (n=35) was "Gamified learning motivate me..." (scale 1="yes to 5="no"). Group T (n=43) answered a corresponding question: "text-based learning motivate me..." with the same scale as group G. 80% of the participants in group G are motivated ("yes" and "tend to yes") in the context of gamified learning. In group T, 38% of the participants are motivated by text-based learning ("yes" and "tend to yes"). The mean value in Group G was 1,8 and 2,77 in group T (Welch's t-test: p=$1.59 \cdot 10^{-5}$, df=75,99, t=-4,61). In **Table 1** there is a comparison of the motivation of the both types of learning.

Table 1. Comparison of motivation gamified learning and text-based learning

Comparison of Motivation	**Gamified learning motivate me (n=35)**	**Text-based learning motivate me (n=43)**
Yes	43%	12%
Tend to yes	37%	26%
Neutral	17%	42%
Tend to no	3%	16%
No	0%	5%

Effectiveness. In the 3rd part, the participants had to classify four previously unseen mails into the categories “phishing” and “not phishing” and thereby apply what they had learned in part 2. The results are summarized in **Table 2.** Success of participants in respective groups. Statistical testing confirms the difference in means, but overall the difference is not large.

Table 2. Success of participants in respective groups. Welch's t-test: p=0.02536, t=-2.279, df=78,986.

	Group T (n=45)	Group G (n=36)
Mean number of correct answers (out of 4)	2,356	1,944
Correct answers in entire group	59%	51%

Satisfaction and Applicability. In answer to the two questions "I will be able to use what I just learned" and "I would use text-based/game-based learning to educate myself in the area of IT-security", we see a trend in favor of game-based learning. In **Table 3** we show how the participants of both groups assess their future-use of the learning-task made in part 2 of the survey. 51% of the participants from group G are sure that they will be able to use the knowledge they have just learned in the future. 23 % say that they "tend to yes", 20% of the participants see it neutral and the others (65%) do not think that they are able to use the knowledge in the future (n=35). The mean answer in group A was 1,829, a clear trend to "Yes". In group T, 26% of the participants in group T are sure that they will be able to use what they have just learned in the future and for 51% it is more likely that they are able to use the knowledge in the future (n=43). The mean is 2.116, also trending to yes, but less strongly so (Welch's t-test: p=0.222100, t=-1.231800, df=71.667). Note that we can only have limited confidence that the difference in means is a true reflection of differences in the underlying populations since p is relatively large.

We also asked participants whether they would be willing to use game-based or text-based learning in the future. The results are shown in **Table 3**. The answers again tend to favor game-based learning. The mean answer for group G was 1.686, whereas it was 2.6 for group T (Welch's t-test: p=$5.29 \cdot 10^{-5}$, t=-4.285400, df=75.884). This time we have higher confidence that the difference in means reflects a difference in means of the underlying population.

Table 3. Applicability of learned skills in future scenarios and satisfaction with learning approach. Numeric scale: Yes=1 to No=5.

I will be able to use what I have just learned in the future							
	Yes	**Tend to yes**	**Neutral**	**Tend to no**	**No**	**N**	**Mean**
Group G	51%	23%	20%	3%	3%	35	1.829
Group T	26%	51%	14%	5%	5%	43	2.116
I would use text-based/game-based learning to educate myself in the area of IT-security							
	Yes	**Tend to yes**	**Neutral**	**Tend to no**	**No**	**N**	**Mean**
Group G	51%	34%	9%	6%	0%	35	1.686
Group T	14%	35%	30%	19%	2%	43	2.605

3.4 Comparison between Technical and Non-technical Scientific Disciplines.

In group G 45,7% of the participants could be identify working in non-technical disciplines and 54,3% in technical disciplines (n=35). 49% of the participants in group T are working in non-technical disciplines and 51% of the participants are in technical disciplines (n=45). We compared the success of the two groups (technical vs. non-technical background) overall, that is regardless of learning method, and then per learning method. The results are shown in **Table 4**. A Welch's t-test shows that the results for the overall comparison (first row in **Table 4**) cannot be treated as evidence of a difference of means in the underlying population (p=0.889500, t=-0.139420, df=74.469). Since the two means are very close in the observed sample, this is not surprising. The differences in means for the technical and non-technical groups show a similar picture. The difference in means for the non-technical group is small and a Welch's t-test (p=0,819100, t=-0,230480, df=33,214) cannot confirm this difference to be indicative of a corresponding difference in the underlying populations. However, within the technical group, the difference in means is larger (approximately 0.364). This difference is more likely to reflect a difference in means between the two underlying populations (p=0.1834, t=-1.3544, df=38.9). We remark that also in this case p is relatively large, but small enough to suggest that a true difference in means cannot be ruled out.

Table 4. Mean number of correct answers in correlation with professional background.

	Technical		**Non-technical**	
Mean	2.195		2.222	
	Group G	***Group T***	***Group G***	***Group T***
Mean	2.0	2.364	2.188	2.25

4 Discussion

Our results show a statistically significant difference in the mean number of correct questions *in favor of text-based learning*. This is somewhat surprising, given that previous work established game-based learning to be superior. However, we must note that the results depend highly on the way in which effectiveness is measured, on the population that was studied, and on the specific type of game or gamified learning environment. In [8], results similar to our own were reported: Text-based training outperformed computer-based training (though the computer-based training worked somewhat differently than our own). Interestingly, this study relied on school students as participants for their study. A possible interpretation is that school students are particularly used to learning from text-based material – they are practiced learners. The same logic applies to the group of test-subjects we chose for our study: people working in higher education. Recall that 68.2% of our sample placed themselves into the "teaching and research" category. This group of people is likely used to digesting complex texts on a daily basis (be it in their role as teachers, researchers or students). Possibly, text-based learning is the optimal choice for this group as a whole.

Additionally, our results indicate that the difference between text-based learning and game-based learning is small to negligible within the non-technical population. However, the subjects with technical background actually performed worse given text-based learning to a statistically significant degree. The difference in success is in line with the corresponding difference observed in the overall group. This result further underscores that professional background or scientific discipline are an influential factor in the success of a given learning method. From the given results, we cannot explain why this particular group seems to perform better with text-based learning than with game-based learning. It is possible that this group benefits from texts especially well.

Or results further indicate that game-based learning motivates subjects to a higher degree than text-based learning. There is a significant difference in the mean score achieved in group G (game-based) compared with group T (text-based). This is in line with results from previous studies, such as [12]. Furthermore, participants report that they would use game-based formats for further education in the future, and that they feel that what they have learned can be applied in practice. These results suggest that game-based learning might be a prudent tool to initiate the learning process, even in groups that would perform better with text-based material. Put differently: Game-based learning might be a gateway to learning, where subjects would not otherwise initiate the learning process at all.

In summary, our results indicate that game-based learning is not generally superior in terms of effectiveness. However, there is some reason to believe it will generally lead to more people learning about IT-security than would if presented with text-based material only. Moreover, the effectiveness of game-based learning depends on the subject´s background and seems to be lower when the subject is from a technical discipline.

5 Conclusions

We conducted a study using subjects from an institution in higher education. Two groups of subjects were formed and trained to spot phishing mails with game-based material and text-based material respectively. Our results showed that the text-based group outperformed the game-based group. Moreover, this effect was observed when clustering subjects according to whether their professional field could be categorized as "technical" in nature, but was hardly observable when that field is of a non-technical kind. We believe these results underscore that a learner's background factors into the choice of the right learning tool. The fact that our subjects largely work in education and research, also underscores this point: They may simply be extremely adept at learning from texts. Additionally, we found that game-based learning is better at motivating subjects, and subjects who used game-based learning report a higher confidence that they will also use that form of learning in the future than those who used text-based material. We therefore believe that game-based learning might be better suited at initiating a learning process, particularly in contexts where learning about IT-security is voluntary and not part of some mandatory training.

For future work, the effects of sampling from an institution of education and research warrant further examination. A classification of what groups will likely benefit from game-based or text-based material could aid choosing the right tool for specific target audiences. Revisiting our observation that game-based learners report a higher probability that they will continue to learn using game-based material, we think there is merit to analyzing the emotional state of learners over time when presented with specific learning material (for similar studies, see e.g. [13, 14]) and determine how to best guide learners to a continuous learning routine.

References

1. Alkhalil, Z., Hewage, C., Nawaf, L., Khan, I. "Phishing Attacks: Recent Comprehensive Study and a New Anatomy.," Frontiers in Computer Science, 3 (2021)
2. Alami, D., Dalpiaz, F., „A gamified tutorial for learning about security requirements engineering," in EEE 25th International Requirements Engineering Conference (2017)
3. Canova, G., Volkamer, M., Bergmann, C., Borza, R. „NoPhish: an anti-phishing education app," in International Workshop on Security and Trust Management., Cham (2014)
4. Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., Volkamer, M. „An investigation of phishing awareness and education over time: When and how to best remind users.," in Symposium on Usable Privacy and Security (2020)
5. Canova, G., Volkamer, M., Bergmann, C., Reinheimer, B. „NoPhish app evaluation: lab and retention study.," in NDSS workshop on usable security (2015)
6. Karumbaiah, S. et al. „Phishing training: a preliminary look at the effects of different types of training.," Proceedings of the 11th Pre-ICIS Workshop on Information Security and Privacy (2016)
7. Scholefield, S., Shepherd, L.A. „Gamification techniques for raising cyber security awareness", International Conference on Human-Computer Interaction. Springer, Cham. (2019)
8. Stockhardt, S. et al. „Teaching Phishing-Security: Which Way is Best?," in ICT Systems Security and Privacy Protection. SEC 2016. (2016)
9. Perrault, E. K. „Using an interactive online quiz to recalibrate college students' attitudes and behavioral intentions about phishing.," Journal of Educational Computing Research 55.8, pp. 1154-1167 (2018)
10. Schreuders, Z.C., Butterfield, E. „Gamification for teaching and learning computer security in higher education.," 2016 {USENIX} Workshop on Advances in Security Education ({ASE} 16) (2016)
11. Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L.F., Hong, J., Nunge E. „Protecting people from phishing: the design and evaluation of an embedded training email system.," In: CHI, ACM, p. pp. 905–914 (2007)
12. Landers, R.N., Callan., R.C. „Casual social games as serious games: The psychology of gamification in undergraduate education and employee training.," in Serious games and edutainment applications, London, Springer, pp. 399-423 (2011)
13. Alsharnouby, M., Alaca, F., Chiasson S. „Why phishing still works: User strategies for combating phishing attacks.," International Journal of Human-Computer Studies, 82,, pp. 69-82 (2015)
14. Wanner, T., Wanner, T., Etzold V. „Effects on Girls' Emotions During Gamification Tasks with Male Priming in STEM Subjects via Eye Tracking," Smart Education and e-Learning 2020, pp. pp.67-78, June 2020 (2020)