International Journal of Human-Computer Studies 144 (2020) 102496

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# The effect of challenge-based gamification on learning: An experiment in the context of statistics education

Nikoletta-Zampeta Legaki[*,a,b], Nannan Xi[b], Juho Hamari[b], Kostas Karpouzis[a], Vassilios Assimakopoulos[a]

[a] School of Electrical and Computer Engineering, National Technical University of Athens, Zografou 15780, Greece
[b] Gamification Group, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere 33100, Finland

ARTICLE INFO

*Keywords:*
Gamification
Applications in education
Statistics education
Teaching forecasting
Human-Computer interface

ABSTRACT

Gamification is increasingly employed in learning environments as a way to increase student motivation and consequent learning outcomes. However, while the research on the effectiveness of gamification in the context of education has been growing, there are blind spots regarding which types of gamification may be suitable for different educational contexts. This study investigates the effects of the challenge-based gamification on learning in the area of statistics education. We developed a gamification approach, called *Horses for Courses*, which is composed of main game design patterns related to the challenge-based gamification; points, levels, challenges and a leaderboard. Having conducted a 2 (read: yes vs. no) x 2 (gamification: yes vs. no) between-subject experiment, we present a quantitative analysis of the performance of 365 students from two different academic majors: Electrical and Computer Engineering (n = 279), and Business Administration (n = 86). The results of our experiments show that the challenge-based gamification had a positive impact on student learning compared to traditional teaching methods (compared to having no treatment and treatment involving reading exercises). The effect was larger for females or for students at the School of Electrical and Computer Engineering.

## 1. Introduction

Gamification approaches are being applied with increasing frequency in an attempt to positively affect behavior and cognitive processes by enhancing the system or service with motivational affordances and eventually by bringing similar experiences as games do (Huotari and Hamari, 2017). Motivational affordances have been widely used in many fields such as business (Alcivar and Abad, 2016; Xi and Hamari, 2020), crowdsourcing (Morschheuser et al., 2017), healthcare (Johnson et al., 2016) and education (Dichev and Dicheva, 2017; Hanus and Fox, 2015; Koivisto and Hamari, 2019; Majuri et al., 2018; Osatuyi et al., 2018; Seaborn and Fels, 2015). Additionally, gamification has been employed in many education related contexts, across different educational levels (Caponetto et al., 2014; Dicheva et al., 2015; Simões et al., 2013; de Sousa Borges et al., 2014) and in various subjects (Dichev and Dicheva, 2017; Dicheva et al., 2015; Kasurinen and Knutas, 2018; Seaborn and Fels, 2015), showing its potential to improve learning outcomes (Koivisto and Hamari, 2019; Seaborn and Fels, 2015).

According to reviews of gamification literature, gamification has been employed mostly in the field of education (Koivisto and Hamari, 2019; Majuri et al., 2018; Seaborn and Fels, 2015). Gamified educational applications have been applied in non-academic areas as well: language teaching (Duolingo counts 300 million active users[1]) or software using (Ribbon Hero by Microsoft). Other popular gamified applications are: Kahoot and Quizizz, which can be easily configured and used in a variety of subjects, bringing game elements into classrooms without any special effort. Although gamification has an important position in education both inside and outside universities, there is still little effective guidance on how to combine different gamification features to enhance learning performance in different educational contexts (Hanus and Fox, 2015; Koivisto and Hamari, 2019; Seaborn and Fels, 2015).

Beyond research problems pertaining to the general interest in gamification and its effect on education, statistics education is an increasingly fundamental skill to understand the world around us. The lack of data literacy has been deemed one of the main causes behind our inability to act against climate change, to properly ratify means towards e.g. COVID-19 or generally as a hindrance for public understanding of science. Therefore, there is a need to make teaching methods in

* Corresponding author.
*E-mail addresses:* zampeta.legaki@tuni.fi, zabbeta@fsu.gr, nzabbeta@gmail.com (N.-Z. Legaki).

[1] https://ai.duolingo.com/

https://doi.org/10.1016/j.ijhcs.2020.102496
Received 12 November 2019; Received in revised form 9 June 2020; Accepted 12 June 2020
Available online 14 June 2020

statistics and forecasting more engaging (Love and Hildebrand, 2002). The acceleration of the daily data production, and ever-greater capacity to store and process this information, has boosted the necessity for students with a strong background in statistics and predictive analytics skills, in business environments or even in everyday life. Consequently, both statistics and forecasting techniques are of vital importance in the economics curriculum (Loomis and Cox Jr, 2003) and in other fields such as business (Makridakis et al., 2008) or social problems, where data may help to make better decisions. However, often forecasting courses are not even offered as an independent course in business schools (Hanke, 1989) and when they are, students are discouraged to participate in the courses because they find the topic too complicated (Albritton and McMullen, 2006; Gardner, 2008; Snider and Eliasson, 2013; Torres et al., 2018) and demanding (Craighead, 2004). Therefore, regarding the education and especially the education in the field of statistics and forecasting, student motivation is crucial for their participation and understanding in order to reach their learning potential, meet business needs and get insights of the data to support the decision-making process.

Despite the long tradition in educational business games, thus far there have only been a few studies on gamification or simple gamified exercises, combined with traditional teaching methods in the area of statistical forecasting. These studies have mostly used: score (Craighead, 2004), spreadsheets (Gardner, 2008), competition (Snider and Eliasson, 2013) and real-world forecasting problems (Buckley and Doyle, 2016a; Gavirneni, 2008) in order to encourage students' participation, without examining gamification effects. Other more quantitative studies have used forecasting in the context of a prediction market as a tool to motivate students rather than to teach forecasting aspects (Buckley and Doyle, 2016a; 2016b; 2017; Buckley et al., 2011). While there are several types of games and gamification designs, the challenge-based gamification (e.g points, levels, leaderboard, clear goals/ tasks), as opposed to the immersion- and the social-based gamification, has been suggested (Dicheva et al., 2015; de Sousa Borges et al., 2014; Zichermann and Cunningham, 2011) and applied to a high degree in practice as gamification design in education (Koivisto and Hamari, 2019; Seaborn and Fels, 2015; de Sousa Borges et al., 2014). One or more of these gamification elements have been used with promising results even in educational topics relative to forecasting (Craighead, 2004; Gavirneni, 2008; Gel et al., 2014; Snider and Eliasson, 2013). However, there is still a lack of effective design guides and empirical data on the combination or integration of these features in the context of educational information systems (Koivisto and Hamari, 2019). Challenge-based gamification introduces a design approach of integrating achievement gamification features, positively related with intrinsic need satisfaction (Xi and Hamari, 2019), in an educational service or application, in order to explore its potential, motivate users and eventually improve learning.

The present study examines the impact of three treatments on students' performance i) reading, ii) use of a challenge-based gamified application, iii) the combination of the two. In order to do that, we consider a variety of student characteristics such as gender, level of studies, academic major, expertise in the English language, and use of personal computers and games. We designed and implemented a web-based gamified application, called *Horses for Courses* and we conducted a series of experiments over the last 4 years. The total sample is composed of 365 students, with 279 undergraduate and MBA students at the School of Electrical and Computer Engineering of the National Technical University of Athens, Greece (hereafter ECE, NTUA) and 86 undergraduate students at the Business Administration Department in the School of Business and Economics of the University of Thessaly, Greece (hereafter Business Administration). Our findings show that challenge-based gamification improves students' learning outcomes on a statistics course, contributing to the knowledge of challenge-based gamification's effect on statistics/stem education and eventually on gamified pedagogy.

## 2. Background

### 2.1. Gamification in education

Gamification refers to a method of designing systems, services, organizations and activities in order to create similar experiences and motivations to those experienced when playing games, with the added educational goal of affecting user behavior (Huotari and Hamari, 2017). Games are known to motivate and engage players (Dichev and Dicheva, 2017) because of the enjoyment and the excitement that this activity offers (Koivisto and Hamari, 2019). In this regard, gamification aspires to create this experience in different contexts. This is usually attempted by using game mechanics or other game-like designs in the target environment (Deterding et al., 2011). Over the last decade, gamification research has affected a variety of domains that deal with education (Koivisto and Hamari, 2019). The educational domain is continuously evolving, incorporating the latest developments in information technology even in elementary schools (Karpouzis et al., 2007). Nonetheless still demands students' commitment and persistence in order for them to gain in-depth knowledge. Consequently, gamification has been of great interest to educators who have been exploring its potential in improving student learning (Dichev and Dicheva, 2017; Dicheva et al., 2015; Hamari, 2013; Koivisto and Hamari, 2019; Majuri et al., 2018; Seaborn and Fels, 2015). This potential has led to a growing literature on the effectiveness of gamification, mainly in universities but also in other academic contexts (Caponetto et al., 2014; Koivisto and Hamari, 2019; Seaborn and Fels, 2015; de Sousa Borges et al., 2014) and in a variety of subjects (Dichev and Dicheva, 2017; Kasurinen and Knutas, 2018). To name a few: information technology (Osatuyi et al., 2018), math/ science (Attali and Arieli-Attali, 2015; Christy and Fox, 2014) and taxation (Buckley and Doyle, 2016a; 2017).

Gamification types and types of game design have commonly been horizontally categorized into main three overarching categories of achievement/challenge-, immersion-, and social-based (Hamari and Tuunanen, 2014; Koivisto and Hamari, 2019; Snodgrass et al., 2013; Xi and Hamari, 2019; Yee, 2006; Yee et al., 2012) beyond the common vertical categorization of e.g. the MDA model that separated game design into mechanics, dynamics and aesthetics (Hunicke et al., 2004). The immersion-based game design attempts primarily to engulf the player or user into a story, roleplay and audiovisual richness. The social-based game design is commonly focused on different forms of competition and collaboration. Finally, the achievement/challenge-based game design is focused on overcoming challenges, progressing and earning rewards and feeling competent. Within the achievement/ challenge-based gamification, the most commonly embodied mechanics have been points, challenges, leaderboards, levels and badges (Koivisto and Hamari, 2019; Majuri et al., 2018; Pedreira et al., 2015). According to the self-determination theory, the use of these elements, which are considered as achievement related features and immediate performance indicators, is associated with intrinsic motivation for students (Xi and Hamari, 2019). In this regard, these elements form challenge-based gamification underpinnings in order to motivate students to maximize their knowledge acquisition.

Review studies about the effectiveness of gamification are generally optimistic, mainly listing either positive or mixed results of applied gamified strategies (Buckley and Doyle, 2017; Caponetto et al., 2014; Dicheva et al., 2015; Koivisto and Hamari, 2019; Lambruschini and Pizarro, 2015; Majuri et al., 2018; Nah et al., 2014; Osatuyi et al., 2018; Reiners et al., 2012; Seaborn and Fels, 2015; de Sousa Borges et al., 2014). Nevertheless, they mention the need for more controlled experimental research on the impact of gamification, independently of the application domain or used gamified strategy (Buckley and Doyle, 2017; Caponetto et al., 2014; Dichev and Dicheva, 2017; Dicheva et al., 2015; Hanus and Fox, 2015; Koivisto and Hamari, 2019; Lambruschini and Pizarro, 2015; Landers et al., 2018; Majuri et al., 2018; Nah et al., 2014; Osatuyi et al., 2018; Reiners et al., 2012; Seaborn and Fels, 2015; de Sousa Borges et al., 2014).

The effects of gamification are bound together with the target audience and the context (Buckley and Doyle, 2017; Dichev and Dicheva, 2017; Hanus and Fox, 2015; Koivisto and Hamari, 2019; Seaborn and Fels, 2015). Hence, the results of gamification vary regarding the subject and the field of application (Hanus and Fox, 2015; Sánchez-Martín et al., 2017). Therefore, researchers generally agree on the need for stronger empirical results (Buckley and Doyle, 2017; Hanus and Fox, 2015; Koivisto and Hamari, 2019; Landers et al., 2018; Maican et al., 2016; Morschheuser et al., 2017). This study contributes to this body of research with empirical data drawn from a series of experiments on the impact of gamification with control and treatment groups in the context of a forecasting course.

### 2.2. Teaching forecasting in higher education

As described in Garfield and Ben-Zvi (2007), statistics and statistical literacy are of paramount importance, especially in the rapidly changing business environments. As interest in the available technology and statistics is growing, forecasting skills are becoming more sophisticated (Kros and Rowe, 2016) and the process of teaching forecasting is becoming more difficult and demanding. Statistics courses focus on data analysis (Cobb, 1992), as competitive business environments require graduate students to interpret data and be able to use statistical and judgmental forecasting methods and applications (Giullian et al., 2000; Kros and Rowe, 2016). The importance of forecasting skills is not a new discovery (Albritton and McMullen, 2006; Craighead, 2004; Giullian et al., 2000; Kros and Rowe, 2016; Loomis and Cox Jr, 2003; Makridakis et al., 2008; Snider and Eliasson, 2013). However, recently these skills have have become even more important since business decision-making must be supported by data-based evidence and projections (Giullian et al., 2000). Another aspect of forecasting that highlights its importance is its multidisciplinary nature, since the forecasting techniques are an essential component in a number of fields such as business statistics (Tabatabai and Gamble, 1997), supply chain management (Gavirneni, 2008) and management science (Makridakis et al., 2008).

However, the eagerness of the business sector to equip students with a strong background in forecasting techniques is only partially reflected in the education that universities and business schools provide. Thirty-five years ago, 58% of the surveyed universities offered an independent forecasting course (Hanke, 1984). The percentage is reduced to 34.48% of the surveyed business schools based on a more recent study by Kros and Rowe (2016) and it is almost the same (50%) regarding the top 50 US Business Programs, which requires a forecasting time-series course. Moreover, there is a variety of “e-learning-in-statistics initiatives”, but even these modules do not focus on time-series and forecasting methodologies (Gel et al., 2014).

Despite the growing popularity of and need for forecasting skills, business schools slowly address this demand, and they generally disregard the need for increasing student motivation (Debnath et al., 2007). Business forecasting or statistical forecasting methods are usually considered complicated (Albritton and McMullen, 2006; Craighead, 2004; Gardner, 2008; Snider and Eliasson, 2013; Torres et al., 2018), making it difficult for students to remain motivated (Craighead, 2004). Taking this into account, Chu (2007); Donihue (1995); Loomis and Cox (2000); Loomis and Cox Jr (2003); McEwen (1994) suggest alternative teaching guidelines such as the use of a software or new technology in combination with real data and forecasting problems. Active learning has also been proposed (Love and Hildebrand, 2002) in order to address the need to update the forecasting educational process. So the digitalization, which we experience has boosted the statistical skills’ importance. However, universities and business schools have not responded immediately to this challenge and they have been criticized for not placing enough focus on the specific skills that will improve the students’ future job performance (McEwen, 1994) and career success (Pfeffer and Fong, 2002).

### 2.3. Gamification and teaching statistical forecasting

This study puts emphasis on gamification only in terms of simple educational activities or systems, usually including game mechanics (Bunchball, 2010; Deterding et al., 2011). In this direction, we reviewed journal articles that discuss simple active learning events, gamified exercises or games in the context of a forecasting course. Our results indicate that the use of score (Craighead, 2004), spreadsheets (Gardner, 2008) and competition (Snider and Eliasson, 2013) during lectures has positive effects regarding students’ attitude, but strong empirical data is not presented. The use of a customized software (Spithourakis et al., 2015) and students’ participation in prediction of a basketball score appeared beneficial in the context of an undergraduate forecasting techniques course.

Other active learning exercises have used competition based on students’ forecasting accuracy in order to increase students’ participation and improve learning outcomes in a management course (Buckley et al., 2011) and in a taxation course (Buckley and Doyle, 2016a). Another simple game, named: *FREDCAST* has been designed and recently used in order to teach forecasting in a macroeconomic course (Mendez-Carbajo, 2018). All of these examples, show that forecasting due to its nature can be considered as a kind of an artistic field (Gavirneni, 2008), where gamification could be efficiently integrated in order to make it attractive to its audience.

Our preliminary review mainly positions gamification as a beneficial tool in education of forecasting and related fields such as management (Buckley et al., 2011; Makridakis et al., 2008), decision-making (Makridakis et al., 2008), taxation (Buckley and Doyle, 2016b), supply chain management (Gavirneni, 2008) but highlights the need for more empirical results. Additionally, an overview of teaching forecasting shows the importance of forecasting courses in economics syllabus (Loomis and Cox Jr, 2003) or business school curriculum (Buckley et al., 2011; Gavirneni, 2008). The need for forecasting skills is increasing as well, due to technological changes, as management seeks data-based approaches in dealing with decision-making on market opportunities, environmental factors and technological resources. However, forecasting courses are not adequately supported by students' participation or university and business school programming (Albritton and McMullen, 2006; Snider and Eliasson, 2013). A possible approach to address this issue could be gamification, which under proper design guidelines has produced promising results regarding student motivation and learning outcomes in management courses (Buckley and Doyle, 2016a; Craighead, 2004; Gardner, 2008; Snider and Eliasson, 2013) and in a forecasting module (Gavirneni, 2008). Nevertheless, thus far, there have not been a lot of studies on the effects of gamification on learning outcomes, in the specific area of statistical forecasting. Therefore, this study experimentally examines the potential of the challenge-based gamification, by designing from scratch and using a gamified application in order to improve student learning in a forecasting course.

## 3. Material and methods

### 3.1. Participants

A series of experiments were conducted at the ECE, NTUA and at the Business Administration. More precisely, we performed the experiments in different classes and academic majors, as follows:

- 49 undergraduate students (class of 2015) at the ECE, NTUA.
- 37 MBA students (class of 2015) at the ECE, NTUA.

- 60 undergraduate students (class of 2016) at the ECE, NTUA.
- 52 undergraduate students (class of 2018) at the ECE, NTUA.
- 21 MBA students (class of 2018) at the ECE, NTUA.
- 86 undergraduate students (class of 2018) at the Business Administration.
- 60 undergraduate students (class of 2019) at the ECE, NTUA.

The total sample is composed of 365 students; 270 students are males and 95 are females. The experiments were performed in the context of a forecasting course, with fourth-year undergraduate students and second-year MBA students, respectively, at the ECE, NTUA. At the Business Administration, the experiment was conducted in the context of an information technology course, with first-year students. However, the Business Administration's curriculum contains an operational research course, which includes forecasting techniques. The experimental design of our experiments was followed strictly, independently of the academic level of the students, as described in the next section.

### 3.2. Experimental design

We conducted a 2 (read: yes vs. no) x 2 (gamification: yes vs no) factorial experiment. The dependent variable was student performance in the learning task. Participants were randomly assigned to one of the conditions of the experiment: i) Group Control: no treatment, ii) Group Read: treatment of reading a research paper, named thenceforth as task Read (see 3.3.3), iii) Group Play: treatment of using challenge-based gamification, named thenceforth as task Play (see 3.3.4) and iv) Group Read&Play: both tasks: "Read" and "Play". Time was controlled and was equal to 15 minutes for each task. Table 1, depicts the design of the evaluation of our experiment and all the treatments are explained in the following section.

### 3.3. Materials

#### 3.3.1. Lecture

The learning objectives of this lecture were for students to understand and be able to apply the "Method Selection Protocols" for regular/fast-moving and intermittent demand time-series based on specific academic work of Petropoulos et al. (2014). During the lecture, the aim of the Petropoulos et al. (2014) research was mentioned, along with the data, and the research methodology used. Special attention was paid on the results, the practical implications and the conclusions of this study about the "Method Selection Protocols". More precisely, we further explained the relation between the time-series features and one strategic decision and the forecasting accuracy of the proposed methods for both regular/fast-moving and intermittent demand time-series. The visual material of the lecture was composed of 17 slides and lasted 15 minutes. The lecture content was focused on specialized knowledge of the research of Petropoulos et al. (2014) that both undergraduate and MBA students in any of the stages of their studies would not have otherwise been taught.

#### 3.3.2. Final evaluation form

An evaluation form at the end was used to measure students' learning performance via 30 close-ended questions (i.e. questions where the participants would select the right answer among possible answers) of equivalent grade about the findings of Petropoulos et al. (2014). This was the last task for all participants, independently of the group to which they were assigned. The answers to all the questions were covered in the lecture material. All the questions were about topics that have been discussed in the lecture described at subsection 3.3.1, and therefore, in principle it would be possible to attain the highest score by only participating in the lecture.

#### 3.3.3. Task read

The material of the reading task was: "Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'horses for courses' in demand forecasting. European Journal of Operational Research 237, 152 -163.". The paper is 12 pages and is the foundational material of the lecture content. The students read the article using a computer in the computer lab.

#### 3.3.4. Task play: Challenge-based gamification

Since there is a lack of free, computationally non-complex gamified applications, specifically created to teach statistical forecasting, we developed a gamification approach called *Horses for Courses*. It is a simple gamified application, which is composed of main design patterns related to challenge-based gamification, as described in Section 3.3.4.2. *Horses for Courses* aims to motivate students' participation, to improve their learning outcomes regarding the choice of simple but accurate statistical forecasting methods, and consequently enhance their forecasting skills. It is structured based on the above-mentioned foundational material: "Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'horses for courses' in demand forecasting. European Journal of Operational Research 237, 152 -163.".

*3.3.4.1 Horses for Courses Architecture*

In order to implement *Horses for Courses*, we considered the methods and design principles of both gamification (Morschheuser et al., 2018; Zichermann and Cunningham, 2011) and software development (Barnett et al., 2005; Gallaugher and Ramanathan, 1996; Lewandowski, 1998). As far as the architecture of the application is concerned, a focus on flexibility, accessibility, high-level programming and the ability to be integrated in different platforms led us to build a web application on the Microsoft.NET framework (Barnett et al., 2005) and to use an MS-SQL database. *Horses for Courses* is a web-based gamified application, structured as a three-tier system with simple application logic layer, which is fully accessible to registered users via a browser. Users register with an email and a password in order to save their progress in a database scheme, which serves as a data tier. A "Data Module" is used to retrieve data from the database and to build it into functional objects. A class named "Actions", enables the interaction between users and the "Data Module" in order to save updated data and the user's progress back into the database, composing the logic tier of the application. The presentation tier consists of a graphical user interface, including time-series, data visualization and system functions. A compact graphical representation of the *Horses for Courses* application is found in Fig. 1.

*3.3.4.2 Horses for Courses Design*

Guidelines for the design of the challenge-based gamification and consequently of *Horses for Courses* application were divided into two main directions: (1) the effective use of motivational affordances in

**Table 1**
Design of the evaluation of the experiments.

| Task Description | Group Control | Group Read | Group Play | Group Read&Play |
|---|---|---|---|---|
| Attend Lecture (see 3.3.1) | ✓ | ✓ | ✓ | ✓ |
| Read the Paper (see 3.3.3) | | ✓ | | ✓ |
| Play (see 3.3.4) | | | ✓ | ✓ |
| Evaluation Form (see 3.3.2) | ✓ | ✓ | ✓ | ✓ |

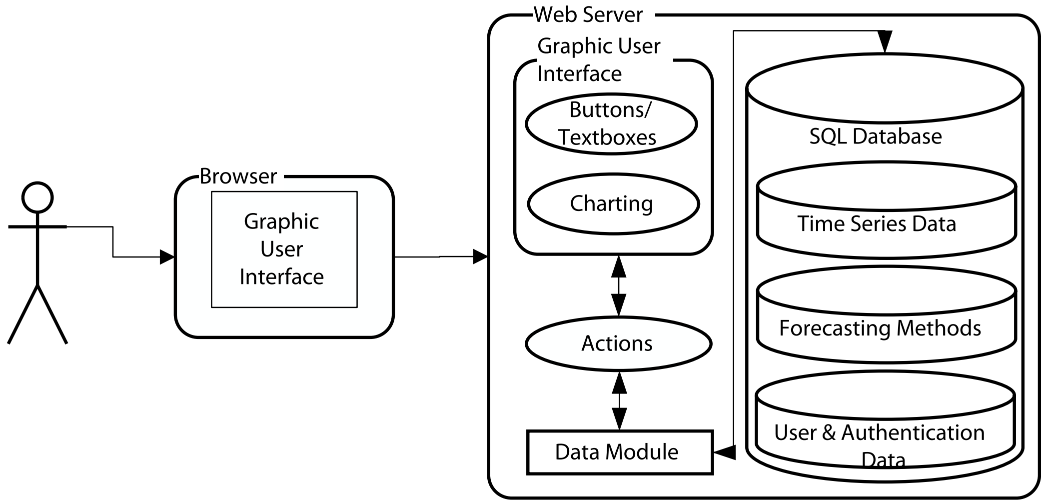


**Fig. 1.** *Horses for Courses* architecture.

**Table 2**
Integrated motivational affordances in *Horses for Courses* application.

| Affordance | Definition | Purpose in *Horses for Courses* |
|---|---|---|
| Points | Numeric measure of players' performances. | Reward for the correct application of method selection protocol. |
| Levels | Difficulty moderated based on players' expertise. | Indicator of progression and difficulty. |
| Challenges | Predefined quests and increasingly more difficult objectives. | Positive impetus to keep players engaged to maximize their points. |
| Leaderboard | Direct comparison of players' performance. | Increase of competition among students. |

Source: (Buckley and Doyle, 2017; Bunchball, 2010; Kapp, 2013; Maican et al., 2016; Nah et al., 2014; Seaborn and Fels, 2015; Zichermann and Cunningham, 2011).

learning environments (Deterding et al., 2011; Dicheva et al., 2015; DomíNguez et al., 2013; González and Area, 2013; Hanus and Fox, 2015; Maican et al., 2016; Nah et al., 2014; Pedreira et al., 2015; da Rocha Seixas et al., 2016; Yildirim, 2017) and (2) the design and development of gamified applications (Kapp, 2013; Morschheuser et al., 2018; Zichermann and Cunningham, 2011). The most frequently used and assessed motivational affordances in education and in general, so far are: points, levels, badges/achievements and leaderboards (Alhammad and Moreno, 2018; Koivisto and Hamari, 2019; Majuri et al., 2018; Pedreira et al., 2015). Thus, we incorporated points, a level setting, challenges and a leaderboard into our application in order to maximize the external validity of the experiment - i.e. to mimic a possible real world implementation of this gamification style. More precisely, Table 2 describes the motivational affordances in *Horses for Courses*, along with their definitions from the literature and the purpose they serve.

Apart from these motivational affordances, our design decisions regarding *Horses for Courses* were also determined by a desire to create a user-friendly and agile interface and work flow, with clear player guidance and instructions (Kapp, 2013). Fig. 2 describes a full round of the game. Initially, students had to register or sign in. Later, for each level they had to select the most suitable forecasting method based on the provided data and information, as it is depicted in Fig. 3. Then participants win points according to their choices. Instructions are easily accessible as well as the"Method Selection Protocols", through colorful buttons as Fig. 3 illustrates, as well. New challenges arise at each level, for example to identify time-series components for the real data time-series as it is depicted in Fig. 4, encouraging the students to assess their knowledge and win more points. The aim is to achieve a high ranking on the final leaderboard, based on the collected points.

### 3.4. Procedure

The experiments took place as a replacement of a normal lecture of the respective course. Therefore, participants (students) arrived at the lecture as normal at the designated computer lab at the standard time. After arriving at the class, the participants were informed about the experiment and their informed consent was obtained. At ECE, NTUA, participation was voluntary, however, the incentive for participation was a bonus of 0.5/10 in the course's final grade, instead of an equivalent exercise in the final examination. In such manner, every student, participating or not in the experiment, could receive the highest grade in the final examination. The participation of undergraduate students at Business Administration was mandatory as part of the course and there were no additional incentive for them to take part in the experiment.

Participants were instructed that they should pick a computer station at the classroom (a computer lab), attend a lecture and then complete an evaluation form, which was based on the content of the lecture. They were informed that their performance in the evaluation form would not affect their course grade, however, they should try to correctly answer the questions based on their understanding of the topic described in the lecture. The experimenter mentioned that the participants would be randomly divided into four groups, without further information, but the importance of the evaluation form along with the time constraints for all groups was highlighted.

First part of the experiment procedure was a 15-minute lecture about the findings of Petropoulos et al. (2014) research (see 3.3.1). After the lecture, the participants were randomly assigned to different conditions of the experiment: Group Control, Group Read, Group Play and Group Read&Play. All students, independently of their group received the same incentive in order to eliminate the recruitment bias. Then, the experimenter informed the participants about their next task and the available time. Instructions were given for each group respectively. Each group had 15 minutes to complete the respective task. As described above, Group Control did not have an extra task, Group Read had 15 minutes to read the paper (task Read), Group Play had 15 minutes to fulfill a full round in the gamified application (task Play). Group Read&Play had 30 (15 + 15) minutes to fulfill the task Read and then the task Play. Finally, all groups had to complete the evaluation

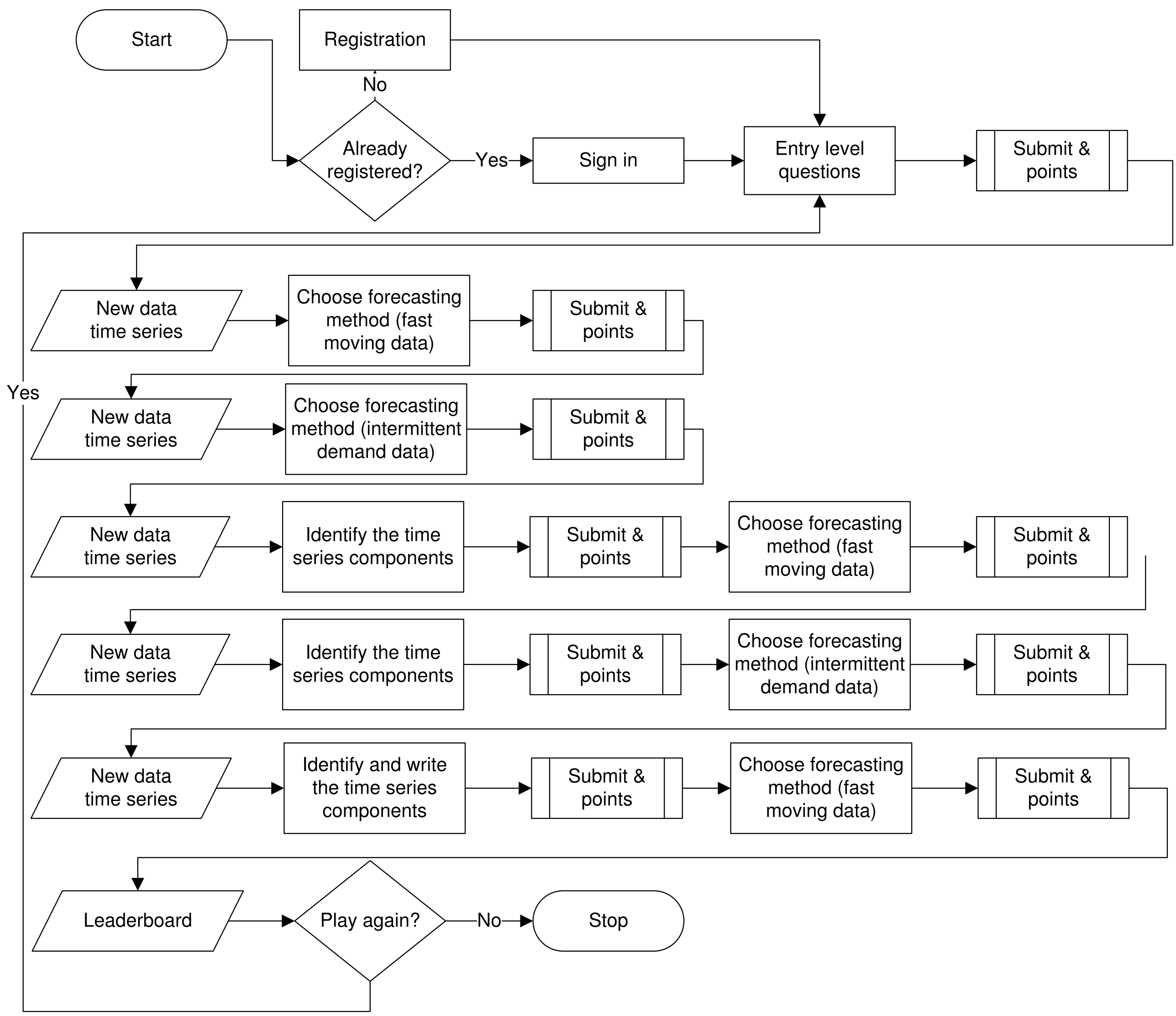


**Fig. 2.** *Horses for Courses*' flowchart of a full game round.

form within 15 minutes, which measured their performance (see Materials 3.3.2). Participants completed all of the tasks' stages independently and were not allowed to communicate with other participants. The procedure was the same at both participating campuses, including the materials and the experimenter.

## 4. Results

The objective of this study is to identify the impact of the challenge-based gamification on students' performances in the evaluation form and consequently on their comprehension. Thus, we examine the relative performance of a control group in comparison with that of the treatment groups. Students' performances on the evaluation form were calculated as the sum of the right answers on the questionnaire, normalized to a maximum of 100. Summary statistics of results are presented in Table 3 for each treatment along with the students' gender, their academic major: ECE, NTUA or Business Administration and their educational level: Undergraduate or MBA. The distribution of the performances is illustrated in percentiles with box-plot diagrams in Fig. 5, which are further separated into different treatments, including gender, major and educational level. Finally, for a subset of the sample (N = 146), three extra variables have been examined along with the treatment groups, namely students' expertise in English, and use of personal computers and games. Students' responses were ranging from 1 = beginner to 5 = proficient. *Horses for Courses* is a web-based gamified application which demands the use of a personal computer and the language of its interface is English. Therefore, we examined the relationship between these variables and students' performances by testing if these variables would be statistically significant in students' performances in conjunction with the treatment received.

The analysis of the results was conducted in three steps. First, we investigated the mean values of the students' performances and their statistically significant differences with respect to specific treatments. Table 3 presents the number of students per treatment, the mean value of the students' performances, and the standard deviation in each treatment group, regarding their gender, academic major and educational level. Overall, the groups that experienced the challenge-based gamification achieved greater mean values of performances than the other groups. More precisely, Group Read&Play, which read the respective paper and used the gamified application, reached the highest mean performance of 58.05 out of 100, and had the second lower level of standard deviation in results (SD = 17.00). Group Play, which only experienced the gamified application, had the second highest mean performance of 52.55 out of 100 and the highest level of standard deviation in results (SD = 19.74). Group Read had a lower mean performance of 46.13 out of 100 (SD = 18.68). Finally, Group Control had the

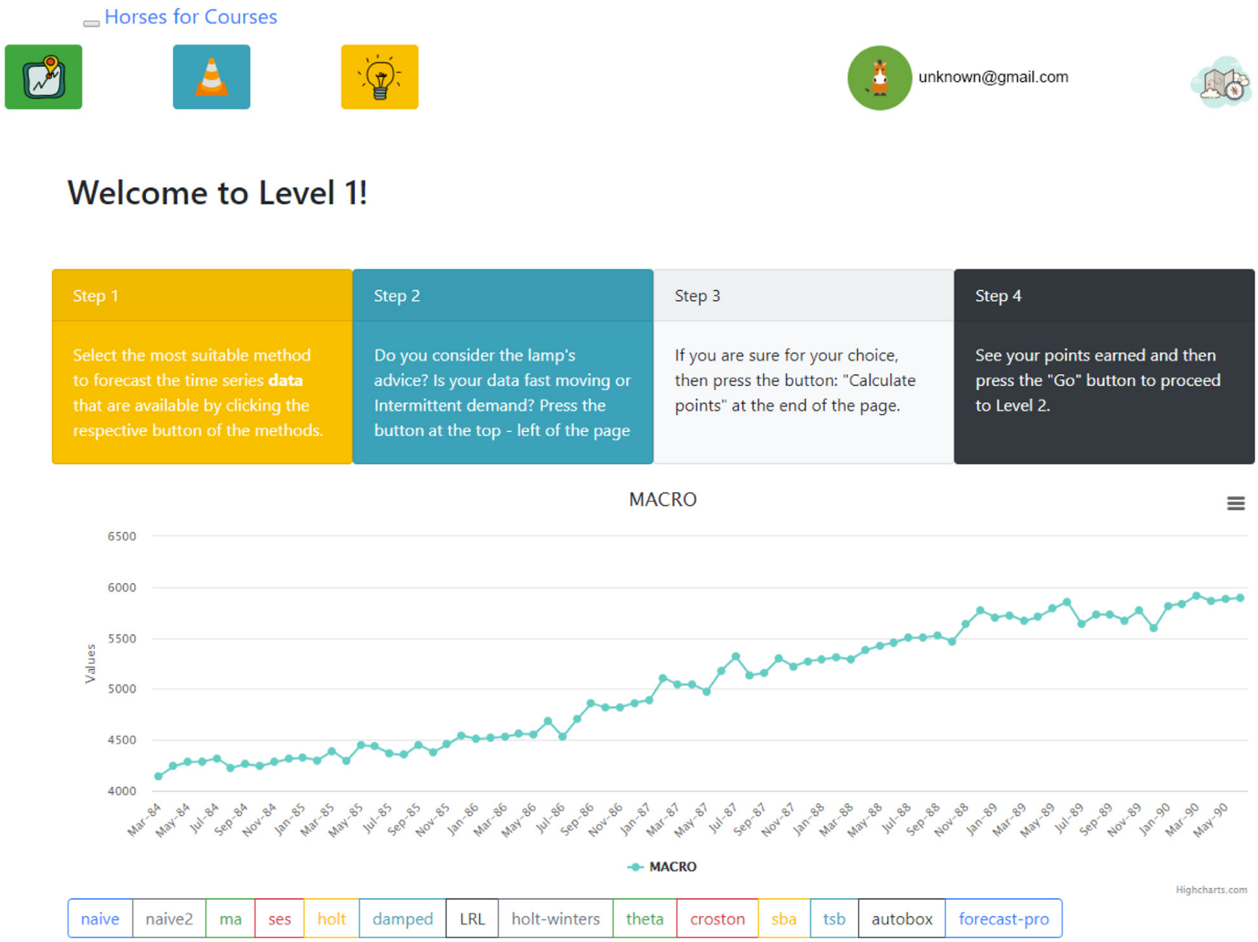


**Fig. 3.** View of 1st level challenges of *Horses for Courses* application.

lowest mean performance, 36.13 out of 100, but also the lowest standard deviation (SD = 11.57).

Based on the Shapiro-Wilk test on the ANOVA residuals, the assumption of normality was violated. In order to study the significant differences in the average values of all the groups' performances, we ran the non-parametric Kruskal-Wallis rank sum test (Kruskal and Wallis, 1952). The null hypothesis of equal differences is rejected (chi-squared = 70.842, df = 3, $p < 0.001$) and we can therefore establish significant differences between the groups. Furthermore, we conducted pairwise multiple comparisons without making assumptions about normality, using the Dunn procedure (Dinno, 2015; Dunn, 1961; Zar et al., 1999), with a confidence interval equal to 95%. Kruskal-Wallis with Dunn's post-test was chosen to test the significant differences because data was not normally distributed in all cases. Table 4 presents the outcomes concluding that all treatment groups resulted in significantly higher performances compared to Group Control (p. adj < 0.001, Kruskal-Wallis with Dunn's post test). Additionally, Table 4 displays the respective effect size of these treatments compared to Group Control, based on non-parametric Cliff's Delta estimator (Cliff, 2014; Macbeth et al., 2011; Wilcox, 2006). The only pairwise comparison without statistically significant differences in students' performances is Group Play versus Group Read&Play. Finally, Group Read&Play outperformed all the other groups. This group noted the highest improvement regarding the mean values of performances of Group Control equal to 60.67%. Group Play and Group Read follow with improvement equal to 45.45% and 27.68% respectively.

The performances of all groups were compared directly, focusing on the assessment of questions in the evaluation form, despite that treatments did not have the same duration. In order to deal with this limitation of our study, we used independent binary variables for the tasks Read and Play respectively. The value of the variable Read is equal to 1 if the respective group completed the task, and 0 otherwise. The same applies for the variable Play. Then, the Scheirer-Ray-Hare test was performed (Scheirer et al., 1976), using students' performances and these variables. Results show that each of the tasks: Read the research ($H = 16.014$, $p < 0.001$) or Play with *Horses for Courses* application ($H = 52.81$, $p < 0.001$) had a significant impact on students' performances, but their interaction was not significant ($H = 2.019$, $p = 0.156$).

Along with the impact of different treatments on students' performances, the independent variables gender, academic major and educational level were examined using the Scheirer-Ray-Hare test. While the students' gender and their academic major appeared to have a great impact on their performances, based on the results presented in Table 5, the interactions between them and the respective treatments they underwent did not have significant impact. The students' educational level was not an important variable, nor was its interaction with the treatments.

Regarding the impact of additional variables on a subset of our sample, only the impact of the students' expertise in English resulted in statistically significant differences in students' performances. The last rows of Table 5 show the results of the Scheirer-Ray-Hare test for the respective samples. Results about the mean values and the standard deviations of these extra variables are presented in Table 6. However, these variables will not be further analyzed because students reported their answers only in the recent experiments.

Table 7 demonstrates the impact of different treatments on students' performances, regarding the statistically significant variables of gender and academic major. We calculated the improvement of more specified

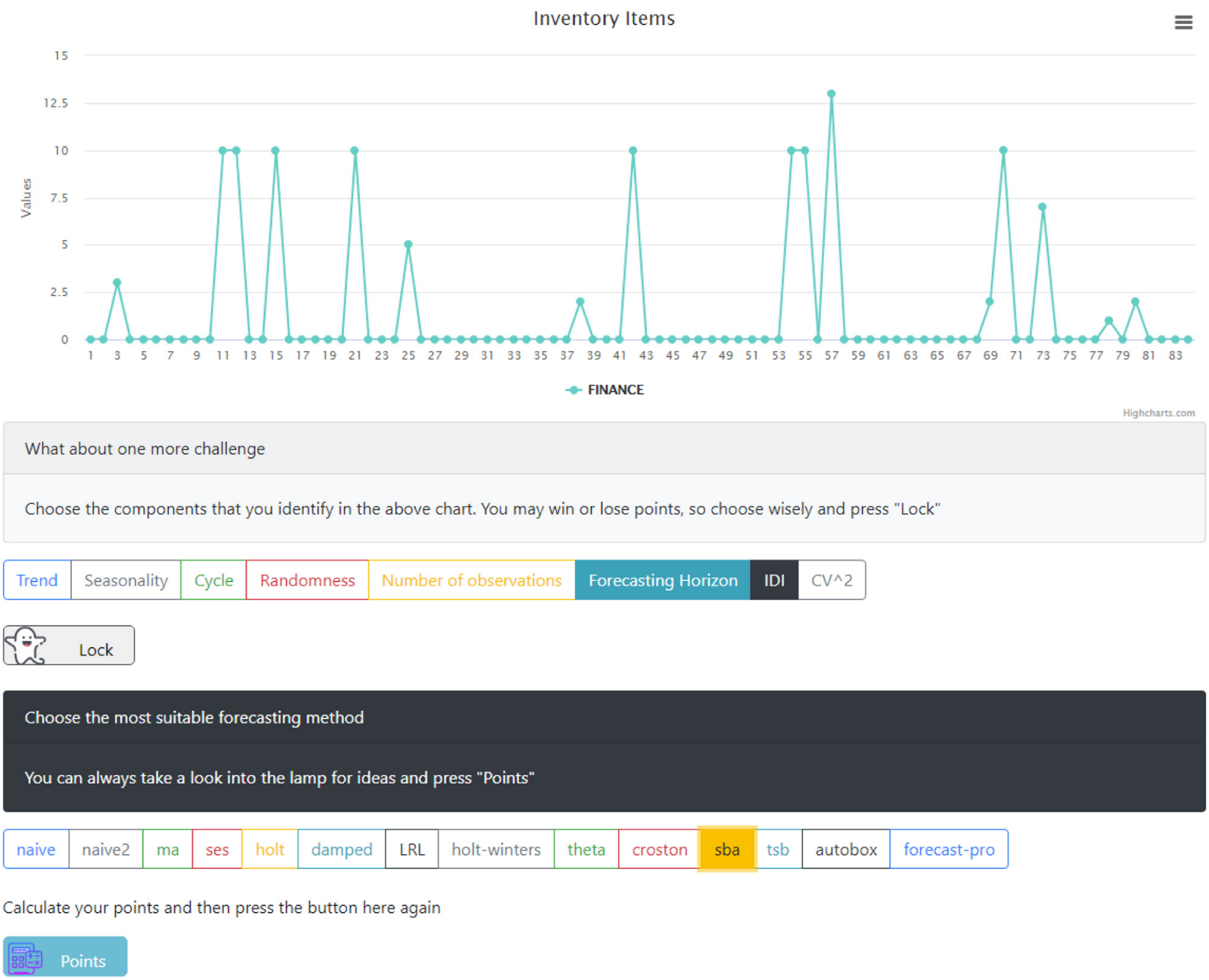


**Fig. 4.** View of $4^{th}$ level of *Horses for Courses* application.

groups regarding these variables compared to the mean value of the students' performances of Group Control (equal to 36.13) as a benchmark value. Students at the ECE, NTUA have noted the highest improvement regarding the mean values of their performances in the evaluation form. Furthermore, female students, independently of their major, who had used the gamified application benefited more from gamification; their mean performances are higher from those of the respective groups composed of male participants. These findings do not apply in non-gamified groups.

To conclude, we divided all the data of students' performances into two larger groups, instead of four: the non-gamified group, composed of 165 students (M = 41.04, SD = 16.22), who did not use the *Horses for Courses* application (Group Control and Group Read) and 200 students (M = 55.30, SD = 18.58) who used it (Group Play and Group Read&Play), the gamified group. We adopted this approach in order to examine the overall impact of the challenge-based gamification on students' learning outcomes. Fig. 6 illustrates the students' performances for each group in percentiles with box-plot diagrams. Normality is not confirmed, thus Wilcoxon-Mann-Whitney rank sum test was performed, with a confidence interval equal to 95%. The null hypothesis of equal differences in means is rejected ($W = 23821$, $p < 0.001$), while the use of *Horses for Courses* presents a moderate level of impact

**Table 3**
The challenge-based gamification results per treatment and variable.

| Variable | Group Control | | | Group Read | | | Group Play | | | Group Read&Play | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n | M | SD | n | M | SD | n | M | SD | n | M | SD |
| Gender | | | | | | | | | | | | |
| Female | 15 | 27.08 | 10.07 | 19 | 35.14 | 19.19 | 28 | 49.08 | 20.46 | 33 | 57.81 | 18.75 |
| Male | 69 | 38.10 | 10.98 | 62 | 49.49 | 17.32 | 72 | 53.90 | 19.43 | 67 | 58.16 | 16.21 |
| Academic major | | | | | | | | | | | | |
| ECE, NTUA | 61 | 39.56 | 10.58 | 65 | 50.56 | 16.55 | 74 | 59.78 | 15.67 | 79 | 61.53 | 15.28 |
| Business Administration | 23 | 27.04 | 8.97 | 16 | 28.13 | 16.18 | 26 | 31.97 | 15.19 | 21 | 44.94 | 17.07 |
| Educational Level | | | | | | | | | | | | |
| UG | 71 | 36.28 | 11.87 | 68 | 47.59 | 19.10 | 85 | 52.45 | 21.03 | 83 | 56.00 | 16.90 |
| MBA | 13 | 35.34 | 10.16 | 13 | 38.46 | 14.62 | 15 | 53.13 | 10.02 | 17 | 68.01 | 14.00 |
| Total | 84 | 36.13 | 11.57 | 81 | 46.13 | 18.68 | 100 | 52.55 | 19.74 | 100 | 58.05 | 17.00 |

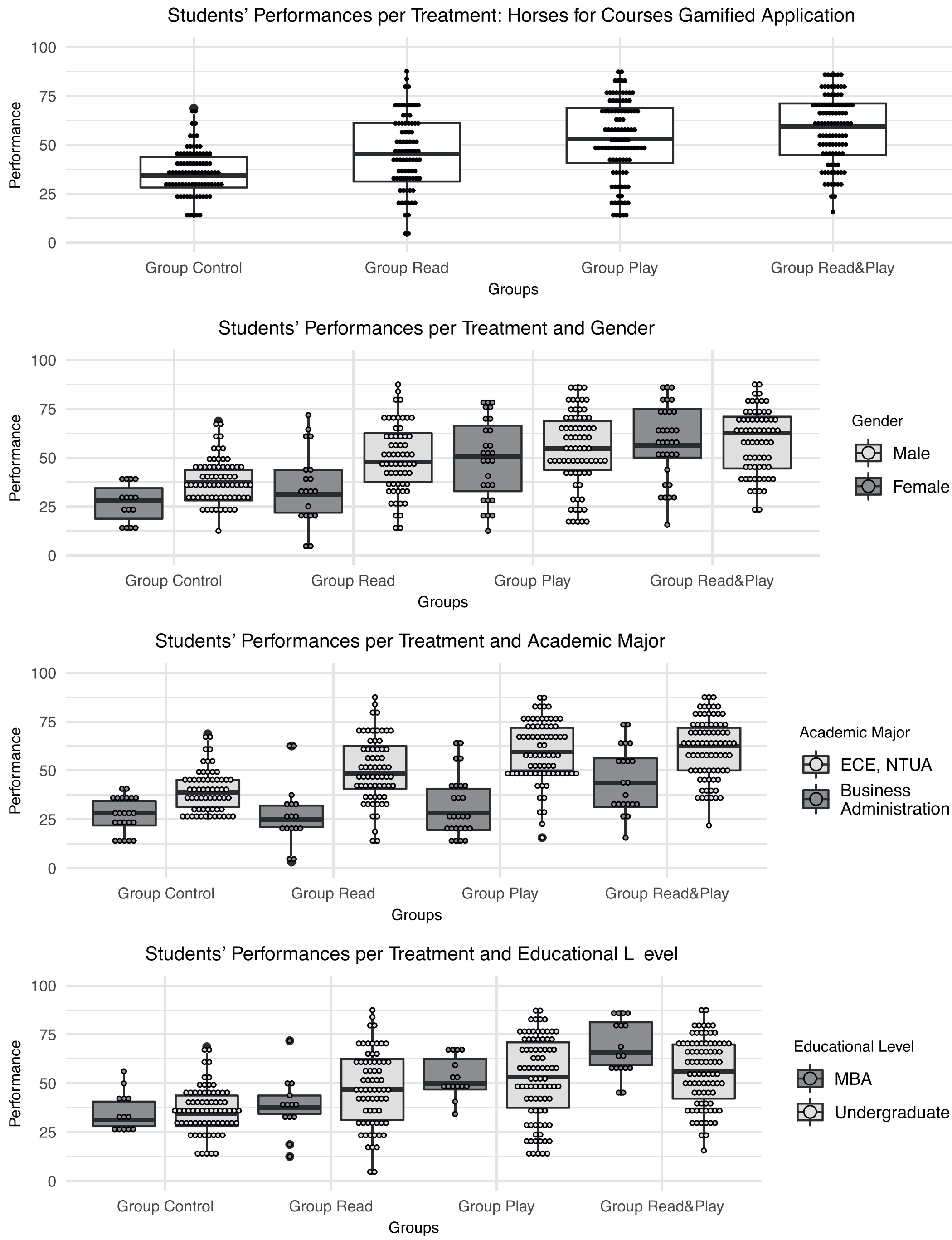


**Fig. 5.** Students' performances per treatment and variable.

**Table 4**
Pairwise multiple comparisons among the groups based on Kruskal-Wallis with Dunn's post test and Cliff Delta effect size.

| Groups | | | Z | P.adj | Delta estimate | Improvement (%) |
|---|---|---|---|---|---|---|
| Control | vs. | Read | -3.70 | 0.001 | 0.35 (medium) | 27.68% |
| Control | vs. | Play | -6.16 | < 0.001 | 0.51 (large) | 45.45% |
| Control | vs. | Read&Play | -8.04 | < 0.001 | 0.69 (large) | 60.67% |
| Play | vs. | Read | 2.25 | 0.049 | - | - |
| Play | vs. | Read&Play | -1.96 | 0.05 | - | - |
| Read&Play | vs. | Read | 4.10 | < 0.001 | - | - |

based on non-parametric Cliff's Delta estimator (delta estimate = 0.44 (medium)) and an improvement regarding mean values of performances, equal to 34.75%.

Students' performances in the final evaluation form (questionnaire) should not be confused with their game performances. Weak positive correlation was found between students' performances at the final evaluation form and their game performances for students who experienced the challenge-based gamification ($r(146) = 0.339$, $p < 0.001$). However, the value of the Pearson's Correlation Coefficient is calculated only for a subset of students ($N = 148$), who used the gamified application, since there was no specific instruction to students to use the same personal details in the gamified application and in the evaluation form.

**Table 5**
Impact of the treatments, variables and their interactions.

| Variables | Gender | | | Academic major | | | Educational Level | | |
|---|---|---|---|---|---|---|---|---|---|
| Groups (N = 365) | df | H | Sig. | df | H | Sign. | df | H | Sign. |
| Treatment (df = 3) (H = 70.84, p < 0.001) | 1 | 7.74 | 0.005 | 1 | 73.41 | < 0.001 | 1 | 0.20 | 0.657 |
| Interaction | 3 | 5.10 | 0.164 | 3 | 7.02 | 0.071 | 3 | 7.63 | 0.054 |
| Variables | | English Proficiency | | | PC Expertise | | | Game Expertise | |
| Groups (n = 146) | df | H | Sig. | df | H | Sign. | df | H | Sign. |
| Treatment (df = 3) (H = 24.77, p < 0.001) | 4 | 15.26 | 0.004 | 4 | 7.83 | 0.098 | 4 | 0.51 | 0.972 |
| Interaction | 9 | 2.58 | 0.995 | 9 | 5.41 | 0.797 | 12 | 11.31 | 0.502 |

**Table 6**
Students' performances per treatment and extra variables.

| Group | Performance | English Proficiency | PC Expertise | Game Expertise |
|---|---|---|---|---|
| Control (n = 28) | M = 32.3 SD = 11.4 | M = 3.36 SD = 1.37 | M = 3.43 SD = 1.14 | M = 3.29 SD = 1.49 |
| Read (n = 27) | M = 44.4 SD = 18.4 | M = 3.59 SD = 1.39 | M = 4 SD = 0.83 | M = 3.37 SD = 1.04 |
| Play (n = 47) | M = 43.2 SD = 20.6 | M = 3.77 sd = 1.37 | M = 3.96 SD = 1.02 | M = 3.26 SD = 1.15 |
| Read&Play (n = 44) | M = 56.7 SD = 20.0 | M = 4.23 SD = 0.96 | M = 4.07 SD = 0.90 | M = 3.27 SD = 1.34 |

## 5. Discussion of results

Overall, the results suggest that the challenge-based gamification improves learning outcomes in a forecasting course. This type of gamification presents the greatest improvement in students' performances when it is combined with traditional teaching methods. However, our results show that even this gamified application alone, integrated in a lecture, may have a positive impact on learning outcomes.

In general, groups who experienced the challenge-based gamification have greater performances than the groups that only participated in traditional teaching methods such as only attending the lecture (Group Control) or reading the paper (Group Read). More precisely, the group whose participants read the respective paper and used the gamified application had the highest performance regardless of their gender, academic major and level of studies. However, the mean value of the performances of this group is not statistically significant different from the group who only used the gamified application. Despite the fact that Group Read&Play had extra 15 minutes to read the respective research and then use the gamified application, the interaction of these two tasks did not seem to have a great impact on students' performances; while each of the tasks, Read or Play did. According to Fisher et al. (1981), the amount of time that students are focused or engaged in an activity is generally positively associated with their learning outcomes. Given that, we might speculate that the students who had to complete two tasks may not have been fully engaged throughout the duration of the tasks. Nevertheless, the aim of our study is to investigate the impact of gamification on students' performances. So, based on our analysis the use of this gamified application presents an improvement regarding the mean values of performances, equal to 34.75%. In addition, the challenge-based gamification may improve students' performance by up to 89.45% compared to only being present at a lecture. Under certain conditions, the use of gamification within less time, may have almost the same impact as reading and using the gamified application, as far as learning outcomes in forecasting are concerned.

Moreover, we can state that a gamified application combined in a lecture may improve learning outcomes at both schools. However, its impact is even more important in the case of engineering students, where both female and male participants had significantly better performances. This finding is in agreement with the fact that gamification has already been incorporated into software engineering and math/science education to a greater degree (Alhammad and Moreno, 2018; Dicheva et al., 2015; Pedreira et al., 2015). Dicheva et al. (2015) argue that one possible explanation could be the lack of fully customized gamified applications in a variety of educational fields or the fact that instructors in these schools are more qualified to develop such applications. Another explanation might be that gamification helps to increase student interest in difficult concepts in engineering (Markopoulos et al., 2015). Thus, engineering students may benefit more by experiencing these subjects as more manageable. Although the research in this field is at a preliminary stage (Alhammad and Moreno, 2018; Pedreira et al., 2015), Pedreira et al. (2015) support that gamification has great potential in software engineering education mainly because of the nature of tasks, which demand high motivation

**Table 7**
Improvement of students' performances per treatment, gender and academic major.

| Group | Academic major | Gender | n | M | SD | Improvement (%) | Delta Est. |
|---|---|---|---|---|---|---|---|
| Control (0%) | ECE, NTUA (9.49%) | Female | 4 | 32.81 | 5.41 | -9.18 | -0.173 (small) |
| | | Male | 57 | 40.03 | 10.72 | 10.80 | 0.195 (small) |
| | Business Administration (-25.17%) | Female | 11 | 25.00 | 10.73 | -30.81 | -0.524 (large) |
| | | Male | 12 | 28.91 | 6.94 | -20.00 | -0.388 (medium) |
| Read (27.66%) | ECE, NTUA (39.93%) | Female | 12 | 44.44 | 16.74 | 23.00 | 0.283 (small) |
| | | Male | 53 | 51.94 | 16.35 | 43.76 | 0.594 (large) |
| | Business Administration (-22.16%) | Female | 7 | 19.20 | 11.04 | -46.87 | -0.731 (large) |
| | | Male | 9 | 35.07 | 16.59 | -2.94 | -0.193 (small) |
| Play (45.44%) | ECE, NTUA (65.45%) | Female | 15 | 59.95 | 16.07 | 65.91 | 0.775 (large) |
| | | Male | 59 | 59.74 | 15.70 | 65.34 | 0.779 (large) |
| | Business Administration (-11.51%) | Female | 13 | 36.54 | 17.97 | 1.13 | -0.068 (negligible) |
| | | Male | 13 | 27.40 | 10.61 | -24.15 | -0.424 (medium) |
| Read&Play (60.65%) | ECE, NTUA (70.30%) | Female | 16 | 68.45 | 13.70 | 89.45 | 0.935 (large) |
| | | Male | 63 | 59.77 | 15.26 | 65.43 | 0.764 (large) |
| | Business Administration (24.38%) | Female | 17 | 47.79 | 17.53 | 32.28 | 0.400 (medium) |
| | | Male | 4 | 32.81 | 7.86 | -9.18 | -0.179 (small) |

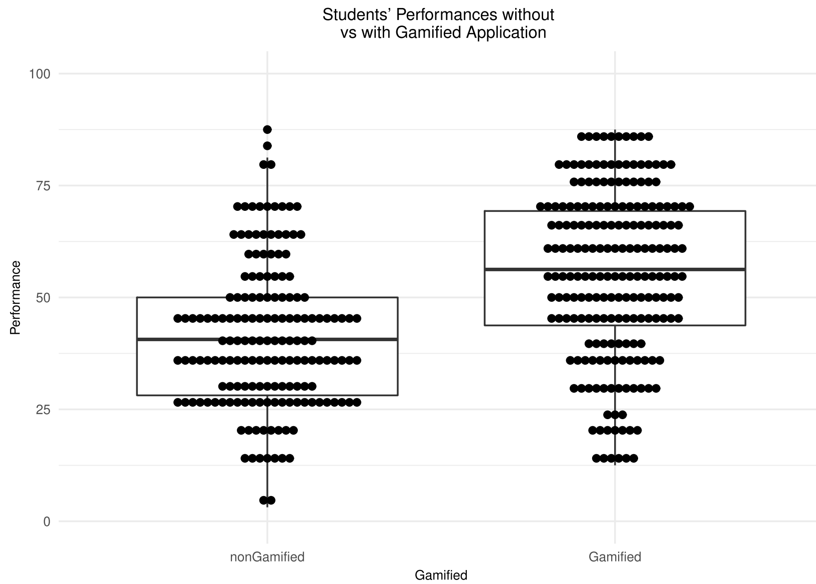


**Fig. 6.** Students' performances of non-gamified and gamified groups.

by students (Pedreira et al., 2015). Our results strengthen this statement, indicating that challenge-based gamification was even more effective in engineering students' performances, without neglecting its potential in business schools, as well.

Another finding, based on the results presented in Table 7, is that female users of this gamified application, independently of their educational background, had higher performances and a higher level of improvement compared to their male counterparts. However, this finding does not apply in non-gamified groups (Group Control and Group Read). More specifically, the female participants of gamified groups at the ECE, NTUA have achieved the highest performances and the highest improvement. Differences in motivations (Carr, 2005) and activities that engage players (Codish and Ravid, 2017; Koivisto and Hamari, 2014), within genders have already been mentioned. Thus, since female participants are generally more motivated by challenge than by competition (McDaniel et al., 2012), females students would probably be even more motivated by challenge-based gamification, which would lead to better performance. Another possible explanation could be that female students receive higher levels of playfulness in a gamified educational content (Codish and Ravid, 2015), which increases their motivation and improves their learning outcomes. It is important to note that while the difference in the sample size of the genders may be the cause of the results, this assumption requires further exploration, and for the time being we may conclude that both female and male students appear to benefit from gamification (O'Donovan et al., 2013).

Last but not least, it is surprising that the variables of students' expertise in use of personal computers and games were not statistically significant for the subset of the participants who reported this data (engineering students = 60, business school students = 66 and MBA students = 20). However, based on the results of Table 6, these groups have similar mean values regarding these variables, sharing probably common characteristics. The students' expertise in English is another parameter which had a statistically significant impact on their performances, since the slides, the research and the interface of the application are all in English. The difference between the mean values of this variable, achieved by the engineering students ($M = 4.45$, $SD = 0.86$) and business school students ($M = 3$, $SD = 1.29$) could justify the lower performances of the business school students in the evaluation form for all groups.

### 5.1. Limitations

While challenge-based gamification has a positive impact on learning outcomes for both engineering and business school students, some limitations should be acknowledged. We gathered and compared performances of all groups although the treatments did not have the same duration as the tasks did. However, this study focuses on the assessment of the evaluation form (questionnaire), which was the same for all groups and experiments, in order to evaluate the challenge-based gamification impact. Although conducting both pre- and post-test would provide further methodological rigor, in the scope of the present study, the extant knowledge of the participants could be assumed to be homogeneous because of the specialized content of the lecture. The content of the lecture, and consequently the topic of the gamified application and the evaluation form are focusing on the specific topic of "Method Selection Protocols" (Petropoulos et al., 2014). Thus, students in any of the stages of their studies would have not otherwise learned this topic. Therefore, the present study's design was economized by conducting only the post-test of knowledge on the topic. Furthermore, engineering students noted better performances than business school students. This fact is probably due to the discrepancy between engineering and business school students in their proficiency in English and in their years of studies.

Additionally, another possible limitation is that there is a difference between the two schools in terms of the incentives to participate in our experiments. Finally, although non-parametric tests have been conducted because of the fact that data was non-normal or heteroscedastic or both, the differences in the sample size of different groups should be considered as a limitation (i.e. 307 undergraduate students versus 58 MBA students). Thus, further experiments could contribute to the conclusions of this study.

## 6. Conclusion

Overall this study contributes to the core literature of how gamification affects desired outcomes (i.e. skills, knowledge, motivations and behavior). According to several state-of-the-art analyses of the field (Koivisto and Hamari, 2019; Majuri et al., 2018; Nacke and Deterding, 2017; Rapp et al., 2019), there has been a relative gap of randomized controlled experiments that could reliably show effects of gamification. Therefore, this study contributes to the corpus via such an experiment by showing that challenge-based gamification (i.e. implementation including points, levels, challenges and a leaderboard) improves learning outcomes. Our findings contribute to the literature of serious games (Connolly et al., 2012) and game-based learning (Hamari et al., 2016; Squire, 2003), as well. More specifically, the study informs the area of scientific and statistics education (Chu, 2007; Love and Hildebrand, 2002), because the object of learning in the experiment was in forecasting techniques. The contribution in this area gets more importance by considering the need for updating and making more engaging the traditional teaching methods (Love and Hildebrand, 2002; Surendeleg et al., 2014), since the statistical skills are fundamentals to get insight of the available data in order to increase awareness towards social problems and understanding our world.

In this study, we conducted a factorial design experiment, using a developed gamification approach named *Horses for Courses*, which provides valuable empirical evidence on how challenge-based gamification and reading differently influence learning performance. The findings of our empirical study, based on a quantitative analysis of our results, are in line with the positive effects of gamification on learning (Buckley and Doyle, 2016a; Hamari et al., 2016; Kuo and Chuang, 2016; Maican et al., 2016; da Rocha Seixas et al., 2016; Simões et al., 2013; Yildirim, 2017) as well as on software engineering education (Alhammad and Moreno, 2018; Pedreira et al., 2015). The results demonstrate that the challenge-based gamification improves students' performances by 34.75% regarding a statistical forecasting topic and that the effect was larger for females or engineering students. The greatest improvements take place when gamification is combined with traditional methods such as reading, however even simply integrating a gamified application into a lecture benefits students.

This research sheds light upon the effect of challenge-based gamification on statistics education by demonstrating improvement in learning outcomes. Apart from the theoretical contribution, this study also provides practical implications to gamification designers and educators. Challenge-based gamification (i.e. points, levels, challenges and leaderboard), can be effectively combined with traditional teaching methods such as lectures and reading in order to improve the learning outcomes in a variety of educational fields related to statistics and stem education. Finally, gamification designers should take into account students' profiles, since our results show that benefits differ across students' characteristics.

With our study, we position challenge-based gamification as a useful educational tool in statistics education in different academic majors under certain circumstances, but also we acknowledge its limitations. Further investigation of the effects of individual game elements or different gamified approaches in statistics or data-related courses with a larger sample is necessary, in order to enhance the scope of the research and further refine its findings. An extension of our research could be to investigate the impact of additional motivational affordances combined or compared with challenge-based gamification, under proper and cautious design. These additional affordances could be related to the actual content of the course or the actual behaviors that the instructors want to promote, e.g. social sharing or responding to forum questions.

## CRediT authorship contribution statement

**Nikoletta-Zampeta Legaki:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Validation, Funding acquisition. **Nannan Xi:** Conceptualization, Writing - original draft, Writing - review & editing. **Juho Hamari:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Kostas Karpouzis:** Methodology, Conceptualization. **Vassilios Assimakopoulos:** Conceptualization, Methodology, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

Special thanks to the students of the forecasting courses at the School of Electrical and Computer Engineering of the National Technical University of Athens, Greece and to the students at the Business Administration Department in the School of Business and Economics of the University of Thessaly, Greece who participated in our experiments and helped to investigate the effect of challenge-based gamification on student learning.

An early version of this study was presented at the International GamiFIN Conference 2018, at the HICSS-52 2019 52nd Hawaii International Conference on System Sciences and at the International GamiFIN Conference 2019. This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie [grant agreement ID 840809]; Business of Finland [Grant No. 5654/31/2018]; Business of Finland [Grant No. 4708/31/2019]; Liikesivistysrahasto [Grant No. 14-7798].

## References

Albritton, M.D., McMullen, P.R., 2006. Classroom integration of statistics and management science via forecasting. Decision Sci. J. Innovat. Educ. 4 (2), 331–336.
Alcivar, I., Abad, A.G., 2016. Design and evaluation of a gamified system for erp training. Comput. Human. Behav. 58, 109–118.
Alhammad, M.M., Moreno, A.M., 2018. Gamification in software engineering education: asystematic mapping. J. Syst. Softw. 141, 131–150.
Attali, Y., Arieli-Attali, M., 2015. Gamification in assessment: do points affect test performance? Comput. Educ. 83, 57–63.
Barnett, B., Kirtland, M., Ganapathy, M., 2005. An architecture for distributed applications on the internet: overview of the microsoft?. net framework. J. STEM Educ. 4 (1).
Buckley, P., Doyle, E., 2016. Gamification and student motivation. Interact. Learn. Environ. 24 (6), 1162–1175.
Buckley, P., Doyle, E., 2016. Using web-based collaborative forecasting to enhance information literacy and disciplinary knowledge. Interact. Learn. Environ. 24 (7), 1574–1589.
Buckley, P., Doyle, E., 2017. Individualising gamification: an investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. Comput. Educ. 106, 43–55.
Buckley, P., Garvey, J., McGrath, F., 2011. A case study on using prediction markets as a rich environment for active learning. Comput. Educ. 56 (2), 418–428.
Bunchball, I., 2010. Gamification 101: an introduction to the use of game dynamics to influence behavior. White paper 9.
Caponetto, I., Earp, J., Ott, M., 2014. Gamification and education: A literature review. European Conference on Games Based Learning. 1. Academic Conferences International Limited, pp. 50.
Carr, D., 2005. Contexts, gaming pleasures, and gendered preferences. Simulat. Gaming 36 (4), 464–482.
Christy, K.R., Fox, J., 2014. Leaderboards in a virtual classroom: a test of stereotype threat and social comparison explanations for women's math performance. Comput. Educat. 78, 66–77.
Chu, S., 2007. Some initiatives in a business forecasting course. J. Stat. Educ. 15 (2).
Cliff, N., 2014. Ordinal Methods for Behavioral Data Analysis. Psychology Press.
Cobb, G., 1992. Teaching statistics. Heed. Call Change 22, 3–43.
Codish, D., Ravid, G., 2015. Detecting playfulness in educational gamification through behavior patterns. IBM J. Res. Dev. 59 (6), 6–11.
Codish, D., Ravid, G., 2017. Gender Moderation in Gamification: Does One Size Fit All?.
Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M., 2012. A systematic literature review of empirical evidence on computer games and serious games. Comput. Educ. 59 (2), 661–686.
Craighead, C.W., 2004. Right on target for time-series forecasting. Decis. Sci. J. Innovat. Educ. 2 (2), 207–212.
Debnath, S.C., Tandon, S., Pointer, L.V., 2007. Designing business school courses to promote student motivation: an application of the job characteristics model. J. Manag. Educ. 31 (6), 812–831.
Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From game design elements to gamefulness: Defining "gamification". Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. ACM, New York, NY, USA, pp. 9–15.
Dichev, C., Dicheva, D., 2017. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. Int. J. Educ. Technol. Higher Educ. 14 (1), 9.

Dicheva, D., Dichev, C., Agre, G., Angelova, G., 2015. Gamification in education: a systematic mapping study. Educ. Technol. Soc. 18 (3), 75–89.
Dinno, A., 2015. Nonparametric pairwise multiple comparisons in independent groups using dunn's test. Stata. J. 15 (1), 292–300.
DomíNguez, A., Saenz-De-Navarrete, J., De-Marcos, L., FernáNdez-Sanz, L., PagéS, C., MartíNez-HerráIz, J.-J., 2013. Gamifying learning experiences: practical implications and outcomes. Comput. Educ. 63, 380–392.
Donihue, M.R., 1995. Teaching economic forecasting to undergraduates. J. Econ. Educ. 26 (2), 113–121.
Dunn, O.J., 1961. Multiple comparisons among means. J. Am. Stat. Assoc. 56 (293), 52–64.
Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R., Cahen, L.S., Dishaw, M.M., 1981. Teaching behaviors, academic learning time, and student achievement: an overview. J. Classroom Interaction 17 (1), 2–15.
Gallaugher, J.M., Ramanathan, S.C., 1996. Choosing a client/server architecture: a comparison of two-and three-tier systems. Inf. Syst. Manag. 13 (2), 7–13.
Gardner, L., 2008. Using a spreadsheet for active learning projects in operations management. INFORMS Trans. Educ. 8 (2), 75–88.
Garfield, J., Ben-Zvi, D., 2007. How students learn statistics revisited: a current review of research on teaching and learning statistics. Int. Stat. Rev. 75 (3), 372–396.
Gavirneni, S., 2008. Teaching the subjective aspect of forecasting through the use of basketball scores. Decision Sci. J. Innovat. Educ. 6 (1), 187–195.
Gel, Y.R., O'Hara Hines, R.J., Chen, H., Noguchi, K., Schoner, V., 2014. Developing and assessing e-learning techniques for teaching forecasting. J. Educ. Bus. 89 (5), 215–221.
Giullian, M.A., Odom, M.D., Totaro, M.W., 2000. Developing essential skills for success in the business world: a look at forecasting. J. Appl. Bus. Res. 16 (3), 51–62.
González, C., Area, M., 2013. Breaking the rules: Gamification of learning and educational materials. Proceedings of the 2nd international workshop on interaction design in educational environments. pp. 7–53.
Hamari, J., 2013. Transforming homo economicus into homo ludens: afield experiment on gamification in a utilitarian peer-to-peer trading service. Electron. Commer. Res. Appl. 12 (4), 236–245.
Hamari, J., Shernoff, D.J., Rowe, E., Coller, B., Asbell-Clarke, J., Edwards, T., 2016. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput. Human. Behav. 54, 170–179.
Hamari, J., Tuunanen, J., 2014. Player Types: A Meta-synthesis.
Hanke, J., 1984. The teachers/practitioners corner forecasting in business schools: a survey. J. Forecast. 3 (2), 229–234.
Hanke, J., 1989. Forecasting in business schools: a follow-up survey. Int. J. Forecast. 5 (2), 259–262.
Hanus, M.D., Fox, J., 2015. Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Comput. Educ. 80, 152–161.
Hunicke, R., LeBlanc, M., Zubek, R., 2004. Mda: a formal approach to game design and game research. Proceedings of the AAAI Workshop on Challenges in Game AI. 4. pp. 1722.
Huotari, K., Hamari, J., 2017. A definition for gamification: anchoring gamification in the service marketing literature. Electron. Markets 27 (1), 21–31.
Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., Hides, L., 2016. Gamification for health and wellbeing: a systematic review of the literature. Internet Intervent. 6, 89–106.
Kapp, K.M., 2013. The Gamification of Learning and Instruction Fieldbook: Ideas into Practice. John Wiley & Sons.
Karpouzis, K., Caridakis, G., Fotinea, S.-E., Efthimiou, E., 2007. Educational resources and implementation of a greek sign language synthesis architecture. Comput. Educ. 49 (1), 54–74.
Kasurinen, J., Knutas, A., 2018. Publication trends in gamification: a systematic mapping study. Comput. Sci. Rev. 27, 33–44.
Koivisto, J., Hamari, J., 2014. Demographic differences in perceived benefits from gamification. Comput. Human. Behav. 35, 179–188.
Koivisto, J., Hamari, J., 2019. The rise of motivational information systems: areview of gamification research. Int. J. Inf. Manage. 45, 191–210.
Kros, J.F., Rowe, W.J., 2016. Business school forecasting for the real world. Advances in Business and Management Forecasting. Emerald Group Publishing Limited, pp. 149–161.
Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47 (260), 583–621.
Kuo, M.-S., Chuang, T.-Y., 2016. How gamification motivates visits and engagement for online academic dissemination - an empirical study. Comput. Human. Behav. 55, 16–27.
Lambruschini, B.B., Pizarro, W.G., 2015. Tech-gamification in university engineering education: Captivating students, generating knowledge. 2015 10th International Conference on Computer Science & Education (ICCSE). IEEE, pp. 295–299.
Landers, R.N., Auer, E.M., Collmus, A.B., Armstrong, M.B., 2018. Gamification science, its history and future: Definitions and a research agenda. Simulat. Gaming. 1046878118774385
Lewandowski, S.M., 1998. Frameworks for component-based client/server computing. ACM Comput. Surv. (CSUR) 30 (1), 3–27.
Loomis, D.G., Cox, J.E., 2000. A course in economic forecasting: rationale and content. J. Econ. Educ. 31 (4), 349–357.
Loomis, D.G., Cox Jr, J.E., 2003. Principles for teaching economic forecasting. Int. Rev. Econ. Educ. 2 (1), 69–79.
Love, T.E., Hildebrand, D.K., 2002. Statistics education and the making statistics more effective in schools of business conferences. Am. Stat. 56 (2), 107–112.
Macbeth, G., Razumiejczyk, E., Ledesma, R.D., 2011. Cliff'S delta calculator: a non-parametric effect size program for two groups of observations. Universitas Psychol. 10 (2), 545–555.
Maican, C., Lixandroiu, R., Constantin, C., 2016. Interactivia.ro - a study of a gamification framework using zero-cost tools. Comput. Human. Behav. 61, 186–197.
Majuri, J., Koivisto, J., Hamari, J., 2018. Gamification of education and learning: A review of empirical literature. Proceedings of the 2nd International GamiFIN Conference, GamiFIN 2018. CEUR-WS.
Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 2008. Forecasting methods and applications. John wiley & sons.
Markopoulos, A.P., Fragkou, A., Kasidiaris, P.D., Davim, J.P., 2015. Gamification in engineering education and professional training. Int. J. Mech. Eng. Educ. 43 (2), 118–131.
McDaniel, R., Lindgren, R., Friskics, J., 2012. Using badges for shaping interactions in online learning environments. 2012 IEEE International Professional Communication Conference. IEEE, pp. 1–4.
McEwen, B.C., 1994. Teaching critical thinking skills in business education. J. Educ. Bus. 70 (2), 99–103.
Mendez-Carbajo, D., 2018. Experiential learning in intermediate macroeconomics through fredcast. Int. Rev. Econ. Educ.
Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A., 2017. Gamified crowdsourcing: conceptualization, literature review, and future agenda. Int. J. Hum. Comput. Stud. 106, 26–43.
Morschheuser, B., Hassan, L., Werder, K., Hamari, J., 2018. How to design gamification? a method for engineering gamified software. Inf. Softw. Technol. 95, 219–237.
Nacke, L.E., Deterding, C.S., 2017. The maturing of gamification research. Comput. Hum. Behav. 450–454.
Nah, F.F.-H., Zeng, Q., Telaprolu, V.R., Ayyappa, A.P., Eschenbrenner, B., 2014. Gamification of education: a review of literature. International Conference on hci in Business. Springer, pp. 401–409.
O'Donovan, S., Gain, J., Marais, P., 2013. A case study in the gamification of a university-level games development course. Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference. ACM, pp. 242–251.
Osatuyi, B., Osatuyi, T., de la Rosa, R., 2018. Systematic review of gamification research in is education: a multi-method approach. CAIS 42, 5.
Pedreira, O., García, F., Brisaboa, N., Piattini, M., 2015. Gamification in software engineering - a systematic mapping. Inf. Softw. Technol. 57, 157–168.
Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. Horses for courses' in demand forecasting. Eur. J. Oper. Res. 237 (1), 152–163.
Pfeffer, J., Fong, C.T., 2002. The end of business schools? less success than meets the eye. Acad. Manag. Learn. Educ. 1 (1), 78–95.
Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., Cena, F., 2019. Strengthening gamification studies: current trends and future opportunities of gamification research. Int. J. Hum. Comput. Stud. 127, 1–6.
Reiners, T., Wood, L.C., Chang, V., Gütl, C., Herrington, J., Teräs, H., Gregory, S., 2012. Operationalising Gamification in an Educational Authentic Environment.
da Rocha Seixas, L., Gomes, A.S., de Melo Filho, I.J., 2016. Effectiveness of gamification in the engagement of students. Comput. Human. Behav. 58, 48–63.
Sánchez-Martín, J., Dávila-Acedo, M.A., et al., 2017. Just a game? Gamifying a general science class at university: collaborative and competitive work implications. Think. Skill. Creat. 26, 51–59.
Scheirer, C.J., Ray, W.S., Hare, N., 1976. The analysis of ranked data derived from completely randomized factorial designs. Biometrics 32 (2), 429–434.
Seaborn, K., Fels, D.I., 2015. Gamification in theory and action: a survey. Int. J. Hum. Comput. Stud. 74, 14–31.
Simões, J., Redondo, R.D., Vilas, A.F., 2013. A social gamification framework for a k-6 learning platform. Comput. Hum. Behav. 29 (2), 345–353.
Snider, B.R., Eliasson, J.B., 2013. Beat the instructor: an introductory forecasting game. Decis. Sci. J. Innovat. Educ. 11 (2), 147–157.
Snodgrass, J.G., Dengah, H.F., Lacy, M.G., Fagan, J., 2013. A formal anthropological view of motivation models of problematic mmo play: achievement, social, and immersion factors in the context of culture. Transcult. Psychiatry 50 (2), 235–262.
de Sousa Borges, S., Durelli, V.H., Reis, H.M., Isotani, S., 2014. A systematic mapping on gamification applied to education. Proceedings of the 29th annual ACM symposium on applied computing. ACM, pp. 216–222.
Spithourakis, G.P., Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2015. Amplifying the learning effects via a forecasting and foresight support system. Int. J. Forecast. 31 (1), 20–32.
Squire, K., 2003. Video games in education. Int. J. Intell. Games Simulat. 2 (1), 49–62.
Surendeleg, G., Murwa, V., Yun, H.-K., Kim, Y.S., 2014. The role of gamification in education-a literature review. Contempor. Eng. Sci. 7 (29), 1609–1616.
Tabatabai, M., Gamble, R., 1997. Business statistics education: content and software in undergraduate business statistics courses. J. Educ. Bus. 73 (1), 48–53.
Torres, C., Babo, L., Mendonça, J., 2018. Engaging students to learn forecasting methods. 10th annual International Technology, Conference on Education and New Learning Technologies. INTED, pp. 10337–10343.
Wilcox, R.R., 2006. Graphical methods for assessing effect size: some alternatives to cohen's d. J. Exp. Educ. 74 (4), 351–367.
Xi, N., Hamari, J., 2019. Does gamification satisfy needs? a study on the relationship between gamification features and intrinsic need satisfaction. Int. J. Inf. Manage. 46, 210–221.
Xi, N., Hamari, J., 2020. Does gamification affect brand engagement and equity? a study in online brand communities. J. Bus. Res. 109, 449–460.
Yee, N., 2006. Motivations for play in online games. CyberPsychol. Behav. 9 (6), 772–775.
Yee, N., Ducheneaut, N., Nelson, L., 2012. Online gaming motivations scale: development and validation. Proceedings of the SIGCHI conference on human factors in computing

systems. ACM, pp. 2803–2806.
Yildirim, I., 2017. The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. Internet Higher Educ. 33, 86–92.
Zar, J.H., et al., 1999. Biostatistical analysis. Pearson Education India.
Zichermann, G., Cunningham, C., 2011. Gamification by design: Implementing game mechanics in web and mobile apps. " O'Reilly Media, Inc.".

**Nikoletta-Zampeta Legaki** is a Marie - Curie Researcher at Gamification Group, Tampere University. She pursuits her PhD in School of Electrical and Computer Engineering of National Technical University of Athens (NTUA), Greece. She has been a researcher and teaching assistant in Forecasting and Strategy Unit, School of Electrical and Computer Engineering since 2012. During this period, she has participated in various research projects about forecasting and data analytics and she has worked as a consultant in Financial Services and Risk Management, in EY, Greece. Her research interests lie on time series forecasting, business forecasting information systems, gamification and educational methods in teaching forecasting.

**Dr. Nannan Xi** is a Postdoctoral researcher in Gamification Group. She got her PhD in marketing management from Zhongnan University of Economic and Law (ZUEL), China and holds M.Sc. in International Business from the University of Lincoln, UK. Her dissertation was awarded as the excellent Ph.D. dissertation in ZUEL. Xi's current research focuses on gamification in marketing, especially in gamified interaction in brand management. In addition, her research interests include customer management in gamification and virtual reality/augmented reality/mixed reality in business and sharing economy.

**Dr. Juho Hamari** is a Professor of Gamification at the Faculty of Information Technology and Communications, Tampere University. He leads the Gamification Group. His research covers several forms of information technologies such as games, motivational information systems, new media (social networking services, eSports), peer-to-peer economies (sharing economy, crowdsourcing), and virtual economies. Dr. Hamari has authored several seminal empirical, theoretical and meta-analytical scholarly articles on these topics from perspective of consumer behavior, human-computer interaction, game studies and information systems science. His research has been published in a variety of prestigious venues.

**Dr. Kostas Karpouzis** is currently an Associate Researcher at the Institute of Communication and Computer Systems (ICCS) of the National Technical University of Athens (NTUA) in Greece. His research interests lie in the areas of human computer interaction, emotion understanding, affective and natural interaction, serious games and games based assessment and learning. He has participated in more than twenty research projects at Greek and European level. He is also a member of the Student Activities Chair for the IEEE Greece Section and a member of the Editorial Board for international journals.

**Prof. Vassilios Assimakopoulos** is a professor of Forecasting Systems at the School of Electrical and Computer Engineering of the National Technical University of Athens (NTUA). He is the author of over than 60 original publications and papers in international journals and conferences. Moreover, he has conducted research on innovative tools for management support and decision systems design. He has participated and led numerous projects, funded by National and European institutes. He specializes in various fields of Strategic Management, Design and Development of Information systems, Business Resource Management, Statistical and Forecasting Techniques using time series.