Technological Forecasting & Social Change 167 (2021) 120725

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Gamification to avoid cognitive biases: An experiment of gamifying a forecasting course

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

[a] Gamification Group, Tampere University, Kalevantie 4, Tampere 33100, Finland
[b] Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Heroon Polytechniou 9, Zografou 15780, Greece
[c] Intelligent Systems, Content and Interaction Lab, National Technical University of Athens, Heroon Polytechniou 9, Zografou 15780, Greece

ARTICLE INFO

Keywords:
Forecasting
Judgment Bias
Teaching
Gamification
Decision Making
Education

ABSTRACT

In their daily lives, people are confronted with situations where they need to form a schema of possible future scenarios and the likelihood of them occurring, be it about climate change, economic up- or downturn, or even the potential success of a romantic date. Be these issues of mundane or universal importance, this judgmental forecasting poses people with a difficult pervasive cognitive challenge. Commonly, judgmental forecasting is taught in forecasting courses syllabi, and the pedagogy surrounding it is challenging. However, gamification and game-based learning have risen as promising tools to simulate different kinds of scenarios and stimulate cognitive problem solving. This study investigates the effects of a gamified application with points, levels, challenges, storytelling and leaderboard for teaching judgmental forecasting by conducting a 2×2 between-subjects experiment (treatments: i) read: yes vs no, and ii) gamification: yes vs no), with a sample of 285 students of a School of Electrical and Computer Engineering and a Business Administration Department. The findings indicate that the gamified application improved learning outcomes regarding the heuristics and biases that affect judgmental forecasting by almost 15%, supporting the use of gamification in forecasting education.

1. Introduction

How many forecasts have you made today? In our everyday life we make lots of forecasts at individual- or even business-level based on our judgment, and this can be characterized as judgmental forecasting. However, these forecasts are rarely evaluated, mainly because we generally try avoid taking responsibility for any forecasting errors (Makridakis et al., 2008). Unfortunately, there is no such option in the business environment. Judgmental forecasting in organizations remains an important and high-risk topic regarding decision making, and is linked to a variety of departments such as marketing, sales, management and others, and ultimately has potentially great cost implications (Makridakis, 1996; Makridakis et al., 2008).

As a process of producing forecasts based mainly on human judgment, judgmental forecasting is of paramount importance in business planning, especially when no data exists or important changes are forthcoming (Makridakis et al., 2008). The case of long-term technological forecasts is another field that underlines the significance of judgmental forecasting in order to analyze future opportunities and timely make plans for future investments and improve decision making (Albright, 2002). Human judgement though is often characterized by important biases, limitations and fallacies in this field, regarding short- or long-term forecasting, and it may lead to great cost. For example, statements by experts that failed to forecast the future such as: *There is no reason anyone would want a computer in their home* by the founder of the Digital Equipment Corporation, (1977) have remained as evidence of prejudicial judgmental forecasts that have affected a firms future. On the other hand, the elimination of judgmental forecasting has cost millions of dollars for prestigious companies such as Nike and Goodyear (Worthen, 2003), so these cases remind us of the importance of judgment in situations where planning and decisions are necessary.

Intense research has been conducted regarding the improvement of judgmental forecasting accuracy (Han et al., 2019; Lawrence et al., 2006), the comparison between statistical and judgmental forecasting accuracy (Lawrence et al., 2006), judgmental adjustment (Lawrence et al., 2006; Leitner and Leopold-Wildburger, 2011; Makridakis et al., 2008), and judgment bias (Eroglu and Croxton, 2010; Lawrence et al., 2006). The importance of judgmental forecasting is reflected in forecasting courses' syllabi (Cox Jr and Loomis, 2006), textbooks (Hyndman and Koehler, 2006; Makridakis et al., 2008) and in long-term forecasts (Albright, 2002). Since 1987, forecasting books have contained coverage of judgmental forecasting as a separate topic, since human

* Corresponding author.

https://doi.org/10.1016/j.techfore.2021.120725
Received 31 March 2020; Received in revised form 27 January 2021; Accepted 3 March 2021
Available online 17 March 2021

judgment is subjected to biases and limitations which we should be made aware of (Lawrence et al., 2006). Similarly, decision making research proposes interventions to debias and improve decision making towards high-risk global challenges for the future, such as climate change, management of clean water or pandemics for more than 40 years (Bhargava and Loewenstein, 2015; Morewedge et al., 2015; Soll et al., 2015; Tversky and Kahneman, 1974). Games and consequently gamification have been suggested as effective in training behaviors in heavy cognitive tasks such as decision making, because of their potential to be positively related with intrinsic motivation. But only a few studies have focused on innovative teaching and disseminating these biases using game-based learning approaches. Recently a few serious games have been designed and suggested in this field, speculating that the experience of a game can be efficient and fruitful for both strengthen the knowledge of and mitigating the cognitive biases in order to support the rational thinking (Dunbar et al., 2014; Jacobson and Dargue, 2019; McKernan et al., 2015; Morewedge et al., 2015).

Due to the great cost that serious games' development may be accompanied (Papastergiou, 2009), a possible way to address the above-mentioned lack issue could be the use of gamification to make the learning process of judgmental heuristics and eventually judgmental forecasting easier and more efficient. During recent years, gamification research has grown exponentially, highlighting the impact and the potential of adopting this approach. Several systematic reviews dedicated to the topic offer us invaluable mappings of the strategies and effects of gamification, especially in the educational domain (Dicheva et al., 2015; Koivisto and Hamari, 2019; Seaborn and Fels, 2015a; de Sousa Borges et al., 2014). As a method to enhance users experience through affordances in a way similar to games (Huotari and Hamari, 2017), gamification is famous today both inside and outside of universities and schools. For example, *Doulingo* and *Kahoot* are some of the most successful gamified applications worldwide. Gamification has been integrated into a variety of educational subjects, with mostly positive results in a wide range of settings (Dicheva et al., 2015; Koivisto and Hamari, 2019; Seaborn and Fels, 2015a; de Sousa Borges et al., 2014). Most empirical studies have used points, badges and leaderboards (Koivisto and Hamari, 2019; Majuri et al., 2018; de Sousa Borges et al., 2014). However, forecasting courses have only slightly integrated gamification approaches in their teaching methods, despite the importance of this topic within organizations (Makridakis et al., 2008), and also the need to continually update teaching methods (Cox Jr and Loomis, 2006). A few initiatives such as using spreadsheets, competitions, scores and simple games (Craighead, 2004; Gavirneni, 2008; Mendez-Carbajo, 2019; Snider and Eliasson, 2013) have had positive outcomes, which is encouraging for the use of gamification in teaching forecasting. Considering the multidisciplinary nature of forecasting, more detailed researches have combined gamification strategies with forecasting per se, in order to teach some of the more broader concepts of risk management (Buckley et al., 2011), taxation (Buckley and Doyle, 2016a; 2016b; 2017) or to examine the origins of asset price bubbles (Bao et al., 2017).

Thus, while the integration of gamification in the process of teaching forecasting is not a newfangled notion, the impact of gamification regarding learning outcomes in forecasting education has not been experimentally or extensively explored, nor has judgmental forecasting been addressed as a specific topic. Empirical researches stand mainly positive regarding gamification in management studies, but still there is a lack of effective design regarding gamification and controlled experimental research in education (Koivisto and Hamari, 2019) that would further support gamification adoption in teaching methods, especially about cognitive biases. Therefore, this study integrates a gamified strategy which is composed of points, levels, challenges, storytelling and leaderboard, into the design of a web-application named *JudgeIt*. This strategy was exclusively developed for the needs of this study. Hence, the scenario and content of this gamified application is based on the research of Tversky and Kahneman (1974), and aspires to provide a complimentary educational tool that teaches and raises awareness regarding the heuristics and biases which have a great impact on human judgment.

This study aims to examine the effect of the proposed gamified strategy, which includes points, levels, challenges, storytelling and a leaderboard regarding the students learning outcomes about the biases and heuristics of human judgment. We conducted a series of 2×2 between subject experiments, using treatments: i) read: yes vs no, and ii) gamification: yes vs no. The impact of these treatments on students learning outcomes were also investigated, along with demographic characteristics such as gender, school, educational level. The total sample (N=285) is composed of 184 undergraduate students and 20 MBA students from the Electrical and Computer Engineering School of the National Technical University of Athens, Greece (hereafter mentioned as ECE) and 81 undergraduate students from the Business Administration Department in the School of Business and Economics of the University of Thessaly, Greece (hereafter Bus.Adm.). Our findings show that the proposed gamification strategy improves student learning outcomes regarding judgment bias, contributing to the investigation of gamification effects on education of forecasting and decision making, and eventually supporting rational thinking in decision making for a sustainable future under proper prediction.

2. Conceptual background

2.1. Research needs in judgmental forecasting

Forecasting refers to the process of predicting the future as accurately as possible. Historical data or information about relevant future events that might affect this process should be considered in forecasting process (Hyndman and Athanasopoulos, 2018). Given the progress of technology, nowadays there is a plethora of available statistical methods and machine learning algorithms that produce accurate statistical forecasts, taking advantage of the increased production of data (Makridakis et al., 2018; 2020). However historical data is not always available, or if so, the data does not necessarily include the human knowledge regarding possible relevant future events or changes that impact on the attempted forecasting (Makridakis et al., 2008). Technological forecasting which refers to long-term forecasting, also demands thinking outside the box, because of the imminent changes in the existing patterns and relationships in the long run (Makridakis et al., 2008). In such cases, human judgment is the only alternative means to support forecasting, even when it concerns planning or major decision making in the business environment. Thus, judgmental forecasting, when described as forecasting that relies on human judgment, is necessary not only at an individual level but also in the business environment (Makridakis et al., 2008). Furthermore, judgmental forecasting is strongly linked with planning and eventually decision making in a variety of business departments such as sales, inventory, supply chain and others (Hyndman and Athanasopoulos, 2018; Lawrence et al., 2006; Makridakis et al., 2008).

Judgmental forecasting is linked to decision making, especially in the business environment where planning and scheduling are necessary activities. It has a crucial role since usually it is inevitably and strongly affiliated with great cost in regard to error prevention (Lawrence et al., 2006; Makridakis et al., 2008). Therefore, for more than 30 years, judgmental forecasting has gained researchers attention (Cox Jr and Loomis, 2006) and has been introduced into textbooks and forecasting courses, and great progress has been made to improve its accuracy (Cox Jr and Loomis, 2006; Lawrence et al., 2006). A plethora of judgmental forecasting methods are taught in forecasting courses such as Delphi, structured analogies, scenario forecasting, and rule-based forecasting (Hyndman and Athanasopoulos, 2018; Makridakis et al., 2008). Research also focuses on the improvement of judgmental forecasting accuracy. The remarkable work of Tetlock and Gardner (2016) named the *Good Judgment Project* showed that there are people termed as superforecasters whose judgmental forecasts are of high accuracy. However, despite this noted progress (Lawrence et al., 2006), judgmental forecasting necessarily entails forecasting errors, mainly because of the limitations and bias of human judgment (Bolger and Wright, 2017; Makridakis, 1996; Tversky and Kahneman, 1974).

The bias of human judgment is an important aspect of judgmental forecasting (Tversky and Kahneman, 1974), and people are mainly unaware of their existence or operation (Dunbar et al., 2014). Decisions makers and experts are prone to biases, even unconsciously, in case of decisions of forecasts estimation with great uncertainty as much as ordinary people. Hence, most of the textbooks which are relevant to forecasting topics mention common judgmental biases and propose ways to avoid or mitigate them. In this regard, Lawrence et al. (2006) suggest that the dissemination of knowledge on human bias should be strengthened. This suggestion could eventually lead to further improvements in the accuracy of judgmental forecasting. However, there is a lack of research and courses that focus on teaching the cognitive biases in forecasting process, or that motivate students or practitioners to learn judgmental forecasting methods and improve their knowledge regarding the biases of human judgment. A few serious games have been shown more effective than traditional teaching methods (Dunbar et al., 2014) about the knowledge of cognitive biases but the majority of teaching principles regarding forecasting focus on statistical forecasting (Cox Jr and Loomis, 2006).

2.2. Gamification in education

While no special attention has been given to motivate people to explore the topic of human judgment bias, a new process of using affordances for gameful experiences in order to engage the target audience called gamification (Huotari and Hamari, 2017) is becoming increasingly well known (Koivisto and Hamari, 2019). Gamification has been used in a variety of life aspects, and intensively within the education sector (Dicheva et al., 2015; Hanus and Fox, 2015; Koivisto and Hamari, 2019; Nah et al., 2014; Seaborn and Fels, 2015a). Literature reviews regarding the adoption of gamification in education mainly mention the positive effects it has on a variety of behavioral and psychological outcomes (Caponetto et al., 2014; Koivisto and Hamari, 2019; Majuri et al., 2018; Nah et al., 2014; Osatuyi et al., 2018; Reiners et al., 2012; Seaborn and Fels, 2015a; Yildirim, 2017). The plethora of well-known commercial gamified applications reinforces the potential of gamification in the educational sector, and gamification has been effectively applied in various educational subjects and to different target audiences ranging from elementary up to lifelong learning.

Despite the successful adoption of gamification in educational subjects, there is a need for more controlled experimental research to investigate how gamification can be effectively integrated into education for specific taught subjects or audiences (Dichev and Dicheva, 2017; Koivisto and Hamari, 2019). Up to now, most of the studies that refer to gamified interventions in the educational process mention the use of points, badges and leaderboards (Dicheva et al., 2015; Koivisto and Hamari, 2019; Majuri et al., 2018; Pedreira et al., 2015). All of these elements have been recently categorized as a form of achievement-based gamification (Xi and Hamari, 2019) which focuses on challenges and on fostering a feeling of competence. Based on the same research, this type of gamification is positively related with an overall intrinsic satisfaction need, including elements of autonomy, relatedness and competence. From a pedagogical perspective and in accordance with the study of Xi and Hamari (2019), students who are intrinsically motivated by competitive and challenging active learning experiences are more likely to experience better learning outcomes and boost their problem-solving thinking (Markopoulos et al., 2015; Zepke and Leach, 2010). Although these gamification elements seem effective in an educational context, reported negative outcomes warn about the need for a cautious design of gamification interventions, since its effects may vary in relation to the subject, learning type, or the time-duration of the intervention (Buckley and Doyle, 2017; Dicheva et al., 2015; Hamari et al., 2014; Hanus and Fox, 2015; Koivisto and Hamari, 2019; da Rocha Seixas et al., 2016; Seaborn and Fels, 2015a).

Apart from the need for more empirical data regarding achievement-based gamification, the impact of other motivational affordances and other types of gamification in education seems promising and indicates a need for further controlled experimental research (Dichev and Dicheva, 2017; Dicheva et al., 2015; Koivisto and Hamari, 2019). Given the categorization of Xi and Hamari (2019), two other types of gamification social- and immersion-based gamification are also positively related with autonomy need satisfaction, and eventually with intrinsic motivation. Social-based gamification mainly focuses on competition and/or collaboration, while immersion-based gamification puts an emphasis on role playing and narrative enhancement. So, these three types of gamification cover the majority of the motivational affordances used in an educational context. Thus, when exploring the motivational affordances seen in each category, competition seems to be a common element between achievement- and social-based gamification. This fact can justify the efficient use of a competition setting in a variety of educational contexts (Koivisto and Hamari, 2014; Majuri et al., 2018; da Rocha Seixas et al., 2016), including forecasting subjects (Mendez-Carbajo, 2019; Snider and Eliasson, 2013). On the other hand, immersion-based gamification is strongly related with autonomy, which has been proposed as a pedagogical approach in order to improve students engagement (Zepke and Leach, 2010). In this regard, narrative and storytelling have been used in statistical courses, as well as motivating student participation and comprehension (Mallette and Saldaña, 2019; Novak et al., 2016). Therefore, a combination of motivational affordances such as challenges, competition and storytelling which cover the range of the discussed gamification types could be effective in regard to students motivation, and eventually their learning outcomes in more demanding educational subjects such as forecasting.

2.3. Gamification in cognitive biases, forecasting and decision making

Forecasting and statistical courses in general, have been accused of being complex topics that discourage students to participate (Craighead, 2004; Gel et al., 2014). However, these topics remain of paramount importance for the business environment, and are strongly linked to marketing (Hofacker et al., 2016), human resources (Dale, 2014), and management (Müller et al., 2015). Given the exponential increase of data, forecasting has gained a crucial role. Forecasting methods and systems note great progress by incorporating neural networks and machine learning techniques, aside from using simple statistical methods to support the decision making process (Makridakis et al., 2018). Nonetheless, the need to produce forecasts that rely on human judgment remains an important feature at individual- and business-levels.

Considering the amount of information that we are sounded by, the need to comprehend the bias that human judgment is limited by is even more essential. This need is reflected in forecasting courses that include judgmental forecasting as a topic, and in judgmental forecasting research which notes great progress (Cox Jr and Loomis, 2006; Lawrence et al., 2006; Tetlock and Gardner, 2016). However, higher education and practitioner training does not yet seem to follow this trend. Currently, only 50% of the top 50 US Business Programs provide forecasting courses (Kros and Rowe, 2016) and there is no great difference seen in the e-learning courses that are available, in spite of the variety of existing statistics courses (Gel et al., 2014). Even though human judgment has the potential to produce accurate forecasts, straightening the decision making process (Tetlock and Gardner, 2016), the teaching methods, which aim to improve learning outcomes regarding judgmental forecasting topic are not so popular. Concurrently, gamification strategies have been implemented in a variety of educational domains to e.g. popularize the learning of a foreign language (for example *Duolingo*) to teaching macroeconomics (Mendez-Carbajo, 2019). A lot of similar topics have been touched by gamification, even complex global challenges (Heinonen et al., 2017; McGonigal, 2011) but judgmental forecasting still remains an obscure area of study. Nooriafshar (2005) mentions the effective use of interactive technology in order to teach how to conduct judgmental forecasting methods. Critical action learning, without using games or gamification, has been investigated in the teaching of scenario analysis and exploring didactic pedagogy approaches (Bradfield et al., 2015). A prediction market setting has also been used as a pedagogical tool in the context of management courses, motivating students and disseminating forecasting aspects, but not judgment bias per se (Buckley and Doyle, 2016a; 2017; Buckley et al., 2011).

Regarding the learning and knowledge of judgment biases, not necessarily in a forecasting context, a few serious games have been designed and used for both improving knowledge and mitigating the cognitive biases. Some of them offer game design in a context of adaptive learning system (Jacobson and Dargue, 2019) or intelligence tutoring system (Veinott et al., 2013), suggesting a game environment to deal with biases or even with real-world learning challenges. Other studies have designed and experimentally used serious games such as *"Cycles Carnivale"* (McKernan et al., 2015), *"Heuristica"* (Veinott et al., 2013) and *"MACBETH"* (Dunbar et al., 2014) with a great sample of participants, in order to examine the effectiveness of games variables and components and/or compare to traditional teaching methods such as videos. In addition, a trading game has been efficiently used in financial education, teaching the avoidance of cognitive biases (Martelli, 2013). Despite the differences that these studies present in methodology and implications, their findings demonstrate the potency of game-based learning when compared to a traditional pedagogy, especially for convoluted aspects such as mitigating the natural tendency toward cognitive bias. Apart from the superiority of game-based learning compared to traditional methods (Dunbar et al., 2014), games development and design have been strongly suggested in order to raise awareness about cognitive biases and straighten the rational thinking for improving decision making (Dunbar et al., 2014; Jacobson and Dargue, 2019; Martelli, 2013; McKernan et al., 2015) even as a simple training intervention (Morewedge et al., 2015). However, there is a lack of simple gamified interventions that bring game experience in participants' learning about judgment biases, and gamification has been barely mentioned in this topic. So, on one hand there is a gap in effective teaching in the context of judgmental forecasting and judgment bias, and on the other hand, gamification effects have not been experimentally examined in this topic especially regarding student learning outcomes.

Therefore, we designed and developed a gamified web-based application which embodies an innovative gamified strategy, based on individual achievement, competence and storytelling. It aims to teach heuristics and bias as reported in the research of Tversky and Kahneman (1974), contributing to the multidisciplinary areas of gamification effects on education, teaching judgmental forecasting, and eventually improving rational thinking and decreasing forecasting errors.

3. Methods & data

3.1. Participants

We conducted our experiments in two different schools in two different universities. The participants were male and female, had different educational levels (bachelor undergraduate students - UG, MBA students), were students on different courses of their school, and were in different years of their studies. However, all of the students had forecasting courses within their syllabi. Precisely, we performed the following experiments:

- 204 students in forecasting courses at the ECE (class 2016-2019), in the 4th year of their studies. 20 of whom were MBA students at the same school (class of 2019) in the 2nd year of their studies. Students were given a participation incentive being available as a bonus 0.5 mark out of 10 in the course's final grade.
- 81 students on an information technology course at the Bus.Adm. class of 2018. Student participation was mandatory in the context of a laboratory exercise.

Despite differences in the students' education, none of the students had previously read the research of Tversky and Kahneman (1974), and none had already participated in forecasting courses or in a judgmental forecasting course.

3.2. Materials

All the materials were designed and implemented, in order to teach the main heuristics and biases that have great impact on human's rational thinking, affect judgmental forecasting skills, and eventually the decision making (Armstrong, 2001; Lawrence et al., 2006; Makridakis et al., 2008). Given the importance of judgmental forecasting skills at educational or business context (Cox Jr and Loomis, 2006; Kros and Rowe, 2016), being aware of cognitive biases and human's judgement limitations under uncertain conditions has been suggested as a way to improve judgmental forecasting accuracy and the decision making process (Lawrence et al., 2006) for both laymen and experts alike. Given the magnitude of the cognitive biases fields and the limited time of our experiments, we chose to focus only on one but critical research of Tversky and Kahneman (1974). Despite the fact that this research has been conducted almost 50 years ago, it still counts thousands of citations, it composes the structure of game-based learning approaches, it is cited by forecasting textbooks, and it vividly presents biases that affect forecasting process (Tversky and Kahneman, 1974). Hence the research of Tversky and Kahneman (1974) is the basis of the conducted lecture, the material for the traditional reading task, the scenario of the gamified application and finally the content behind the online 30-questions test, as it follows.

3.2.1. Lecture

The purpose of the lecture was to offer an overview of Tversky and Kahneman (1974) research and deliver detailed information regarding the three heuristics that lead to great biases in assessing probabilities of future events or judgmentally forecasting values. During the lecture the main categories and subcategories of the heuristics were mentioned describing at least one example for each of them, based on the research of Tversky and Kahneman (1974). Finally, the theoretical implications of this study were discussed along with their importance in forecasting process, education and decision making. For these lecture's needs, we created and used a visual material, which was composed of 22 slides and lasted 15 minutes.

3.2.2. Read the research of Tversky and Kahneman (1974)

Students should read the following research: "Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.", using a computer in the computer lab. The above mentioned research is 8 pages and constitutes the underpinnings of the lecture, the gamified application and the 30-questions test.

3.2.3. Use of the gamified application

Since gamification has only been used to a limited degree in forecasting courses, we designed and developed a gamified application from scratch, which aims to communicate the heuristics and biases that have a great impact on human judgment (Tversky and Kahneman, 1974). The gamified application was designed following software design principles (Barnett et al., 2005) and gamification design principles (Kapp, 2013; Morschheuser et al., 2017b; Zichermann and Cunningham, 2011). The content of the application was based on the research of Tversky and Kahneman (1974).

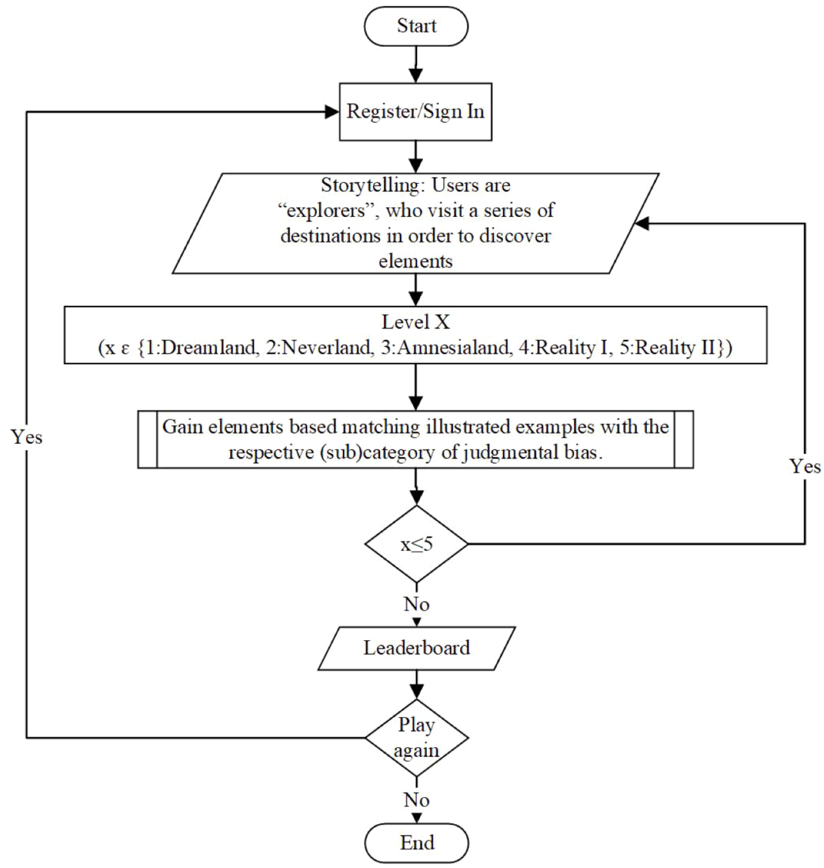

Fig. 1. The flow of the gamified application: *JudgeIt*

Architecture of the gamified application. The gamified application is named *JudgeIt* in order to link with its content regarding the biases on judgment that it aims to teach. Special attention has been paid to the flexibility, accessibility and interoperability of the application. Therefore, *JudgeIt* is a web-based application, publicly available, and built on . NET framework using an MS -SQL Server database. Users need to be registered using their email and a password of their choice, and later signed in to use it. The users progress is saved in the database based on their credentials, in order to use it as frequently as they wish.

Design of the gamified application. The design of *JudgeIt* was based on guidelines about the use of motivational affordances in educational applications (Deterding et al., 2011; Dicheva et al., 2015; Maican et al., 2016; Nah et al., 2014; Pedreira et al., 2015; da Rocha Seixas et al., 2016). The most commonly used motivational affordances in education are points, badges and leaderboards (Koivisto and Hamari, 2019; Majuri et al., 2018), which are affiliated with achievement-based gamification in terms of creating feelings of progressing and achievement. In our study, we integrated competition and progression feeling into the gamified application by using points, levels, challenges and a leaderboard. We further used storytelling as a narrative context behind the

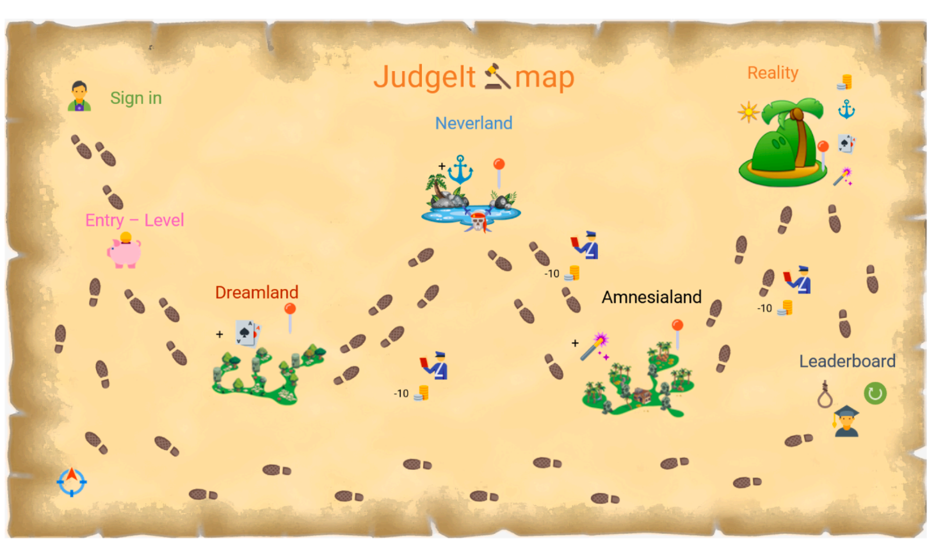

Fig. 2. *JudgeIt*: the map of a full game round.

Table 1
Motivational affordances in the context of *JudgeIt*.

Affordances	Representation in *JudgeIt*	Purpose of Use
Points	Credits to travel,cards, magic wands, anchors.	▪ A numeric reward for identifying the represented bias in videos. ▵ Matching illustrating examples with the subcategories of judgmental biases.
Levels	Entry level,"Dreamland" *Representativeness*, "Amnesialand" *Availability*,"Neverland" *Adjustment & Anchoring*.	▪ Indicator of progression and difficulty. ▵ Recognizing examples of judgmental biases' categories.
Challenges	Reality I & II*(2 levels)*.	▪ Motivation for improvement. ▵ Recognizing examples of all the categories of judgmental biases.
Storytelling	The narrative behind the representations.	▪ Illustrate the applicability of the examples. ▵Memorising easier the categories of judgmental biases.
Leaderboard	Function that combines the collected points and elements.	▪ Direct comparison of players' performance to increase the competition.

▪ contribution to the gamified application, ▵ educational contribution
Source: (Buckley and Doyle, 2017; Bunchball, 2010; Kapp, 2013; Maican et al., 2016; Nah et al., 2014; Seaborn and Fels, 2015a; Zichermann and Cunningham, 2011).

content of the application, covering the whole range of the gamification types (Xi and Hamari, 2019).

We chose these affordances because of the nature of the research of Tversky and Kahneman (1974). Since a lot of information is described, levels help to schematically categorize the information. Points and challenges give the feeling of reward and progress, and storytelling helps students to memorize the concept (Stewart, 2012). In the application, by making users part of a meaningful story, users become "explorers" and aspire to visit a series of destinations in order to discover elements linked with heuristics and biases. Fantasy is nicely interwoven with simple challenges per level, intriguing users' interests (Poncin et al., 2017). Fig. 1 illustrates the flow of a full round of the gamified application, which is composed of 5 levels.

Registered users are informed about the story and the challenges that they will have to deal with. They are informed that they are "explorers" and they are challenged to discover and collect elements in a series of imaginary destinations. Fig. 2 shows the names of these destinations, by illustrating the map of the gamified application. Each of them represents the heuristics, presented in Tversky and Kahneman (1974)'s research, according to Table 1. More precisely, students journey starts from

Fig. 3. *JudgeIt*: Screen of the "Dreamland" level.

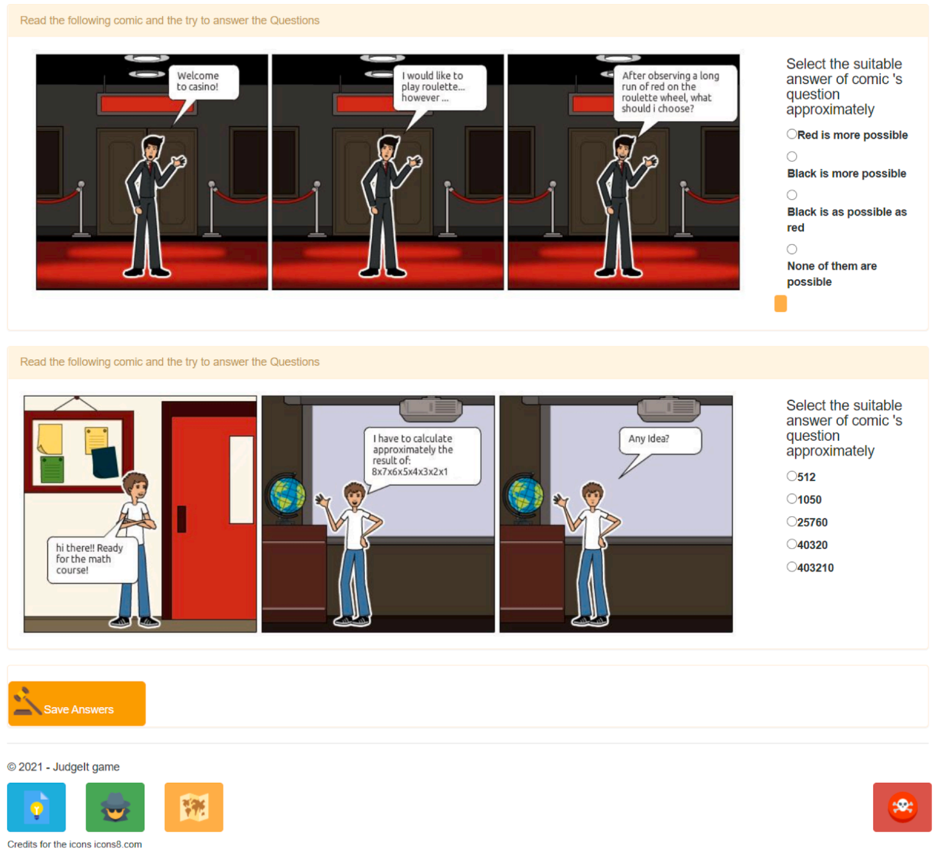

Fig. 4. *JudgeIt*: Screen of the "Reality" level.

"Dreamland", which represents the "Representativeness" bias and they need to answer correctly two multiple-choice questions, by matching the illustrated examples with the respective subcategories of judgmental biases. Fig. 3 depicts a screenshot of this level, showing on the top of the page users' points and different elements, and then below the videos and the questions that users need to answer in order to save their choices and proceed to the next destination/level. The screenshot is divided into rectangles with appropriate messages only to explain the different game elements. The same user interface applies to the next levels later, named: "Neverland" and "Amnesialand" which represent "Availability" and "Adjustment and Anchoring" respectively. However, for each destination/level, there is a different pair of videos and questions regarding the represented biases and heuristics (see Fig. B.1 and Fig. B.2, at the Appendix). Finally, students reach the last level, named "Reality" where they need to recognize the hidden judgement biases, through a short story illustrated by some comics stripes. Fig. 4 depicts two of these comic stripes at "Reality" level.

Overall, students collect different kinds of elements in each imaginary place that they visit, by identifying the respective biases that occur in each place. Useful videos and pictures illustrate the examples presented in the research of Tversky and Kahneman (1974), in order to challenge and guide the explorers through these destinations. Users gain points and collect elements by identifying biases categories based on the videos, targeting the user to achieve a high rank on the leaderboard, so competing with their peers. Table 1 presents the embodied motivational affordances, their representation in the application and the purpose these affordances serve. In this direction, apart from the game experience that *JudgeIt* offers, it provides a user-friendly interface and easy workflow, with clear guidance for the users (Kapp, 2013).

3.2.4. *The online 30-questions test*

An online 30-questions test was used to measure students' learning performance regarding their knowledge of the main categories of heuristics and biases presented in the research of Tversky and Kahneman (1974), their ability to categorize examples in these categories and subcategories and the meaning of some basic concepts about judgement such as: recency effect, inconsistency and regression effect. The online test was composed of 30 closed-ended question format, such as multiple-choice and true-false. Participants provided mostly one answer per question with respect to predetermined multiple response options or true-false (see Appendix Table A.1). All of the questions were of equivalent grade and were based on information provided in the lecture (see 3.2.1), hence all participants were able to achieve the highest score in this test by only attending the lecture. The test lasted for a maximum of 15 minutes and it was the last part of the experiment for all participants.

3.3. *Design*

The assessment of the gamification impact on student learning outcomes regarding the heuristics and biases presented by Tversky and Kahneman (1974) was done through a series of controlled experiments

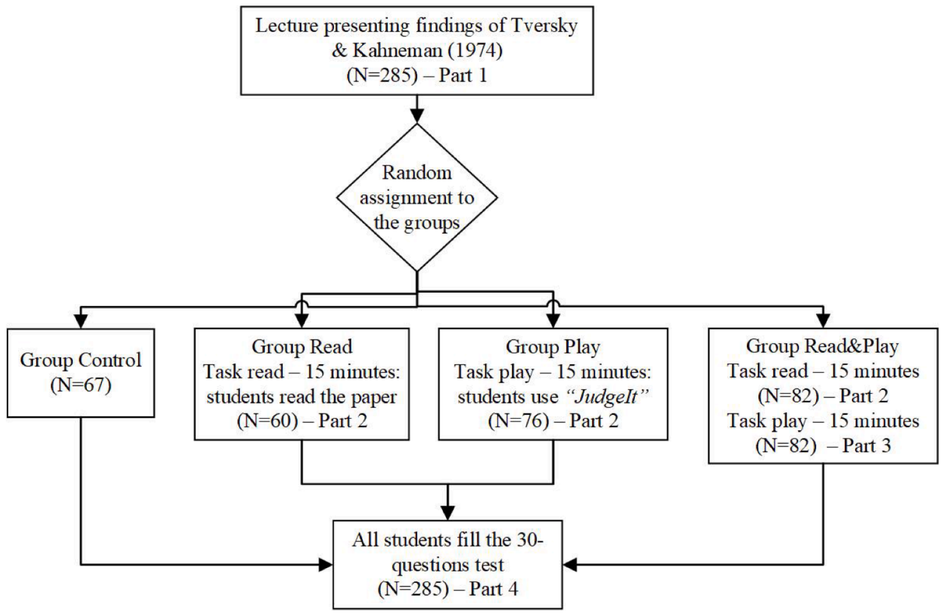

Fig. 5. Design of the experiment

in two different schools, following exactly the same setup. More accurately, we conducted a 2 × 2 between subjects experiment, having as treatments: read: yes vs no and gamification: yes vs no. Participants were randomly assigned to one of the following groups: i) Group Control: no treatment, ii) Group Read: treatment of reading a research paper (see 3.2.2), iii) Group Play: treatment of using the gamification strategy (e.g.points, levels, challenges, storytelling and leaderboard) (see 3.2.3) and iv) Group Read&Play: both treatments. Time was controlled and was equal to 15 minutes for each task: reading or using the gamification strategy. Fig. 5 illustrates the experimental design along with the respective treatment and the sample size per group. The design of the experiment composed of the following parts:

Part 1 All students attended the lecture (see 3.2.1) about the heuristics and biases presented in the research of Tversky and Kahneman (1974), strictly lasting for 15 minutes.

Having attended the lecture, students were randomly assigned to one of the groups: Group Control, Group Read, Group Play and Group Read&Play. Each group receives different treatment. Students who compose each group have different tasks to complete as described in Part 2 and Part 3.

Part 2 The Group Read and the Group Read&Play had to read the research of Tversky and Kahneman (1974) in electronic format for 15 minutes (hereafter this task is described as *task read*). Group Play had to use *JudgeIt* and complete a full game round for 15 minutes (hereafter this task is described as *task play*).

Part 3 The Group Read&Play had to complete the *task play* for 15 minutes.

Part 4 All students from all groups had to complete the online 30-questions test (see 3.2.4). The test lasted for a maximum of 15 minutes as well.

3.4. Procedure

Students were informed beforehand that they could participate in an experiment, instead of attending a regular lecture of the respective course. At the ECE, participants could voluntarily participate. Although there was an incentive for students' participation (a bonus of 0.5/10 in the courses final grade), they were informed that the final exams would contain an equivalent exercise. So all students could reach the highest grade of the course, taking part or not in the experiment. At Bus.Adm. the undergraduate students should participate in the experiment, which was a required part of the course with no further incentive.

At both schools, when participants arrived at the computer lab, their informed consent was received and they were guided to choose a computer station. After that, the experimenter informed participants about the experiment set up, e.g. that participants should attend a lecture, then they would be randomly assigned to one of the four groups and all participants should correctly answer an online 30-questions test based on the lecture's content. The experimenter encouraged participants to complete as correctly as possible the test, independently of their group, even though their score would not affect their final course grade. Finally, the time limits for each activity were mentioned.

The experiment starts with the 15-minutes lecture (Part 1), presenting the heuristics and biases in human judgement (Tversky and Kahneman, 1974). Then, participants were randomly divided into the groups: Group Control, Group Read, Group Play and Group Read&Play. Groups received different treatments, as their names imply. Group Control did not receive any treatment, Group Read and Group Read&Play had to complete the *task read* for 15 minutes (see 3.2.2) and Group Play had to complete a full game round in the gamified application (see 3.2.3), which was the Part 2 of the experiment. Later, only Group Read&Play had 15 minutes to complete the *task play*. Finally all groups completed the online 30-questions test for 15 minutes respectively, which composed the Part 4 of the experiment.

Important notes about the experiments design are that Part 1 and Part 4 were exactly the same for all participants. The duration of the treatments was different, but the duration of each task -*task read, task play*- was equal to 15 minutes. For example, Group Read&Play had 30 minutes available to complete the *tasks* of *read* and *play*, while Group Play had 15 minutes available to complete the *task play*, etc. However, this study examines the impact of gamification based on an assessment of the final 30-questions test, which was the same for all groups. The students were randomly assigned to one of the groups: Group Control, Group Read, Group Play, and Group Read&Play, and were not allowed to collaborate or search for information on the internet. All students received the same incentives provided for their different school, independent of the group that they were assigned to. The design of the experiment was exactly the same for all classes and for both schools and the experiments were organized and conducted by the same researcher.

4. Results

This study aims to investigate the impact of the gamification strategy which is integrated in *JudgeIt* on students learning outcomes. The impact of this gamification strategy is based on an assessment of the 30-questions test (Part 4), regarding the students comprehension of heuristics and the biases of human judgment. The performance for each student is equal to the sum of the right answers in the test, normalized to a maximum of 100.

Some student demographic characteristics were examined, such as the students gender (male, female), their school (ECE or Bus. Adm.) and their educational level (undergraduate students, hereafter UG or MBA students) along with the impacts of tasks read and/or play. The statistical analysis of the results was conducted on three aggregated levels: i) Effects of variables read and play on student performance ii) Effects of variables read and play and demographic differences on student performance, and iii) Traditional versus gamified teaching methods.

4.1. The effects of variables read and play on student performance

The aim of this part of the analysis is to identify statistically significant differences in the mean values of the students performances regarding the different treatments: reading the research, play with gamified application, and the combination of these two along with no treatment. Fig. 6 illustrates the students performances for each group in percentiles with boxplot diagrams, and Table 2 presents the descriptive statistics regarding the treatments. Group Play, which only experienced the gamified application noted the highest mean performance and also the highest standard deviation in the results. Group Read&Play has the second highest mean value of performances and standard deviation. Group Control follows regarding both mean value and standard deviation. The lowest mean value of performance with the lowest value of standard deviation was achieved by Group Read.

An analysis of variance (one-way ANOVA) was conducted to compare the effects of treatments on students performances, using the square root of the normalized values of performances in order to meet the assumption about normality and homogeneity of variances. The null hypothesis: H_0: *Equal means in all groups (4.1.1)* is rejected ($F(3, 281)=3.37$, $p=0.019$), so we can claim significant differences among the groups. In order to further examine the differences, the 2 ×2 between-subjects factorial design of our experiment (treatments: i) read: yes vs no, and ii) gamification: yes vs no), allows us to depict each factor (treatment) with a Boolean variable in order to both examine the impact of each factor and their interaction. Their interaction represents the combination of *tasks read* and *play*, which required double the time of each factor. Hence, we used a variable named play, which is equal to 1 if the respective group completed the *task play* (n=158), and 0 if otherwise (n=127). We used a variable named read, respectively, which is equal to 1 if the group completed the *task read* (n=142), and 0 if otherwise (n=143). Given this allocation, we conducted a multi-factorial analysis of variance (two-way ANOVA) in order to examine the effects of these variables. The null hypothesis for this test is: H_0: *There is no difference in group means of any level of the variable read, or the variable play, and there is no interaction effect between these variables (4.1.2)*. The results show that only the *task play* has a significant impact on student performance ($F(1, 281)=9.10$, $p=0.003$). The *task read* ($F(1, 281)=1.21$, $p=0.272$) and interaction between *tasks read* and *play* are not significant ($F(1, 281)=0.08$, $p=0.784$).

In terms of further analysis, we conducted comparisons, using Tukey HSD test to identify the significant differences between the different variables, so eventually between the groups (confidence interval equal

Table 2
Descriptive statistics per group (treatment).

Groups (N=285)	n	%	M	SD
Group Control	67	23.51	41.34	13.63
Group Read	60	21.05	38.89	13.07
Group Play	76	26.67	47.19	18.40
Group Read&Play	82	28.77	45.28	16.30

Table 3
Pairwise multiple comparisons among the groups based on Tukey HSD test with Cohen's d effect size.

Variables: *read*(0.1) × *play*(0.1)			P.adj (Tukey's HSD)	Cohen's d estimate
read=0, play=0 (Group Control)	vs.	read=1, play=0 (Group Read)	0.784	0.19 (negligible)
read=0, play=0 (Group Control)	vs.	read=0, play=1 (Group Play)	0.208	-0.32 (small)
read=0, play=0 (Group Control)	vs.	read=1, play=1 (Group Read&Play)	0.517	-0.23 (small)
read=0, play=1 (Group Play)	vs.	read=1, play=0 (Group Read)	0.023*	0.48 (medium)
read=0, play=1 (Group Play)	vs.	read=1, play=1 (Group Read&Play)	0.920	-0.09 (negligible)
read=1, play=1 (Group Read&Play)	vs.	read=1, play=0 (Group Read)	0.097	0.40 (small)
*p <0.05,		**p <0.01,	***p <0.001	

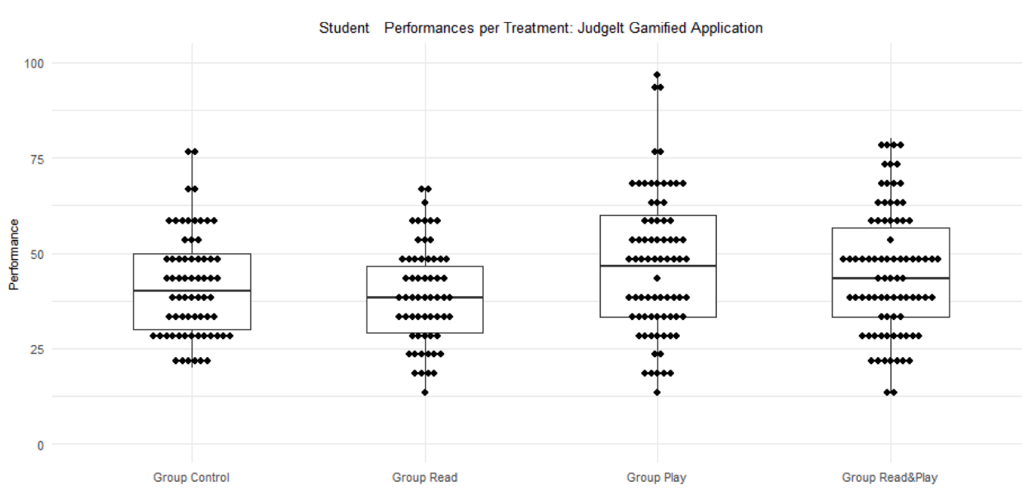

Fig. 6. Students' performances per treatments

to 95%). Table 3 presents the results of the analysis along with the respective effect size of the different tasks, based on Cohen's d estimate. The comparisons indicate that only the mean performance of Group Play (read=0 and play=1) (*M=47.19, SD=18.4*) was significantly different than the mean performance of Group Read (read=0 and play=1) (*M=38.89, SD=13.07*). In addition, based on the effect sizes for this analysis, according to Cohen (1988) convention, small differences were detected in performances between all pairs except for the following: Group Control (read=0 and play=0) versus Group Read (read=1 and play=0), and Group Play (read=0 and play=1) versus Group Read&Play (read=1 and play=1). Thus, Group Play (read=0 and play=1), and more precisely the treatment of only using the gamified application seems to result in statistically significant differences regarding traditional teaching methods.

4.2. The effects of variables read and play and demographic differences on student performance

Apart from the impact of the different tasks *read, play*, and their interaction, the demographic differences of our sample were considered in the analysis. Fig. 7 depicts the distribution of students performances regarding their demographic differences of gender, school and educational level. Table 4 presents descriptive statistics about these variables.

In order to investigate the impact of the different variables (i.e. reading a paper, using the proposed gamified application and doing both), along with students' demographic characteristics, we also used the variables read and play (see 4.1). A three-way ANOVA was conducted on a sample of 285 participants to examine the effect of tasks read, play and one of the variables per time: gender, school and educational level and their interactions on students performances. More precisely, considering one of the demographic variables per time, we examined the main effects of variables read, play and demographic X, and the effects of the two-factor interactions read x play, read x demographic X, and play x demographic X. There is also a three-factor interaction, read x play x demographic X, which was considered. For each of the seven cases the null hypothesis is the same: there is no difference in means. We conducted the test as many times as the number of the demographic variables, having the following null hypotheses:

$H_{0\ \text{read, play, demographic } X}$: *There is no difference in means of students' performances of any level of the variable read, or the variable play, or the variable X and there is no interaction effect among all the possible combinations of the three variables(4.2.X).*

Considering the examination of the demographic variables in this study, the null hypotheses tested are as follows and Table 5 presents the results:

- $H_{0\ \text{read, play, gender}}$ *(4.2.1)*
- $H_{0\ \text{read, play, school}}$ *(4.2.2)*
- $H_{0\ \text{read, play, educational level}}$ *(4.2.3)*

For example, regarding the gender variable, we tested the following null hypothesis: $H_{0\ \text{read, play, gender}}$: *There is no difference in the means of students performances of any level of the variable read, or the variable play, or the variable gender, there is no interaction effect between read × gender, play × gender, read × play, or read × play × gender (4.2.1)*. The same tests were separately conducted, examining the effects of the variables read

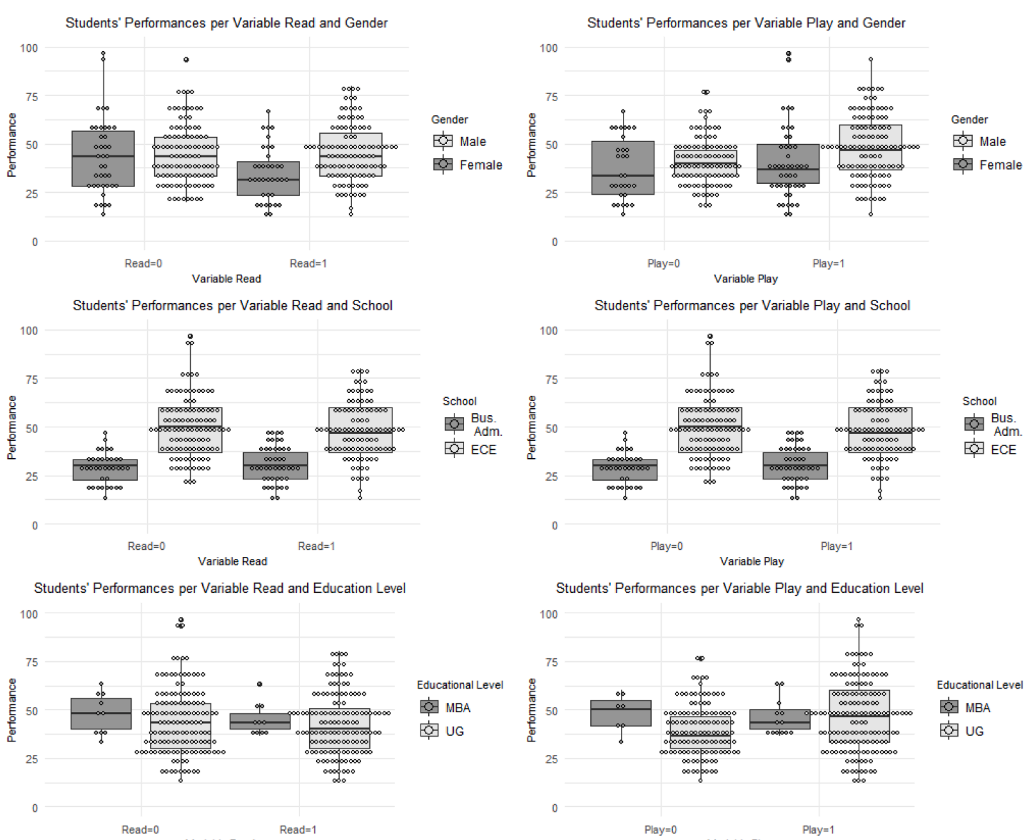

Fig. 7. Students' performances per variable read, play and demographic differences.

Table 4
Students' performances per variable: (read and play), gender, school and educational level.

Variable	Read=0			Read=1			Play=0			Play=1		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD
Gender												
Female	39	43.4	19.6	36	34.4	13.7	30	37.8	15.8	45	40	18.7
Male	104	44.8	15.3	106	45.3	14.9	97	40.9	12.5	113	48.7	16.2
School												
ECE	107	49.7	15.4	97	48	14.5	90	44.4	12.7	114	52.4	15.7
Bus. Adm.	36	28.8	7.73	45	31	9.36	37	29.9	8.7	44	30.1	8.79
Educational Level												
UG	133	44.2	16.9	132	42.4	15.7	120	39.7	13.4	145	46.2	17.9
MBA	10	48	10.3	10	45.3	8.04	7	48.1	9.59	13	45.9	9.14

Table 5
Impact of the variables: read, play, demographic characteristics and their interactions.

Task or interaction	X_1= Gender (4.2.1)			X_2=School (4.2.2)			X_3=Educational Level (4.2.3)		
	df	F	p	df	F	p	df	F	p
read	1	1.35	0.246	1	0.28	0.598	1	1.16	0.282
play	1	11.7	<0.001***	1	12.02	<0.001***	1	8.69	0.003**
X_i, i ϵ {1,2,3}	1	12.40	<0.001***	1	128.68	<0.001***	1	1.12	0.290
read × *play*	1	0.31	0.579	1	0.13	0.716	1	0.06	0.802
read × X_i	1	5.63	0.018*	1	1.70	0.193	1	0.00	0.970
play × X_i	1	0.82	0.366	1	4.75	0.030*	1	1.05	0.307
read × *play* × X_i	1	3.04	0.082	1	0.03	0.872	1	0.01	0.912
*p <0.05,		**p <0.01,	***p <0.001						

Table 6
Pairwise post hoc comparisons of variable play and school using the Tukey HSD test with Cohen's d effect size.

Groups			P.adj (Tukey HSD)	Cohen's d estimate
$ECE_{play=1}$	vs.	$ECE_{play=0}$	<0.001***	0.54 (medium)
$ECE_{play=1}$	vs.	$Bus.Adm_{play=0}$	<0.001***	1.68 (large)
$ECE_{play=1}$	vs.	$Bus.Adm_{play=1}$	<0.001***	1.67 (large)
$ECE_{play=0}$	vs.	$Bus.Adm_{play=0}$	<0.001***	1.28 (large)
$ECE_{play=0}$	vs.	$Bus.Adm_{play=1}$	<0.001***	1.27 (large)
$Bus.Adm_{play=0}$	vs.	$Bus.Adm_{play=1}$	0.999	0.01 (negligible)
*p <0.05,		**p <0.01,	***p <0.001	

and play along with the students school, and educational level and their interactions. Table 5 presents the results in greater detail. Regarding the main independent variables of our study (i.e. read, play), only the *task play* seems to be a statistically significant factor regarding students performances, which is in accordance with the results of the one-way ANOVA and the post hoc comparisons (see 4.1). Students gender and school along with the task *read* and *play* respectively, appear to have statistically important interactions effect on students performances, as well. Based on the results of the three-way ANOVA, students gender ($F(1, 277)=12.40, p<0.001$) and school ($F(1, 277)=128.68, p<0.001$) are statistically important factors regarding students performance, but only the interaction of *read* × *gender* ($F(1, 277)=5.63$, $p=0.018$) and the interaction of *play* × *school* ($F(1, 277)=4.75$, $p=0.030$) resulted in statistically important differences. Since this study aims to examine the effect of the proposed gamified strategy on students learning outcomes, we will further investigate only the interactions related to the *task play*. However, we should mention that the differences in sample sizes between males and females, and particularly the difference in distribution of males and females between the two different schools (ECE: $n_{females}$=30, n_{males}=174, Bus. Adm.: $n_{females}$=45, n_{males}=36) along with the differences in prior knowledge might impact on the above results and we cannot draw any statistically significant result regarding the impact of gender along with the variables read or play.

Therefore, we further investigated the significant differences only between the different schools (i.e. ECE, Bus.Adm.), and the *task play*, by conducting pairwise comparisons. Table 6 presents the results, showing statistically significant differences in the majority of the comparisons, according to post hoc Tukey HSD. The only comparison which did not result in statistically significant differences in means of students' performances is between the groups of Bus.Adm., who used the gamified

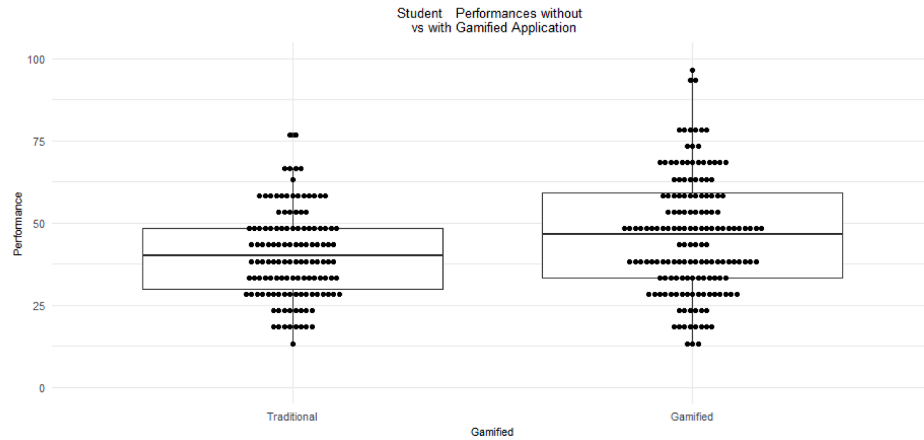

Fig. 8. Students' performances per traditional teaching and gamified groups.

application and those who did not. The effect size for this comparison was found negligible, showing that the effect of task play on students' performances on Bus.Adm. is not statistically significant. While the prior knowledge of students might weaken the comparison between the different schools regarding the effect of *play*, we can support that the use of gamification appears to positively impact on students performances mainly in the engineering school, resulting in a medium effect size (d=0.54).

4.3. *Traditional versus gamified teaching methods*

As a final step of our analysis, we rearranged students performances and created two different groups named: "traditional" and "gamified". The group traditional is composed of 127 students, who did not experience the gamified strategy, including student performances from Group Control and Group Read. The group "gamified" is composed of 158 students who used the *JudgeIt* application, including student performances from Group Play and Group Read&Play. Fig. 8 illustrates the distribution of students performances divided into these new groups, using boxplot diagrams. An independent-samples t-test was conducted to compare students performances in traditional and gamified conditions. The participants who used the gamified application (*M=46.20, SD=17.31*) compared to the students who did not (*M=40.18, SD=3.37*) demonstrated significantly better performances regarding the knowledge about judgment heuristics, $t(282) = 3.04$, $p = 0.003$, and based on a Cohen's d estimator (*d= 0.35 (small)*), there is a small improvement regarding the mean values of performances, equal to 14.98%.

5. Discussion and implications

Overall, based on the conducted statistical analysis, the gamified strategy composed of points, levels, challenges, storytelling and leaderboard improves the learning outcomes regarding the heuristics and bias in human judgment described in the research of Tversky and Kahneman (1974). Thus, the results of this study are in favor of the use of gamification in the lecture setting of a forecasting course, under the described conditions but probably the results also represent positive potential of the use of the approach in education of other topics as well. On a large scale, our findings are in accordance with the game-based learning literature regarding the effectiveness of game-based approaches on improving learning related to complex cognitive tasks such as decision making (McGonigal, 2011; McKernan et al., 2015). However, a few interesting topics are raised, based on the analysis of the students performance in regard to the treatments and their demographic characteristics.

Primarily, the results indicate that the applied gamification approach yielded better learning outcomes compared with both the control (i.e. only lecture) and the reading approach (i.e. lecture and reading) as the groups that had the gamification-based pedagogical approach (Group Play and Group Read&Play) had the highest student performances. Therefore, our results are in general in accordance with the bulk of positive results related to gamification in education seen in the literature (Albritton et al., 2003; Koivisto and Hamari, 2019; Majuri et al., 2018; Nah et al., 2014; da Rocha Seixas et al., 2016; Seaborn and Fels, 2015a), as well as more particularly in the context of a forecasting education (Albritton and McMullen, 2006; Craighead, 2004; Gardner, 2008; Gavirneni, 2008; Kroes et al., 2013; Snider and Eliasson, 2013). Moreover, these results are in line with the positive effects of game-based approaches on learning of cognitive biases reported by Dunbar et al. (2014); McKernan et al. (2015); Morewedge et al. (2015).

A further interesting result is that the activity of reading in our experiment did not affect learning performance neither as a sole activity after the lecture nor when coupled with gamification despite that under the latter condition the participants spend 15 minutes more on the overall learning task. These results may be a preliminary indication that passive reading may not be so effective in students comprehension about judgmental heuristics, and eventually in their performance in a lecture setting about a forecasting topic. Passive learning, such as just attending a lecture or reading an article or even slides may lead students to become bored, demotivated and eventually even minimize students comprehension around the topic (Mann and Robinson, 2009; Ryan, 2006), and according to our findings, seems to apply to judgmental forecasting courses, as well. On the other hand, when a lecture with slides is combined with a more constructive interactive environment, such as a game, it may be more effective (Clark, 2008; Mann and Robinson, 2009). Considering the experimental settings featured in this study, the additional time dedicated in the combination of tasks *read* and *play* may have a net detrimental effect (Mustafa et al., 2014) especially when considering the opportunity cost of increased time investment. In addition, since game-based learning approaches have been shown more effective and beneficial compared to traditional teaching methods such as simple giving instructions or watching a video (Dunbar et al., 2014; Morewedge et al., 2015), our findings directly indicate the potential of the use of game-based learning approaches in the context of cognitive biases in human judgement in a forecasting course and imply positive expectation of its potential also in other similar environments.

Examining the effects of gamification along with students demographic characteristics, a couple of interesting findings emerge. Initially, the treatment *play* (i.e. use of gamified application) affected students performance along with its interaction with students school. Our results support that engineering students might benefit more from integrating gamification in the educational process compared to business school students. This finding is in line with respective literature about the positive impact of gamification in computer science education (Dicheva et al., 2015; Pedreira et al., 2015) and math education. An explanation might be that the problem solving usually associated with engineering is regarded similar with the problem solving in games. So, gamification may be more attuned to the aptitude of engineering students on average (Markopoulos et al., 2015; Pedreira et al., 2015). In this direction, based on our analysis, the high mean performance of students who experienced both traditional and gamified strategies is encouraging in order to further explore gamification potential in business education, apart from the business simulation and business games initiatives that are currently popular (Faria et al., 2009). Additionally, the differences in samples sizes and incentives between the schools in our study should be considered inspiring for further research about gamification impact on business education. Regarding the educational level, the small sample of MBA students (n=20) might have an influence on the validity of the results, since its null impact is not in line with previous studies regarding users perception of gamification (Conaway and Garay, 2014; Jia et al., 2016; Martí-Parreño et al., 2016). However, despite the fact that more empirical data is needed in order to draw conclusions, we argue that the used gamified strategy notes a positive impact on both undergraduate and MBA students, and for both schools as a positive in-lecture intervention.

The conducted statistical analysis indicates that students gender is an important factor for the differences in students' performances regarding the received treatments. Based on Fig. 7 and Table 4, however, pertinently to the treatment of interest, there was an interaction effect between the reading treatment and male gender . The effect may be affected by the diminishing sample size when the granularity of investigation is increased to separately consider genders and schools. Overall, considering that this study focuses on examining the impact of gamification on students' learning outcomes, no important statistical conclusions can be drawn regarding the gender impact along with gamification on students performances. Nevertheless, since females and males have been found to slighlty differ in terms of effects of gamification in prior literature (Carr, 2005; Koivisto and Hamari, 2014) and to differently perceive and accept games (Liu, 2016), further research regarding the impact of gender in combination with gamified strategies is suggested.

5.1. Implications

The theoretical contribution of this study is twofold. It answers the need for more empirical data through controlled experimental research (as indicated e.g. by the reviews of the area (Koivisto and Hamari, 2019; Landers et al., 2018; Majuri et al., 2018; Nacke and Deterding, 2017; Seaborn and Fels, 2015b)) by conducting a 2×2 between-subjects experiment (treatments: i) read: yes vs no, and ii) gamification: yes vs no) and it supports the potential of a gamified strategy in a forecasting course context. Secondly, this study suggests a complementary educational tool that improves students learning outcomes. This new approach, including gamified strategies reflects an active and experiential learning approach (Heinonen et al., 2017), which can be used inside and outside of a classroom, addressing the need to update teaching methods in forecasting courses (Gel et al., 2014; Loomis and Cox, 2000).

Apart from the theoretical contribution, this study also suggests practical implications to gamification designers, educators, managers or corporate trainers. The proposed gamified strategy (i.e. points, levels, challenges, storytelling, and a leaderboard) noted improvements regarding learning outcomes about heuristics that affect judgmental forecasting skill compared to traditional teaching methods. Given the importance of avoiding judgment biases in a plethora of topics such as judgmental forecasting, scenario analysis, decision making, along with their practical applications in business environments, gamified strategies, under specified circumstances seem effective and efficient in terms of time and results. Thus, short active gamified exercises in business training, could motivate and educate both employees and experts independently on their educational level and engineers might benefited even more, via an enjoyable way. Being aware of the judgment biases that irrationally affect humans' decisions, may improve judgmental forecasting and adjustments, and eventually enhance decision making and planning processes. In addition, using a gamified application, with points, levels, challenges, storytelling, and a leaderboard, brings a gaming experience to users, strengthening their creativity (Barata et al., 2013; Salen et al., 2006). This process challenges conventional thinking, and supports the subjective interpretation of data, phenomena or events, which can benefit both technological forecasting and firms' future (Makridakis et al., 2008).

Apart from the advantages in the business environment regarding forecasting, each individual and eventually society can be benefited by being aware of judgment biases. Uncertainty is one of the biggest causes of grief in human lives and has economical consequences and severe repercussions that stem from poor decision making. Being uncertain and feeling uncertain has negative psychophysiological and emotional consequences. Because of this tension towards the future for battling uncertainty, humans make forecasts even unconsciously during their ordinary life in order to make decisions, which eventually have great impact in a variety of societal aspects such as climate change, increased consumption, environmental pollution etc. Regarding these problems of contemporary society, McGonigal (2011) suggests to take advantage of the experience that games bring to players, such as creativity, optimism and emotional activation in order to deal with these challenges. The results of this study support the use of gamification in education of forecasting and cognitive biases, in order to raise awareness of judgmental bias, get insight of the existing data and events and at the same time think outside-of-the-box and be prepared for the forthcoming social changes.

5.2. Limitations

Some limitations about this study should be acknowledged. Students performance on the 30-questions test (part 4) should not be confused with the students score on the gamified application, the results of which fall outside the scope of this research. Student incentives regarding their participation in the experiment were different per school. However, since the incentives were the same for all groups per school, this fact does not seem to affect the validity of the results. Another limitation of this study is the small sample size of some groups which makes the interaction test less reliable (compared to the investigation of the main effects for which the sample sizes are more sufficient), in case we consider more than 2 independent variables such as prior knowledge *tasks read, play* and gender. Since we put emphasis on accurately investigating the interaction of controlled variables, we also need to use small groups where it is more difficult to ensure normal distribution.

All of the instructional materials across the conditions used in our research were designed according to the research of Tversky and Kahneman (1974). Although this research has been conducted almost 50 years ago, it is still a critical research of the field of cognitive biases and heuristics which have great impact on affecting judgmental forecasting. We acknowledge that the time span of 15 minutes for the reading task might not be adequate to read such an in-depth piece of research. However, all students had already attended a lecture regarding the research topic and we strictly kept the same conditions to all groups, so this fact does not endanger the validity of our results. Another limitation of this study is the difference in years of study per different school. Students at ECE were in their 4th year (out of 5) of their studies, so close to graduation. However, the students at Bus.Adm. were in the 1st year (out of 4) of their studies. This fact could explain the differences in students' performances and probably the effects of interactions: *read* × *school*, but it does not influence the vigor of the findings because this study also examined the comparison between treatments in the same schools. In this regard, the small sample of MBA students may in fact weaken the comparison between the student performances regarding the treatments and their educational level, so further research is needed in order to straighten the results.

5.3. Future research

Based on these research results, more gamified applications linked with different gamified strategies in the context of a forecasting course would be an interesting research topic, especially when considering both learner outcomes and forecasting accuracy. The impact of the presented gamified strategy on business forecasters regarding their forecasting accuracy would be an interesting addendum, as well. Since gamification has been effectively used to increase crowdsourcing participation (Morschheuser et al., 2017a), the combination of judgmental forecasting methods such as Delphi (Rowe and Wright, 1999) and the *"Good Judgment Project"* (Tetlock and Gardner, 2016) with the gamified strategies on social and political conflict situations would be a direction for further research. Thus, the investigation of gamification effects on judgmental forecasting accuracy would make a general contribution to a multidisciplinary research field.

6. Conclusions

To conclude, the results of our study argue for the use of this gamification strategy (i.e. points, levels, challenges, storytelling, and a leaderboard) as a complementary tool in a lecture on judgmental forecasting. More precisely, using this gamified strategy may improve learning outcomes up to 14.98%. Even the simple use of this application after the lecture on a judgmental forecasting topic may result in a statistically significant improvement in learning outcomes. Students' school seem to be an important factor regarding the perceived benefit of the gamification experience. Although gamification appears to be effective in terms of motivation in a forecasting course context, considering the highest values of standard deviation of students' performances in the gamified groups, there is a need for cautious design regarding both pedagogical and gamification design principles (Buckley and Doyle, 2017; Dichev and Dicheva, 2017; Hanus and Fox, 2015; Koivisto and Hamari, 2019), and also for more controlled experimental research to be carried out (Dicheva et al., 2015; Koivisto and Hamari, 2019; Landers et al., 2018; Seaborn and Fels, 2015a).

Overall, our study opts for the use of gamification in education regarding judgmental bias strongly related to judgmental forecasting and decision making, but also underlines the need for cautious design. We hope that our findings encourage more experimental researches to investigate the use of gamification and game-based learning in a variety of contexts, connected to forecasting, data-literacy, decision making and particularly those related to challenges such as mitigation of cognitive biases and misconceptions in order to improve the understanding of our world through rational thinking and data.

CRediT authorship contribution statement

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

Declaration of Competing Interest

None.

Acknowledgments

We wish to thank students of forecasting techniques courses of School of Electrical and Computer Engineering of National Technical University of Athens and to students of Business Administration Department in the School of Business and Economics of the University of Thessaly, for participating in our experiments.

This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie [grant agreement ID 840809] and by Academy of Finland Flagship Programme [337653 - Forest-Human-Machine Interplay (UNITE)]. A part of the work-in-progress version of this work was presented in International GamiFIN Conference 2018, 2019 and in HICSS-52 2019 52nd Hawaii International Conference on System Sciences.

Appendix A. 30-questions test and answers

The information where the answers for each question was presented in the material was not available for the students.

Table A1
30-questions test and answers

30-questions test and answers

1. Which are the main heuristics (categories) that are employed in making judgments under uncertainty?
A. Representativeness B. illusion of validity C. Predictability D. Insensitivity **E. Adjustment from an anchor F. Availability** G. Misconception of Regression H. Conservatism I. Chance J. Optimism, Wishful thinking
The answers can be found on the presented categories of biases (Tversky and Kahneman, 1974).
2. Which kind of heuristics is presented: "Relying upon specific events easily recalled from memory to the exclusion of other pertinent information"?
A. Recency **B. Availability** C. Anchoring D. Regression Effects E. Inconsistency
The answer can be found in the section: "Availability" (Tversky and Kahneman, 1974).
3. Which kind of heuristics is presented: Having the most recent events dominate those in the less recent past, which are downgraded or ignored?
A. Recency B. Availability C. Anchoring D. Regression Effects E. Inconsistency
Definition was given on the lecture (see 3.2.1).
4. Which kind of heuristics is presented: Being unduly influenced by initial information which is given more weight in the forecasting process?
A. Recency B. Availability **C. Anchoring** D. Regression Effects E. Inconsistency
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).
5. Which kind of heuristics is presented: Being unable to apply the same decision criteria in similar situations?
A. Recency B. Availability C. Anchoring D. Regression Effects **E. Inconsistency**
Definition was given on the lecture (see 3.2.1)
6. Which kind of bias is presented: Persistent increases (or decreases) might be due to a genuine trend rather than chance?
A. Recency B. Availability C. Anchoring **D. Regression Effects** E. Inconsistency
Definition was given on the lecture (see 3.2.1)
7. People expect that a sequence of events generated by a random process will present the essential characteristics of that process even when the sequence is short. Which kind of biases are presented?
A. Misconceptions of chance - representativeness B. illusion of validity - representativeness C. Biases due to the retrievability of instances - availability D. Biases of imaginability - availability E. Insufficient adjustment - Adjustment and anchoring F. Anchoring in the assessment of subjective probability distributions - Adjustment and anchoring
The answer can be found in the section: "Representativeness" (Tversky and Kahneman, 1974).
8. Evidently people respond differently when given no evidence and when given worthless evidence. Which king of representativeness is described?
A. Misconceptions of chance - representativeness B. illusion of validity - representativeness C. Insensitivity to sample size - representativeness **D. Insensitivity to prior probability of outcomes - representativeness** E. Insensitivity to predictability - representativeness F. Misconception of regression - representativeness
The answer can be found in the section: "Representativeness" (Tversky and Kahneman, 1974).
9. Based on example of hospital: Which hospital recorded more days on which more than 60% of the babies were boys, ans where students replied about the same for both the large and the small hospital. Which kind of bias is represented?
A. Misconception of chance - representativeness **B. Insensitivity to sample size - representativeness** C. Insensitivity to prior probability of outcomes - representativeness D. Insensitivity to validity - representativeness
The answer can be found in the section: "Representativeness" (Tversky and Kahneman, 1974).
10. People recall easier words that begin with r than words that have r in third position, thus they think that they are more frequent. Which kind of bias is presented?
A. Insensitivity to validity - representativeness B. Insensitivity to sample size - representativeness C. Biases of imaginability - availability **D. Biases due to the effectiveness of a search set - availability**
The answer can be found in the section: "Availability" (Tversky and Kahneman, 1974).
11. People tend to overestimate the probability of conjunctive events and to underestimate the probability of disjunctive events. Which kind of bias is presented?
A. Biases in the evaluation of conjunctive and disjunctive events B. Anchoring in the assessment of subjunctive distributions C. the statement is false D. Insufficient Adjustment
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).
12. Different starting points yield different estimates, which are biased toward the initial values. What is called this phenomenon?
A. Conservatism **B. Anchoring** C. Regression D. Biases of Imaginability E. Availability
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).

Table A1 (*continued*)

30-questions test and answers

13. Gambler's fallacy is another consequence of ?
A. Representativeness B. Availability C. Adjustment and Anchoring
The answer can be found in the section: "Representativeness" (Tversky and Kahneman, 1974).
14. A class of cases are easily retrieved will appear more numerous than a class of equal frequency whose cases are less retrievable. Which kind of bias is presented?
A. Biases due to the retrievability of instances B. Biases due to the effectiveness of a search set C. Biases of imaginability D. illusion of validity E. It is not a kind of bias
The answer can be found in the section: "Availability" (Tversky and Kahneman, 1974).
15. Based on an example of estimation of the percentage of African countries in the United Nations, where subjects gave as answers 25 and 45 for groups that had received 10 and 65, respectively, as starting points. Which kind of biases are presented?
A. Insufficient adjustment - adjustment and anchoring B. Biases due to the effectiveness of a search set - adjustment and anchoring C. Anchoring in the assessment of subjective probability distributions - adjustment and anchoring D. illusory correlation - adjustment and anchoring
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).
16. Anchoring occurs only when the starting point is given to the subject.
True **False**
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).
17. Conservatism is a kind of:
A. Insensitivity to validity - representativeness B. Illusion of validity - representativeness C. Insensitivity to prior probability of outcomes - representativeness D. Insensitivity to predictability - representativeness **E. Insensitivity to sample size - representativeness**
The answer can be found in the section: "Representativeness" (Tversky and Kahneman, 1974).
18. Experience researchers are also prone to the same biases - when they think intuitively.
True False
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
19. Useful heuristics such as representativeness and availability, even though occasionally lead to errors in prediction or estimations, they are retained.
True False
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
20. The majority of people discover the principles of sampling and regression on their own
True **False**
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
21. The empirical analysis of cognitive biases has implications for the theoretical but has no applied role of judged probabilities.
True **False**
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
22. Failure of people to infer from lifelong experience such as fundamental statistical rules as regression toward the mean is surprising for researchers.
True False
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
23. While subjective probabilities can sometimes be inferred from preferences among bets, they are not formed in this fashion.
True False
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
24. When people are asked to assess the frequency of a class, which kind of heuristic is employed?
A. Representativeness **B. Availability** C. Adjustment from an anchor D. None of the above
The answer can be found in the section: "Summary" (Tversky and Kahneman, 1974).
25. When people are asked to judge the probability that an object or event A belongs to class or process B, what kind of heuristic is employed?
A. Representativeness B. Availability C. Adjustment from an anchor D. None of the above
The answer can be found in the section: "Summary" (Tversky and Kahneman, 1974).
26. People's preferences for future outcomes affect their forecasts of such outcomes, what kind of heuristic is employed?
A. Representativeness B. Availability C. Adjustment from an anchor D. None of the above
The answer can be found in the section: "Discussion" (Tversky and Kahneman, 1974).
27. Insufficient adjustment is a kind of?
A. Representativeness B. Availability **C. Adjustment from an anchor** D. None of the above
The answer can be found in the section: "Adjustment and Anchoring" (Tversky and Kahneman, 1974).
28. The illusion of validity is a kind of?

(*continued on next page*)

Table A1 (*continued*)

30-questions test and answers
A. Representativeness B. Availability C. Adjustment from an anchor D. None of the above
The answer can be found in the section: “Representativeness” (Tversky and Kahneman, 1974).
29. Insensitivity to sample size is a kind of ?
A. Representativeness B. Availability C. Adjustment from an anchor D. None of the above
The answer can be found in the section: “Representativeness” (Tversky and Kahneman, 1974).
30. Instances of large classes are recalled better and faster but likely occurrences are not easier to imagine than unlikely ones.
True False
The answer can be found in the section: “Availability” (Tversky and Kahneman, 1974).

Appendix B. Gamified application

Screenshots of the gamified application.

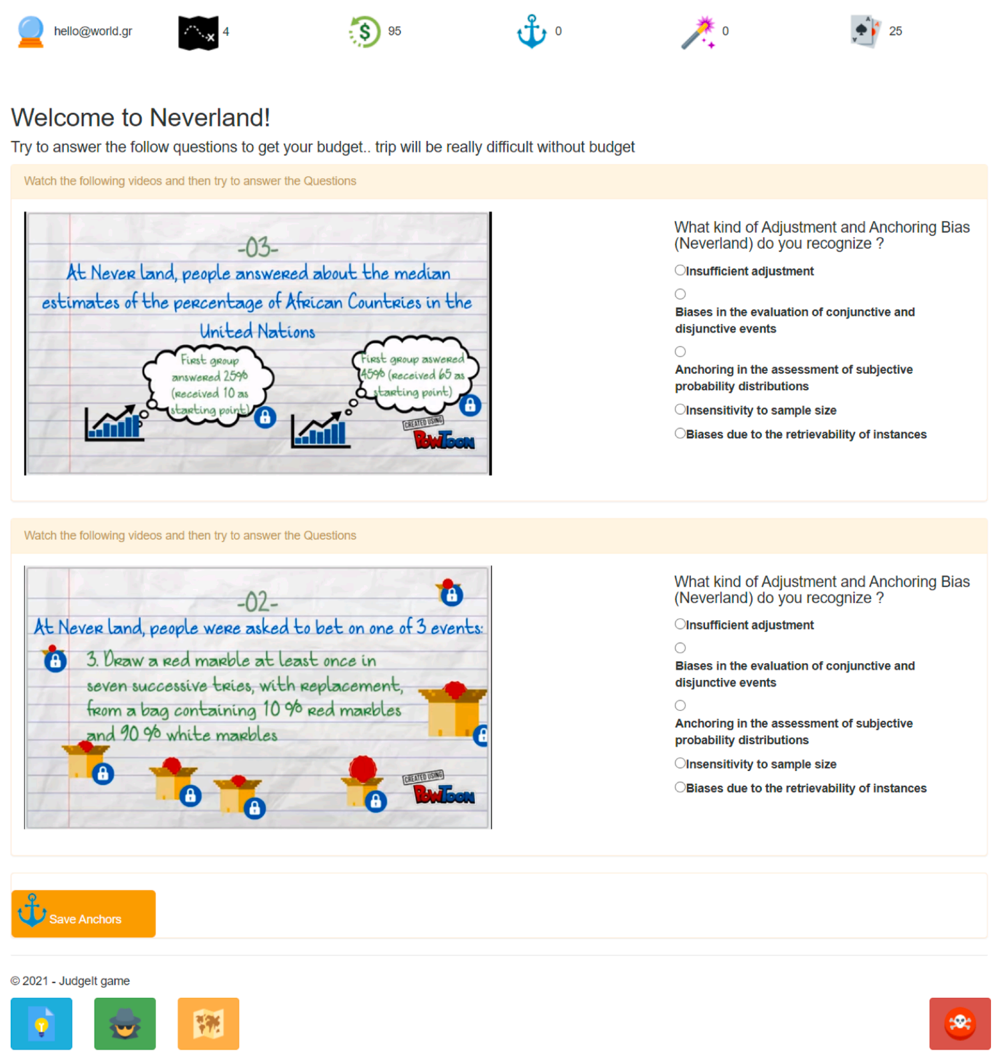

Fig. Appendix B.1. JudgeIt: Screen of the “Neverland” level.

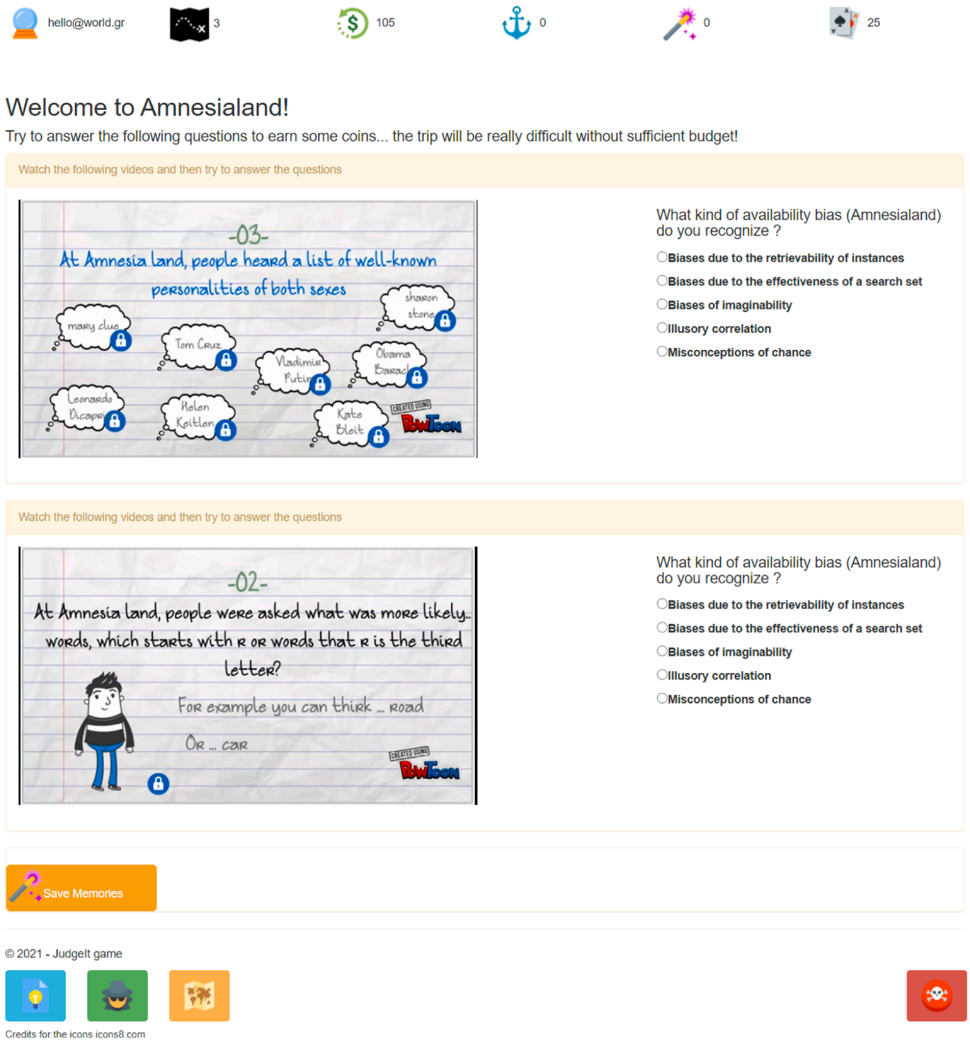

Fig. Appendix B.2. JudgeIt: Screen of the "Amnesialand" level.

References

Albright, R.E., 2002. What can past technology forecasts tell us about the future? Technol. Forecast. Soc. Change 69 (5), 443–464.

Albritton, M.D., McMullen, P.R., 2006. Classroom integration of statistics and management science via forecasting. Decis. Sci. J. Innovat. Educat. 4 (2), 331–336.

Albritton, M.D., McMullen, P.R., Gardiner, L.R., 2003. Or/ms content and visibility in aacsb-accredited us business programs. Interfaces 33 (5), 83–89.

Armstrong, J.S., 2001. Principles of forecasting: a handbook for researchers and practitioners, 30. Springer Science & Business Media.

Bao, T., Hommes, C., Makarewicz, T., 2017. Bubble formation and (in) efficient markets in learning-to-forecast and optimise experiments. Econ. J. 127 (605).

Barata, G., Gama, S., Fonseca, M.J., Goncalves, D., 2013. Improving student creativity with gamification and virtual worlds. Proceedings of the First International Conference on Gameful Design, Research, and Applications, pp. 95–98.

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).

Bhargava, S., Loewenstein, G., 2015. Behavioral economics and public policy 102: beyond nudging. Am. Econ. Rev. 105 (5), 396–401.

Bolger, F., Wright, G., 2017. Use of expert knowledge to anticipate the future: issues, analysis and directions.

Bradfield, R., Cairns, G., Wright, G., 2015. Teaching scenario analysisan action learning pedagogy. Technol. Forecast. Soc. Change 100, 44–52.

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. Education 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, p. 50.

Carr, D., 2005. Contexts, gaming pleasures, and gendered preferences. Simulat. Gaming 36 (4), 464–482.

Clark, J., 2008. Powerpoint and pedagogy: Maintaining student interest in university lectures. College Teach. 56 (1), 39–44.

Cohen, J., 1988. Statistical power analysis for the behavioural sciences. hillsdale, nj: Laurence erlbaum associates.

Conaway, R., Garay, M.C., 2014. Gamification and service marketing. SpringerPlus 3 (1), 653.

Cox Jr, J.E., Loomis, D.G., 2006. Improving forecasting through textbooksa 25 year review. Int. J. Forecast. 22 (3), 617–624.

Craighead, C.W., 2004. Right on target for time-series forecasting. Decis. Sci. J. Innovat. Educ. 2 (2), 207–212.

Dale, S., 2014. Gamification: Making work fun, or making fun of work? Bus. Inf. Rev. 31 (2), 82–90.

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.

Dunbar, N.E., Miller, C.H., Adame, B.J., Elizondo, J., Wilson, S.N., Lane, B.L., Kauffman, A.A., Bessarabova, E., Jensen, M.L., Straub, S.K., et al., 2014. Implicit and

explicit training in the mitigation of cognitive bias through the use of a serious game. Comput. Human Behav. 37, 307–318.
Eroglu, C., Croxton, K.L., 2010. Biases in judgmental adjustments of statistical forecasts: The role of individual differences. Int. J. Forecast. 26 (1), 116–133.
Faria, A.J., Hutchinson, D., Wellington, W.J., Gold, S., 2009. Developments in business gaming: A review of the past 40 years. Simulat. Gaming 40 (4), 464–487.
Gardner, L., 2008. Using a spreadsheet for active learning projects in operations management. INFORMS Trans. Educ. 8 (2), 75–88.
Gavirneni, S., 2008. Teaching the subjective aspect of forecasting through the use of basketball scores. Decis. Sci. J. Innovat. Educ. 6 (1), 187–195.
Gel, Y.R., OHara 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.
Hamari, J., Koivisto, J., Sarsa, H., 2014. Does gamification work?–A literature review of empirical studies on gamification. System Sciences (HICSS), 2014 47th Hawaii International Conference on. IEEE, pp. 3025–3034.
Han, W., Wang, X., Petropoulos, F., Wang, J., 2019. Brain imaging and forecasting: insights from judgmental model selection. Omega 87, 1–9.
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.
Heinonen, S., Minkkinen, M., Karjalainen, J., Inayatullah, S., 2017. Testing transformative energy scenarios through causal layered analysis gaming. Technol. Forecast. Soc. Change 124, 101–113.
Hofacker, C.F., De Ruyter, K., Lurie, N.H., Manchanda, P., Donaldson, J., 2016. Gamification and mobile marketing effectiveness. J. Interact. Market. 34, 25–36.
Huotari, K., Hamari, J., 2017. A definition for gamification: anchoring gamification in the service marketing literature. Electron. Market. 27 (1), 21–31.
Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: principles and practice. OTexts.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.
Jacobson, D., Dargue, B., 2019. Change your mind. International Conference on Human-Computer Interaction. Springer, pp. 108–117.
Jia, Y., Xu, B., Karanam, Y., Voida, S., 2016. Personality-targeted gamification: a survey study on personality traits and motivational affordances. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2001–2013.
Kapp, K.M., 2013. The gamification of learning and instruction fieldbook: Ideas into practice. John Wiley & Sons.
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: A review of gamification research. Int. J. Inf. Manag. 45, 191–210.
Kroes, J.R., Chen, Y., Mangiameli, P., 2013. Improving students' data analysis and presentation skills: The ocean state circuits, inc. forecasting project. Decis. Sci. J. Innovat. Educ. 11 (2), 165–174.
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.
Landers, R.N., Auer, E.M., Collmus, A.B., Armstrong, M.B., 2018. Gamification science, its history and future: Definitions and a research agenda.1046878118774385
Lawrence, M., Goodwin, P., O'Connor, M., Önkal, D., 2006. Judgmental forecasting: A review of progress over the last 25 years. Int. J. Forecast. 22 (3), 493–518.
Leitner, J., Leopold-Wildburger, U., 2011. Experiments on forecasting behavior with several sources of information–a review of the literature. Eur. J. Oper. Res. 213 (3), 459–469.
Liu, C.-C., 2016. Understanding player behavior in online games: The role of gender. Technol. Forecast. Soc. Change 111, 265–274.
Loomis, D.G., Cox, J.E., 2000. A course in economic forecasting: rationale and content. J. Econ. Educ. 31 (4), 349–357.
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., 1996. Forecasting: its role and value for planning and strategy. Int. J. Forecast. 12 (4), 513–537.
Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2018. The m4 competition: Results, findings, conclusion and way forward. Int. J. Forecast. 34 (4), 802–808.
Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The m4 competition: 100,000 time series and 61 forecasting methods. Int. J. Forecast. 36 (1), 54–74.
Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 2008. Forecasting methods and applications. John wiley & sons.
Mallette, L.A., Saldaña, J., 2019. Teaching qualitative data analysis through gaming. Qualit. Inquiry 25 (9-10), 1085–1090.
Mann, S., Robinson, A., 2009. Boredom in the lecture theatre: An investigation into the contributors, moderators and outcomes of boredom amongst university students. Brit. Educ. Res. J. 35 (2), 243–258.
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.
Martelli, D., 2013. Learning from experience or learning for convenience? Qual. Res. Financ. Market.
Martí-Parreño, J., Seguí-Mas, D., Seguí-Mas, E., 2016. Teachers attitude towards and actual use of gamification. Proc.-Soc. Behav. Sci. 228, 682–688.
McGonigal, J., 2011. Reality is broken: Why games make us better and how they can change the world. Penguin.
McKernan, B., Martey, R.M., Stromer-Galley, J., Kenski, K., Clegg, B.A., Folkestad, J.E., Rhodes, M.G., Shaw, A., Saulnier, E.T., Strzalkowski, T., 2015. We dont need no stinkinbadges: The impact of reward features and feeling rewarded in educational games. Comput. Hum. Behav. 45, 299–306.
Mendez-Carbajo, D., 2019. Experiential learning in macroeconomics through fredcast. Int. Rev. Econ. Educ. 30, 100137.
Morewedge, C.K., Yoon, H., Scopelliti, I., Symborski, C.W., Korris, J.H., Kassam, K.S., 2015. Debiasing decisions: Improved decision making with a single training intervention. Policy Insight. Behav. Brain Sci. 2 (1), 129–140.
Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A., 2017. Gamified crowdsourcing: Conceptualization, literature review, and future agenda. Int. J. Human-Comput. Stud. 106, 26–43.
Morschheuser, B., Hassan, L., Werder, K., Hamari, J., 2017. How to design gamification? a method for engineering gamified software. Inf. Softw. Technol.
Müller, B.C., Reise, C., Seliger, G., 2015. Gamification in factory management education–a case study with lego mindstorms. Procedia CIRP 26, 121–126.
Mustafa, T., Farooq, Z., Asad, Z., Amjad, R., Badar, I., Chaudhry, A.M., Khan, M.A.Z., Rafique, F., 2014. Lectures in medical educaton: what students think? J. Ayub Medical College Abbottabad 26 (1), 21–25.
Nacke, L.E., Deterding, C.S., 2017. The maturing of gamification research. Comput. Human 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.
Nooriafshar, M., 2005. Interactive technology in teaching judgemental forecasting. Proceedings of ABBSA Conference 2005 (ABBSA 2005). Australasian Business and Behavioural Sciences Association, pp. 835–840.
Novak, E., Johnson, T.E., Tenenbaum, G., Shute, V.J., 2016. Effects of an instructional gaming characteristic on learning effectiveness, efficiency, and engagement: using a storyline for teaching basic statistical skills. Interact. Learn. Environ. 24 (3), 523–538.
Osatuyi, B., Osatuyi, T., de la Rosa, R., 2018. Systematic review of gamification research in is education: a multi-method approach. CAIS 42, 5.
Papastergiou, M., 2009. Exploring the potential of computer and video games for health and physical education: a literature review. Comput. Educ. 53 (3), 603–622.
Pedreira, O., Garca, F., Brisaboa, N., Piattini, M., 2015. Gamification in software engineering a systematic mapping. Inf. Softw. Technol. 57, 157–168.
Poncin, I., Garnier, M., Mimoun, M.S.B., Leclercq, T., 2017. Smart technologies and shopping experience: Are gamification interfaces effective? the case of the smartstore. Technol. Forecast. Soc. Change 124, 320–331.
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. Hum. Behav. 58, 48–63.
Rowe, G., Wright, G., 1999. The delphi technique as a forecasting tool: issues and analysis. Int. J. Forecast. 15 (4), 353–375.
Ryan, T.E., 2006. Motivating novice students to read their textbooks. J. Instruct. Psychol. 33 (2), 135A.
Salen, K., Tekinbas, K.S., Zimmerman, E., 2006. The game design reader: A rules of play anthology. MIT press.
Seaborn, K., Fels, D.I., 2015. Gamification in theory and action: A survey. Int. J. Human-Comput. Stud. 74, 14–31.
Seaborn, K., Fels, D.I., 2015. Gamification in theory and action: A survey. Int. J. Hum.-Comput. Stud. 74, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006.
Snider, B.R., Eliasson, J.B., 2013. Beat the instructor: An introductory forecasting game. Decis. Sci. J. Innovat. Educat. 11 (2), 147–157.
Soll, J.B., Milkman, K.L., Payne, J.W., 2015. A users guide to debiasing. Wiley Blackwell Handbook Judgment Decis. Making 2, 924–951.
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.
Stewart, M., 2012. Understanding learning: theories and critique. Univ. Teach. Focus. 3–20.
Tetlock, P.E., Gardner, D., 2016. Superforecasting: The art and science of prediction. Random House.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Veinott, E.S., Leonard, J., Papautsky, E.L., Perelman, B., Stankovic, A., Lorince, J., Hotaling, J., Ross, T., Todd, P., Castronova, E., et al., 2013. The effect of camera perspective and session duration on training decision making in a serious video game. 2013 IEEE International Games Innovation Conference (IGIC). IEEE, pp. 256–262.
Worthen, B., 2003. Future results not guaranteed: contrary to what vendors tell you, computer systems alone are incapable of producing accurate forecasts. CIO 16 (19), 1.
Xi, N., Hamari, J., 2019. Does gamification satisfy needs? a study on the relationship between gamification features and intrinsic need satisfaction. Int. J. Inf. Manag. 46, 210–221.
Yildirim, I., 2017. The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. Internet Higher Educ. 33, 86–92.
Zepke, N., Leach, L., 2010. Improving student engagement: ten proposals for action. Act. Learn. Higher Educ. 11 (3), 167–177.
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 postdoctoral researcher at Gamification Group, Tampere University. She received her Ph.D. in School of Electrical and Computer Engineering of National Technical University of Athens. 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, datafication and educational methods in teaching forecasting.

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, in 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.

Prof.Juho Hamari is a Professor of Gamification and leads the Gamification Group at the Faculty of Information Technology and Communication Scienes at Tampere UniversityHis 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. Prof. 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.