Computers in Human Behavior Reports 3 (2021) 100044

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-reports

Leveraging learning with gamification: An experimental case study with bank managers

Felipe Cechella[a,b,1,*], Gardênia Abbad[a], Ralf Wagner[b]

[a] University of Brasília, ICC A1-114, Brasília, DF, Brazil
[b] University of Kassel, Mönchebergstraße 1, Room 4009, Kassel, Hessen, Germany

ARTICLE INFO

Keywords:
Gamification
Training
Learning
Education

ABSTRACT

Gamification has attracted the attention of academics and practitioners as a promising tool for promoting behavioral change. This paper is an experimental case study investigating the effects of gamification on the learning outcomes when used as an instructional procedure for training managers in a Brazilian bank. High-order cognitive learning was assessed with situational tests before and after the training events. Data were collected in a quasi-experimental design involving three groups, two groups in training events and one untrained control group. The results show that gamification had a positive effect on learning but with results similar to training that used an instructional design without gamification. Gamification facilitated learning even with less time available for conceptual explanations and discussions. Competitive game elements were shown to be effective as an aspect of instructional design, especially when a digital feedback system was used as an assessment tool in training. This study innovatively collected data in a Brazilian bank and developed an instrument to measure high-order cognitive skills. Further investigation is required using other game elements that promote cooperation, autonomy, and personalization.

1. Introduction

Gamification and its motivational potential are seen as trends for increasing and promoting user engagement in manifold contexts (Mekler et al., 2017). The commercial and cultural success of using game elements to promote behavioral changes has attracted the attention of researchers and companies (Deterding et al., 2011; Koivisto & Hamari, 2019; Seaborn & Fels, 2015). Despite its popularity as a topic of interest in various fields, such as marketing and education, gamification remains a recent object of study (Robson et al., 2015).

Investment in the training and development of employees gave rise to a paradox in organizations. At the organizational level, investment is high, but companies express considerable skepticism when the focus is on behavioral changes at the individual level (Baldwin et al., 2017). *Training* is defined as the systematic inculcation of knowledge, skills, and attitudes (KSA) that improve performance (Goldstein & Ford, 2002). Learning has two facets: (1) know-how, or the capacity to act, and (2) know-why, or the ability to articulate a conceptual understanding of an experience (Kim, 1998). A complimentary classical definition describes learning as the process of creating knowledge through a transformative experience (Kolb, 1984).

Instructional design is defined as the systematic development of procedures, based on learning and instructional theory, that aim to facilitate instruction (Ozcinar, 2009). Digital transformation technologies have emerged as a dominant paradigm for organizations that want to succeed and remain competitive over time (Jackson, 2019), and this phenomenon has also influenced instructional design in terms of the procedures chosen (sets of tools, devices, and methods) to achieve the instructional objectives (Salas & Cannon-Bowers, 2001).

A primary motivation for using gamification is its promise of motivating individuals to act with greater engagement (Dominguez et al., 2013; Pettit et al., 2015; Tan & Hew, 2016). This promise has attracted businesses' interest in their quest to involve employees in an active learning process (Hew et al., 2016; Ibanez et al., 2014). Although using games for learning can provide various benefits to individuals (De-Marcos et al., 2014), the design and development of full-fledged learning games are generally costly, which has opened space for the use of gamification as a rapidly implemented tool (Ibanez et al., 2014).

Gamification's effectiveness to promote learning and its outcomes remains an ongoing research challenge because most studies have been

* Corresponding author. University of Kassel, Mönchebergstraße 1, Room 4009, Kassel, Hessen, Germany.
E-mail addresses: uk075009@uni-kassel.de (F. Cechella), gardenia@unb.br, gardenia.abbad@unb.br (G. Abbad), rwagner@wirtschaft.uni-kassel.de (R. Wagner).
[1] University of Kassel, Mönchebergstraße 1, Room 3006, Kassel, Hessen, Germany. (present address).

https://doi.org/10.1016/j.chbr.2020.100044
Received 3 March 2020; Received in revised form 10 December 2020; Accepted 10 December 2020
Available online 7 January 2021

done outside business settings (Koivisto & Hamari, 2019). Research carried out in the organizational training context is still rare (Larson, 2020). Studies on learning and gamification are recurrent outside business settings, especially in educational environments (Majuri et al., 2018). To identify which techniques produce the most effective results, it is necessary to build upon theoretical foundations that connect procedures used by instructional designers with the choice of game elements, as indicated by the Theory of Gamified Learning (Landers, 2015). Consequently, this research aims to answer the following research question: What effects does the use of gamification as an instructional procedure have on learning?

The research design is a case study, investigating bank managers in training events in a work environment, a context rarely considered in previous research, thus approaching gamification and learning topics from a unique perspective. Research on gamification and learning mainly measured lower cognitive levels (knowledge and comprehension), with only a few exceptions – some researchers focused on the evaluation of complex factual knowledge, such as those involving decision-making in medical, surgical, or clinical areas (Graafland et al., 2004; Lin et al., 2015). Therefore, this study's learning measures, using a quasi-experimental design, go beyond previous studies by measuring high cognitive level learning tasks (synthesis and evaluation), through the use of situational tests.

This experimental case study aims to investigate the gamification effects on learning in a bank context. The purpose is to evaluate the effectiveness of gamification for employee training. Learning was assessed through open-ended questions (simulating real work situations) in a quasi-experimental design that applied repeated measures in two experimental groups (training treatments) and one control group (untrained).

This study contributes to the literature on gamification in three ways: (1) showing how to use gamification for learning aligned with instructional design theories and principles; (2) developing and applying situational tests to assess high-order cognitive skills; and (3) creating criteria for assessing high-order cognitive skills in with a situational test.

The remainder of this paper is structured as follows. The following section reviews the theoretical foundations of describing learning in business and relates them to gamification design. The materials and methods section describes the procedures for implementing the study, and the results and discussion sections establish a dialogue with the findings of research on this topic. The final section discusses the study's contributions and limitations and proposes a research agenda for the future investigation of gamification and learning.

2. Theoretical framework

2.1. Learning in business settings

Organizations expect that training will result in better performance at work. Employees value feedback, opportunities to develop skills, and challenging tasks that satisfy personal ambitions while also contributing to the achievement of the organization's goals (Noe et al., 2014). In a business environment, the terms *instruction*, *course*, *event*, and *program* refer to learning activities, and *instructor*, *educator*, *learner*, and *participant* are often used to refer to those engaged in the teaching-learning process (Goldstein & Ford, 2002).

Over the past century, *training* was synonymous with an instructor-led, face-to-face event, usually in a classroom. Adopting modern technology made possible more individualized events that are closer in time and space to the application at work. This evolution in the nature of technology and work has led to instantaneously available knowledge and the development of skills on demand using a variety of technologically advanced methods (Ford & Meyer, 2013). Professionals acquire and maintain increasingly complex KSA for problem-solving, teamwork, self-management, and career planning.

Salas et al. (2005) distinguish four basic steps in creating a training solution: (1) training needs assessment, (2) instructional design, (3) training execution, and (4) assessment. The first, needs assessment, provides strategic information (e.g., organizational strategy, trainee profiles, existing solutions) for decision-making before designing the training program. Identifying gaps in KSA is particularly important in this step. Instructors and employees provide perspectives and details of the expected behaviors to be developed through training (Antes, 2014). The instructional design and assessment steps are outlined in the following subsections.

2.2. Instructional design

The instructional design process is both descriptive and prescriptive. It is descriptive because it shows the relationships that occur during the process, explaining the interactions of the design components. It is prescriptive because it guides, assigns methods, and generates strategies in an active, goal-oriented manner adapted to various instructional contexts (Branch & Kopcha, 2014). While learning theories describe basic individual learning processes, instructional theories address how individual differences interact with instruction and context to produce learning outcomes (Reigeluth, 1999). Among the aspects of learning that can be further investigated are the instructor's role and the types of KSA that are best learned online, in the classroom, or with hybrid approaches (Noe et al., 2014).

A result of the training-needs assessment, and the first step of instructional design, is specifying the instructional objectives, that is, describing the behaviors that will address gaps in KSA. The types of KSA and their impact on the effectiveness of the training should be considered to ensure a fit between training delivery, method, and the task being taught (Arthur et al., 2003). Instructional objectives explicitly inform the participants of the expected learning outcomes—what will be expected of them at the end of the training experience—and provide the basis for designing the learning environment and the assessment. These objectives comprise measurable actions represented by verbs, and each sentence must be connected to a KSA. After identifying the objectives, the designers define the sequence of objectives to facilitate learning. At this stage, decisions are made on instructional procedures and the format of activities (Goldstein & Ford, 2002).

The instructional procedures are a set of tools, methods, activities, or techniques designed to impact the participants' KSA and provide utility by improving their work performance (Ittner & Douds, 1997). The learning mode is the type of (inter)action by which content is provided to learners: learning by doing, by watching, or by listening. When trainees acquire or develop KSA by performing a task, they are using the learning-by-doing mode, an approach aligned with the educational philosophy of experiential learning (Kolb & Kolb, 2005). In this study, gamification (combined with other tools and techniques) is used as an instructional procedure aiming to facilitate learning to achieve the instructional objectives.

2.3. Gamification to promote learning

Active learning methodologies introduce practical activities aimed at increasing learner engagement (Prince, 2004). Students who participate in interactive activities learn concepts better, retain them longer, and apply them more effectively in other contexts than students who experience passive instructional designs (Freeman et al., 2014). Gamification can be applied in a learning context in two ways, (1) partially gamified experiences and (2) fully gamified experiences, the latter applying gamification throughout the experience and completely changing the instructional design. In that case, most instructional procedures need to be redesigned to include the game elements (Barata et al., 2017).

Previous research has revealed contradictions in gamification's impact on learning outcomes. Some studies affirm that gamification generates better learning performance when assessing the application of practical concepts (Alcivar & Abad, 2016; Hamari et al., 2016). By contrast, other studies find that gamification does not yield positive results in terms of factual or conceptual learning, even in written assessments (De-Marcos et al., 2014; Dominguez et al., 2013). However, others have found positive effects associated with theoretical knowledge acquisition (Filsecker & Hickey, 2014; Ibanez et al., 2014). Accordingly, gamification needs to be integrated with complementarity to other learning tools (Buckley & Doyle, 2017; De-Marcos et al., 2016) rather than being introduced in isolation in the learning environment.

Although gamification is a tool to promote learning, it nevertheless needs further empirical exploration to determine its efficacy. There is a need to assess the high complexity cognitive levels (e.g., evaluation, synthesis, and analysis, according to learning taxonomies) because most available previous research measured learning using multiple-choice items or self-reporting instruments. This may be due to the designers' and scholars' difficulty in creating or accessing environments in which complex knowledge is developed (Attali & Arieli-Attali, 2015; Lin et al., 2015; Tan & Hew, 2016). Researchers are facing challenges including the assessment of these cognitive levels developing instruments that are appropriate to what they are intended to measure (Landers & Landers, 2015; Tan & Hew, 2016). Scholars and businesses practitioners can take advantage from clarifying how gamification can influence workers' behavioral changes as measured with instruments that simulate work situations (Graafland et al., 2014).

2.4. Assessing high-level cognitive skills

To assure that instructional objectives are attained, instruction and learning assessment must be aligned (Gulikers et al., 2004). This process is based on instructional objectives and is focused on developing KSA. The purpose of assessment is also to verify the acquisition of high-order mental processes, not merely the memorization of facts and the comprehension of concepts. A successful, authentic learning assessment positively influences learning outcomes and participants' motivation (Herrington & Herrington, 1998).

Two important reasons for using KSA-based assessments are their (1) construct validity and (2) consequent validity in the nomological network (Bagozzi, 1982), that is, their impact on learning outcomes at the various cognitive levels (Bloom, 1956). Construct validity is achieved when a measurement is evaluating what it is supposed to measure (Peter, 1981). In KSA assessment, this means that the task should appropriately reflect what needs to be assessed, with the content of the assessment involving authentic tasks that represent real-life problems or situations (Gielen et al., 2003).

The crucial element of an authentic performance assessment is the task's degree of fidelity to the conditions under which the performance will generally occur. An authentic assessment requires learners to use the combinations of KSA under the same criteria demanded in workplace situations. In real life, employees usually know the required criteria for their performances, and authentic assessments must follow the same logic. Learners need to know in advance the quality standards for the products that are expected of them (Gulikers et al., 2004).

Typically, learning criteria are measures of training outcomes, although they are not working performance measures. They are typically operationalized as situational tests with open-ended questions, e.g., using paper and pencil or answering online. Working performance measures help to identify training effects by assessing an individual's current job function characteristics (Arthur et al., 2003).

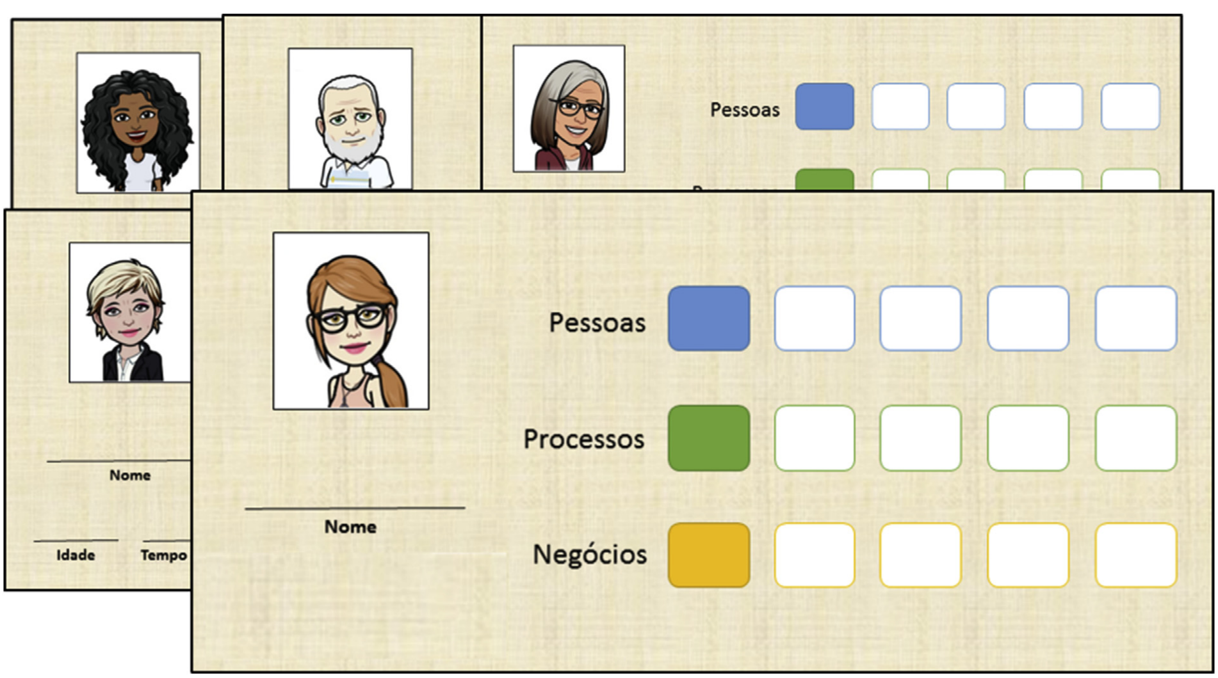

Fig. 1. Avatars.

(b) Leaderboard and team points: Following the same logic as the upgrade points, team points were earned as a performance reward in practical activities. The participants could acquire a different number of points in the same activity, adding even more value to the excellence of their performances. The team points were displayed on a leaderboard that compared the teams' performances at the end of the practical activities.

(c) Digital feedback system: In all practical activities, the participants had to present products (individually or as a team). Because it was a training for managers with the theme of performance appraisal, a digital performance-assessment tool was used in all the practical activities. A digital tool was available on the participants' tablets that allowed them to assign scores to the performances presented throughout the training. The score was derived from the evaluations that were carried out by the participants themselves. A 7-point Likert scale was adopted, which was similar to that used by the participants in the execution of their performance-appraisal tasks in the workplace. The best-evaluated performances were rewarded with team points and upgrade points in line with the criteria established and previously presented to the participants. In the event that did not use gamification, the product evaluation of the practical activities was conducted qualitatively by the instructor and the other participants in group discussions.

3. Materials and methods

This research is positioned and designed as an experimental case study conducted in a Brazilian bank. The research participants were employees of the studied bank, who held a management position – responsible for conducting team processes. As it was not possible to randomly assign individuals to experimental groups, participation was for convenience, this research is characterized as a quasi-experiment (White & Sabarwal, 2014).

The threats to the experiment internal validity were treated as the following: (1) ambiguous temporal precedence – pre-test and post-test implementation; (2) selection – pre-test scores differences were not statistically for the three groups; (3) history and maturation – the presence of a control group; (4) regression – scores were not extreme; (5) attrition – no loss of subjects; (6) testing – different but equivalent situational tests; (7) instrumentation – there was no change in the nature of the measure (Shadish et al., 2002). Additionally, the same instructor could not apply to both training events. During the first one, a diary was written with theoretical and practical tips, trying to prevent unbalance between the training events. Every time that was necessary, the instructor from the second course was advised about the tips used in the first.

3.1. Instructional design development

The application domain of this experimental case study was a classroom course internally developed by a Brazilian bank. The selected training is called the Performance Appraisal Course, a 16-h classroom training conducted on two consecutive days in sessions of 8 h each. The training's target audience is around 30,000 employees who hold management positions in the bank. This study has a descriptive and exploratory first stage (instructional design, instrument, and assessment criteria development) followed by a quantitative confirmatory second stage (a quasi-experiment for testing the theory-based hypotheses).

To create the quasi-experiment, it was necessary to develop two distinct training events, using gamification as an instructional element in only one of them. A workshop was conducted with bank experts to create the courses' instructional designs. The instructional design with gamification reconfigured all the instructional procedures using game elements. Both courses' instructional designs had the same instructional objectives, which were rewritten during the same workshop. The next step in the instructional design was ordering the activities, which was done based on the sequence of performance appraisal activities in the workplace. The quasi-experiment used mobile devices (tablets) in practical activities for the creation and sharing of learning products.

The instructional design of one of the training events used gamification throughout the event to offer a fully gamified experience. Gamification was absent from the other instructional design, replaced by longer discussions and explanations. This procedure was adopted to maintain equivalence between the training events, with both having the same instructional objectives but differing in their instructional procedures. Therefore, the only experimentally manipulated aspect was the instructional procedure. All other instructional design aspects remained constant so as not to affect the quasi-experiment internal validity (Appendix A).

Since the use of game elements required more detailed instruction beforehand and more complex procedures in executing and finalizing the activities, the time remaining in each class activity in the training without gamification was used to prolong group discussions and lecturer dialogue-framed demonstrations. On average, the participants in the non-gamified (traditional) group had about 20 min more per activity than those in the gamified group to discuss concepts and the outcomes of products presented in the practical activities. This represented about 2 additional hours to discuss the course content with the instructor and among the participants. The steps taken in the instructional design stage included (1) rewriting the instructional objectives, (2) ordering the course activities, and (3) creating the instructional procedures with and without gamification.

Because gamification design is the central theme of this study and its use as an instructional element was the basis of the quasi-experiment, the details of which game elements were chosen and how they were adopted in the instructional design of the training event are described in the following paragraphs.

(a) Avatars and upgrade points: Avatars were adopted with the purpose of creating a bond between the participants and the symbolic representations chosen by them. In total, 37 models of avatar were offered, representing different genders, ages, and ethnicities. Each avatar had three empty progress bars that were completed by the participants during the event in response to their performance in the practical activities. To complete the progress bars, the participants had purchase upgrade points, which were distributed according to their performance in the practical activities. See Fig. 1.
(b) Leaderboard and team points: Following the same logic as the upgrade points, team points were earned as a performance reward in practical activities. The participants could acquire a different number of points in the same activity, adding even more value to the excellence of their performances. The team points were displayed on a leaderboard that compared the teams' performances at the end of the practical activities.
(c) Digital feedback system: In all practical activities, the participants had to present products (individually or as a team). Because it was a training for managers with the theme of performance appraisal, a digital performance-assessment tool was used in all the practical activities. A digital tool was available on the participants' tablets that allowed them to assign scores to the performances presented throughout the training. The score was derived from the evaluations that were carried out by the participants themselves. A 7-point Likert scale was adopted, which was similar to that used by the participants in the execution of their performance-appraisal tasks in the workplace. The best-evaluated performances were rewarded with team points and upgrade points in line with the criteria established and previously presented to the participants. In the event that did not use gamification, the product evaluation of the practical activities was conducted qualitatively by the instructor and the other participants in group discussions. See Fig. 2.
(d) Super avatars: Two content-review activities were conducted during the instructional event. The first aimed to activate the participants' memory of their prior knowledge. The second content review was also the final team-competition challenge. Before it started, a super avatar was provided for each team. This game element had two attributes. See Fig. 3.

(1) XP: The accumulated team point scores (acquired by the teams throughout the event) shown on the leaderboard before the activity began.
(2) Combo: The sum of all upgrade points acquired by team members throughout the event. With every correct answer, the groups purchased the equivalent of the Combo of their super avatar, adding this value to the XP (starting score).

In the event without gamification, the content review, assessment made with the digital feedcabk system in the gamification event, took the form of group discussion, with the instructor asking questions and participants who wanted to respond raising their hands and giving the answer.

3.2. Instruments

To measure the learning outcome in this study, it was necessary to develop specific instruments to assess the participants' achievements

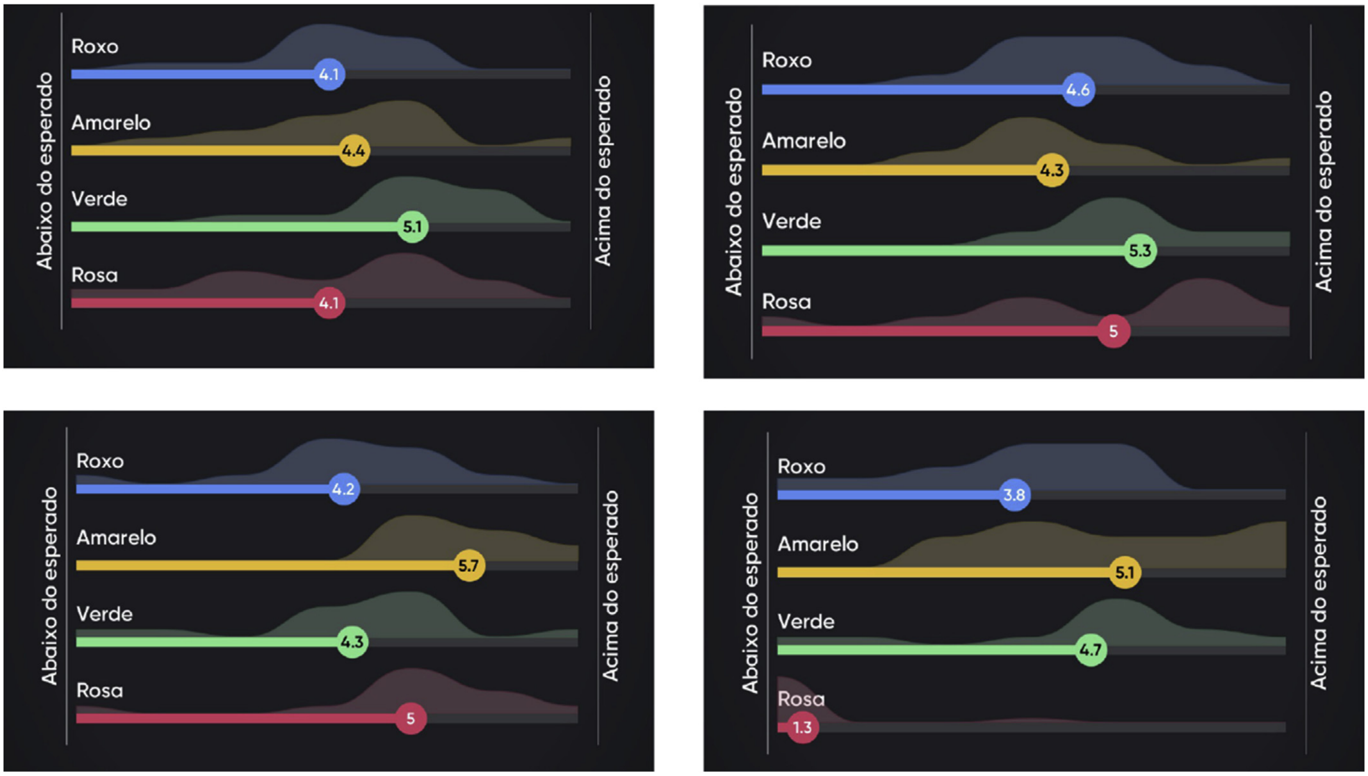

Fig. 2. The digital feedback system.

(d) Super avatars: Two content-review activitis were conducted during the instructional event. The first aimed to activate the participants' memory of their prior knowledge. The second content review was also the final team-competition challenge. Before it started, a super avatar was provided for each team. This game element had two attributes:

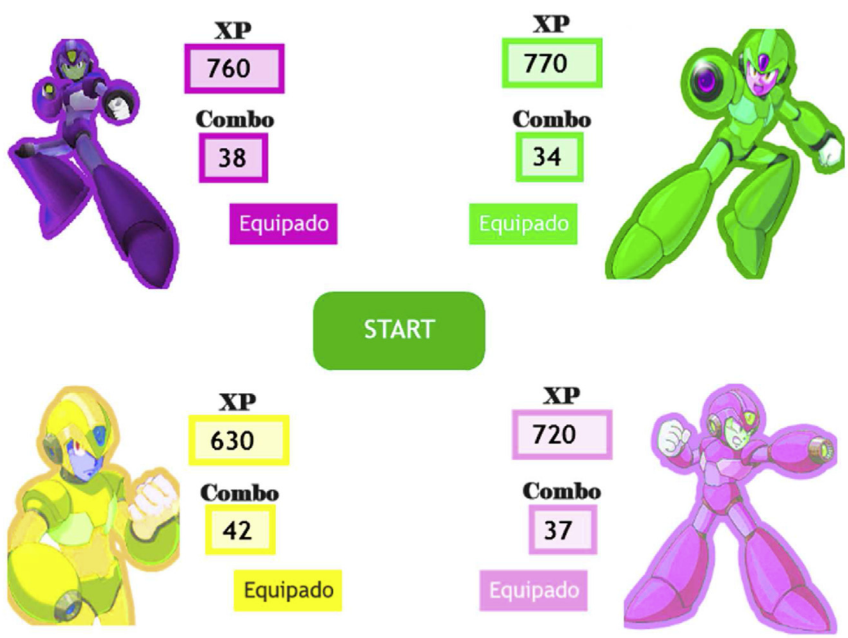

Fig. 3. Super avatars.

according to the instructional objectives discussed above. Managers with expertise in performance appraisal were invited to participate in a second workshop to develop the tests to measure high-order cognitive skills. The assessment instruments aimed to simulate work situations experienced by the participants.

The purpose of these assessments was to measure performance in high-order cognitive tasks. The objectives chosen in creating the tests involved tasks such as understanding and analyzing scenarios, writing, synthesizing ideas, and evaluating simulated performances (fictitious employee descriptions and bank situations were created to simulate authentic work tasks). Thus, the creation of *situational* tests with open-ended questions became vital. In both training events, the activities that explained and reviewed concepts were prerequisites for the subsequent practical activities. Consequently, the questions also delivered information to the participants about situations and specifications that allowed them (in combination with the conceptual knowledge acquired) to answer what was being asked.

To assess the participants' learning in the quasi-experimental setting, it was necessary to validate two instruments, one to be used before (pre-test) and another to be used after (post-test) the training events (Appendix B). In the untrained control group, the tests were given sequentially in one session. To verify the clarity of the test's instructions, to determine the test's equivalence, and to gain evidence of the construct validity, two pilot applications were conducted—the first with two managers belonging to the training target audience and the second with two experts from the board responsible for training content in the bank—and adjustments were made after each one.

The next step was the creation of criteria to assess learning outcomes. A pilot training event was conducted with the aim of testing and adjusting the instructional designs and collecting data from the created instruments. In the pilot group, the instruments were also applied before and after the event. Furthermore, the answers were subjected to an analysis to identify and establish the criteria for judging the responses given by the participants. It was necessary to create a correction sheet indicating the performances that would be crucial for the participants to receive a score for each question assessed.

For each instructional objective, the underlying criteria were identified and transcribed. The identified criteria were transformed into questions to facilitate the correction of the learning assessments. The questions allowed only two answers: (1) *Yes* (the participant was considered to have correctly answered the question and received the score appropriate to it) or (2) *No* (the participant was considered not to have answered the question correctly and did not receive the score). Each correct answer received 4 or 5 points, depending on the complexity level assessed. The maximum possible overall score in the pre-test and the post-test was 100 points (Appendix C). The grades were assigned by only one rater. The correction sheet used was fixed before the final dataset.

3.3. Quasi-experimental setting and data sample

The quasi-experiment took a quantitative approach with repeated measures (two-time points). The independent variable was categorical and composed of three levels represented by each group (experimental group 1, experimental group 2, and the control group). The dependent variable was the learning outcome, which was measured by a pre-test score and a post-test score (both from 0 to 100 points) based on the established criteria. In total, 53 managers participated in the experiment. The sample was divided (non-random assignment) into three groups: experimental group 1 (use of gamification; $N = 19$), experimental group 2 (no gamification; $N = 19$), and the control group (no training; $N = 15$).

To recruit for the training events (experimental groups 1 and 2), e-mails were sent to approximately 300 individuals, and the acceptance criterion was the first respondents indicating availability. Data collection started with the control group, to which both the pre-test and the post-test were given sequentially in one session. This group was composed of untrained managers accessible in their work environments. The group was 46.7% male and 53.3% female; 66.7% were located in business units and 33.3% in support units; the ages ranged from 37 to 53 years ($M = 45.60$, $SD = 5.35$).

The two training events were delivered in sequence, first to experimental group 2 (instructional design without gamification), which was 68.4% male and 31.6% female; 57.9% were business unit managers and 42.1% support unit managers; the ages ranged from 28 to 61 years ($M = 45.3$, $SD = 8.36$). The second event was conducted with experimental group 1 (with gamification as an instructional element). That group was 52.6% male and 47.4% female; 63.2% were business unit managers and 36.8% support unit managers; the ages ranged from 31 to 60 years ($M = 45.31$, $SD = 8.81$). Fig. 4 shows the instruments' time of application to the three groups.

Below are presented the hypotheses for the comparisons within and between the groups.

H1. Learning will occur for the participants in the training with gamification (experimental group 1).

H2. Learning will occur for the participants in the training without gamification (experimental group 2).

H3. Learning will not occur for the control group participants.

H4. Participants in the training with gamification will have a higher learning outcome than the participants in the training without gamification.

4. Results

The descriptive statistics showed no outliers detected through histogram visualization. The assumption of normality was not violated as assessed by the Shapiro-Wilk test for the pre-test ($p = .405$) and the post-test ($p = .690$) or by graphical analysis of the boxplots. The paired-samples *t*-test was adopted to determine whether there was a statistically significant mean difference between the participants' scores on the pre-test and post-test for all three groups. The participants of both training sessions, with and without gamification (groups 1 and 2), showed a statistically significant increase between the scores obtained from the pre-test and those of the post-test, confirming that the learning occurred. By contrast to the first two results, the control group participants showed no statistically significant difference between the scores of the pre-test and the post-test. Table 1 shows the results obtained in the paired *t*-tests.

A one-way ANOVA was conducted to determine whether the scores for the pre-test differed between the three groups. There were no outliers, and the tests had already demonstrated data normality. There was a homogeneity of variances assessed by Levene's test ($p = .983$), and the differences between the three groups' scores for the pre-test were not statistically significant ($F\{2, 50\} = 0.825$, $p = .444$). This result is also important because, even though the participants were not randomly assigned to each group, there were no statistical differences in the scores assessed in the pre-test.

An independent-samples *t*-test was run to determine whether there were differences in the learning outcomes measured by the post-test scores for the two quasi-experimental groups that experienced the training event (with and without gamification). Again, the variable was normally distributed, and there was homogeneity of variances ($p = .752$). The scores for the post-test were higher for the participants of the training with gamification (experimental group 1), but the difference was not statistically significant ($M = 4.47$, $t\{36\} = 1.118$, $p = .271$). All the hypotheses were tested, and only H4 was rejected. Fig. 5 summarizes the findings.

5. Discussion and conclusions

The gamification literature is not sufficiently clear when it defines gamification simply as the use of game elements in non-game contexts. This definition is insufficient to support the design of a compelling experience or even to describe the minimum criteria for developing a gamified solution. In this study, gamification was allied to the fundamentals of instructional theories. Here, game elements were used as an instructional element focusing primarily on competition rather than cooperation between the participants. This choice was risky because it is known that competitive gamification can hinder learning (Hanus & Fox, 2015). In this research, the risk was taken because the study was conducted in a bank context, where managers are used to being measured by their daily performance. The method applied for evaluation and reward

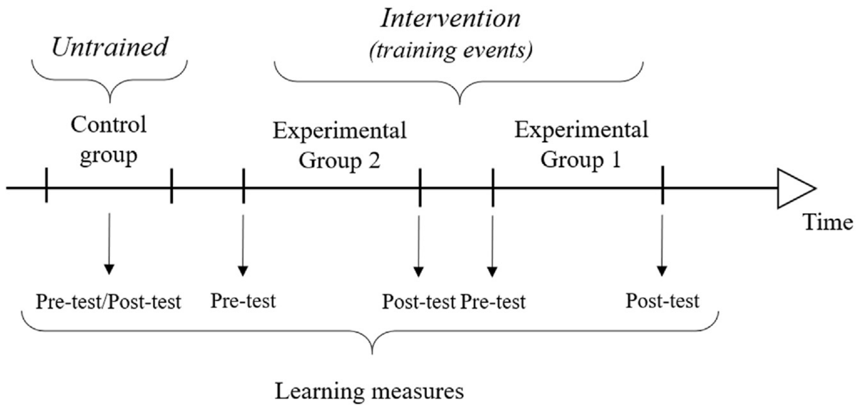

Fig. 4. The instruments' time of application.

Table 1
Paired *t*-test results.

Group	Pre-test		Post-test		*n*	95% CI for mean difference	*p*	*t*	df
	M	SD	M	SD					
Exp. group 1	30.00	13.1	42.32	12.8	19	−20.62, −4.01	.006*	−3.11	18
Exp. group 2	25.16	12.4	38.16	12.5	19	−20.80, −5.19	.003*	−3.49	18
Control group	25.60	12.2	27.27	7.1	15	−7.82, 4.49	.571	−0.58	14

*$p < .05$.

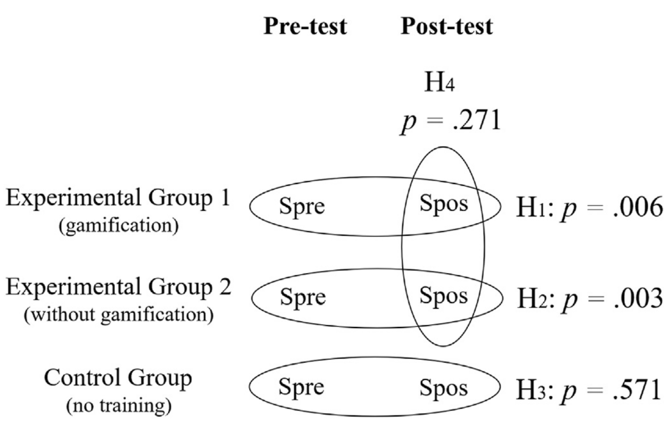

Fig. 5. Results summary.

in competition may have mitigated the possible adverse effects.

The participants themselves were responsible for assessing the practical activities' performance during the course with gamification, and this was done through the use of an interactive digital feedback system. This performance assessment in practical activities may have alleviated the burden of an assessment done by the instructor or a judge.

Theories on gamification do not provide an in-depth explanation of how to use game elements aligned with purpose and behavior (Koivisto & Hamari, 2019). In this study, the theoretical fundamentals provided by the training and development area addressed this shortcoming of gamification studies. On the one hand, training studies have over 100 years of research expertise and sufficient empirical evidence to demonstrate the effectiveness of methods in some situations and their inefficiency in others (Bell et al., 2017). On the other hand, gamification has a recent theoretical basis, with articles implementing gamification design for learning proliferating only for around a decade – not considering game-based learning, simulation games, etc.

Regarding instrument development and the emergence of the assessment criteria, this task was possible only because of the analysis of the pilot group. This indicates that businesses may not be able to predict quality criteria during the instructional design stage. Developing detailed criteria is an exhaustive and laborious task. It is essential to discuss the importance of the pilot events not only to test the instructional design but also to assess the higher learning levels, identifying criteria that allow the instruction to be entirely redesigned based on this underlying information, which was not previously visible.

Although gamification studies demonstrate some inconsistencies in learning performance results, most point positively to the use of gamification (Alcivar & Abad, 2016; De-Marcos et al., 2014, 2016; Dominguez et al., 2013; Filsecker & Hickey, 2014; Hamari et al., 2016; Ibanez et al., 2014), although no study was found measuring high-order cognitive skills using situational tests with open-ended questions. Still, the results obtained in this research were not surprising. The gamification applied in this training was intentionally competitive, requiring long instruction periods before the activities began and a long time to compute the scores obtained on the gamified activities by each participant and group. This extra time that experimental group 2 (training without gamification) had in class was used to prolong group discussions and lecturer dialogue-framed demonstrations. The similar results achieved for both groups indicate that this extra time spent discussing concepts did not result in a higher learning outcome for this group of participants. One explanation may relate to cognitive load; gamification may serve as a relief valve when participants are overloaded with repeated information for extended periods (Su, 2016).

Another relevant result derives from the comparative analyses between the pre-test and post-test scores for the three groups. The results confirmed that the trained groups made improvements in the learning while the group that did not undergo training did not improve (as confirmed by the statistical analysis). This outcome is essential to confirm the quality of the applied training events, but it also showed the equivalence between the instruments used to measure the learning outcomes. This research was unconventional in applying open-format situational performance tests to measure high-order cognitive processes, such as synthesis and evaluation. Another significant result that corroborates the internal validity of the quasi-experiment was the measurement of the initial repertoire levels of the three groups' participants. In the pre-test scores, no significant differences were found. This fact is relevant because the individuals were not randomly assigned to the experimental groups.

It is essential to discuss the relevance of the use of instruments that measure learning in business settings, especially those that measure the more complex levels of the cognitive domain, as was the case in this study. Developing, applying, and assessing learning with open-ended questions requires an extra effort from everyone involved in the creation of the training solution, but the benefits outweigh the costs.

5.1. Contributions

This study's first contribution relates to the use of gamification as an instructional procedure using the fundamentals of instructional design theories and approaches. This contribution arose from aligning the gamification elements with the instructional objectives defined for the training, complementing the other created or selected procedures to facilitate learning. This contribution is valuable because it shows that it is essential, when designing gamified solutions for learning, to identify and write learning objectives according to learning taxonomies. This will allow a better assessment of the achievement of these objectives, adequately measuring the learning outcomes – producing important insight for decision makers.

A second contribution of this work derives from the development of the instrument. Just as the activities included in the instructional design had objectives aligned to the expected performances at work, the situational tests were also created to simulate and approximate the situations that the participants encountered in their management functions in the bank context of this case study.

A third contribution was the construction of the criteria to assess the learning outcomes as measured by the developed instruments (the tests created). The construction of these criteria was based on data analysis of the tests conducted with the pilot group. These criteria were fundamental for the learning measures, empirically demonstrating the importance of constructing quality standards prior to the provision of any training solution for the organization's employees. Most of the criteria identified could not have been predicted by the instructional designers before testing. Workshops to identify KSA gaps and develop training, even with content experts and training target members, are probably unable to identify and categorize all the performance standards that may be derived from analytical work on the responses given in situational tests with open-ended questions. This recommendation can be a valuable contribution to the effectiveness of training solutions. Finally, this research is innovative because it used gamification to deliver training to managers in a bank and used a quasi-experiment design to measure high-cognitive order skills with situational tests.

5.2. Limitations

This study has some limitations deriving from the statistical results, small sample size, and specificities of this case study context – non-generalizable conclusions. Although the omnibus effect is not statistically significant, there were group differences in the pre-test measures of the outcome variable – but not relevant to threaten the internal validity of the study (Shadish et al., 2002). The result and deriving conclusions that gamification does not promote higher learning when compared to traditional training did not consider any power analysis.

Some further arises from the game elements applied in two compounds. Therefore, it was not possible to quantify the effectiveness of one element compared to another. The gamification encompassed the entire learning environment from start to finish of the training event. Elements of inherently competitive games were used. Also, some gamification principles were not adopted in full, such as promoting immersion through the use of narratives. The promotion of collaborative activities was also not explored. Another limitation of the research is that it did not investigate some profile variables of the participants, e.g., their learning preferences and styles.

5.3. Future research

In terms of theoretical considerations, gamification requires very robust theoretical models in developing its taxonomies, organizing the game elements, and ensuring that their use conforms to the level and type of behavior expected of the participants after the experience. The models found in the literature do not contemplate how each element can be best developed when used in a particular situation, participant profile, environment, or type of KSA that the designers wish to develop in the participants. Prescriptive models have been offered that effectively facilitate designers' work in creating and developing instructional solutions within organizations.

Concerning the design of gamification, it was observed that the literature agrees in affirming that some designs are neglected in empirical research. There is a gap in the use of cooperative environments, in which the participants promote their self-management by cooperating and building knowledge while using gamification to support the collaborative process. There is a need to further investigate the differing effects of competitive gamification (the design adopted in this research) and cooperative gamification (not adopted here). Also, one research focus might be on the participants' orientation regarding competition and cooperation in the gamified environment. Competitively oriented individuals might derive a greater benefit from competitive gamification. Identifying this individual profile could indicate more appropriate paths in the design of gamification.

Regarding methodological approaches, it is necessary to continue to conduct multi-method research and experiments. The combination of self-reporting instruments (with large samples) and experiments that manipulate various gamification designs can help to consolidate the relationships proposed by the current theoretical models, indicating more appropriate ways to design gamification that considers individuals' preferences and characteristics.

Finally, gamification applied in learning research within business settings remains a little-explored subject but one that should continue to be investigated because many organizations look for innovative ways to improve individual and team performance. Gamification has proven effective, but it needs support from scientific methods and results to expand its use in organizations. Thus, the participation of researchers is crucial to provide a theoretical and empirical base to help companies that sell and buy gamification to design, apply, and assess solutions that promote behavioral change and lead to the expected positive outcomes.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: By the time of the Research, the author Felipe Cechella was an employee of the bank where the study was done.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2020.100044.

References

Alcivar, I., & Abad, A. G. (2016). Design and evaluation of a gamified system for ERP training. *Computers in Human Behavior, 58*, 109–118.
Antes, A. (2014). A systematic approach to instruction in research ethics. *Accountability in Research, 21*, 50–67.
Arthur, W., Bennett, W., Edens, P., & Bell, S. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234–245.
Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education, 83*, 57–63.
Bagozzi, R. P. (1982). A field investigation of causal relations among cognitions, affect, intentions, and behavior. *Journal of Marketing Research, 19*(4), 562–584.

Baldwin, T. T., Kevin Ford, J., & Blume, B. D. (2017). The state of transfer of training research: Moving toward more consumer-centric inquiry. *Human Resource Development Quarterly, 28*(1), 17–28.

Barata, G., Gama, S., Jorge, J., & Goncalves, D. (2017). Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior, 71*, 550–585.

Bell, B., Tannenbaum, S., Ford, J., Noe, R., & Kraiger, K. (2017). 100 years of training and development research: What we know and where we should go. *Journal of Applied Psychology, 102*, 305–323.

Bloom, B. (1956). *Taxonomy of educational objectives*. N.Y.: Longmans, Green.

Branch, R. M., & Kopcha, T. J. (2014). Instructional design models. In *Handbook of research on educational communications and technology* (pp. 77–87). Springer.

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. *Computers & Education, 106*(6), 43–55.

De-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education, 75*, 82–91.

De-Marcos, L., Garcia-Lopez, E., & Garcia-Cabot, A. (2016). On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers & Education, 95*, 99–113.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th international academic MindTrek conference* (pp. 9–15).

Dominguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernandez-Sanz, L., Pages, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education, 63*, 380–392.

Filsecker, M., & Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. *Computers & Education, 75*, 136–148.

Ford, J. K., & Meyer, T. (2013). Advances in training technology: Meeting the workplace challenges of talent development, deep specialization, and collaborative learning. In L. T. Convert (Ed.), *The psychology of workplace technology* (pp. 43–76). Routledge.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.

Gielen, S., Dochy, F., & Dierick, S. (2003). The influence of assessment on learning. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 37–54). Kluwer Academic Publishers.

Goldstein, L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation*. Wadsworth.

Graafland, M., Vollebergh, M. F., Lagarde, S. M., van Haperen, M., Bemelman, W. A., & Schijven, M. P. (2014). A serious game can be a valid method to train clinical decision-making in surgery. *World Journal of Surgery, 38*(12), 3056–3062.

Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179.

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. *Computers & Education, 80*, 152–161.

Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Education Research and Development, 17*(3), 305–322.

Hew, K. F., Huang, B., Chu, K. W. S., & Chiu, D. K. W. (2016). Engaging Asian students through game mechanics: Findings from two experiment studies. *Computers & Education, 92–93*, 221–236.

Ibanez, M.-B., Di-Serio, A., & Delgado-Kloos, C. (2014). Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on Learning Technologies, 7*(3), 291–301.

Ittner, P. L., & Douds, A. F. (1997). *Train-the-trainer: Instructor's guide*. Human Resource Development Press.

Jackson, N. C. (2019). Managing for competency with innovation change in higher education: Examining the pitfalls and pivots of digital transformation. *Business Horizons, 62*(6), 761–772.

Kim, D. H. (1998). The link between individual and organizational learning. *Strategic Management of Intellectual Capital, 41*, 62.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191–210.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *The Academy of Management Learning and Education, 4*, 193–212.

Landers, R. N. (2015). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752–768.

Landers, R., & Landers, A. (2015). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming, 45*(6), 769.

Larson, K. (2020). Serious games and gamification in the corporate training environment: A literature review. *TechTrends, 64*(2), 319–328.

Lin, D. T., Park, J., Liebert, C. A., & Lau, J. N. (2015). Validity evidence for surgical improvement of clinical knowledge ops: A novel gaming platform to assess surgical decision making. *The American Journal of Surgery, 209*(1), 79–85.

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

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior, 71*, 525–534.

Noe, R., Clarke, A., & Klein, H. (2014). Learning in the twenty-first-century workplace. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 245–275.

Ozcinar, Z. (2009). The topic of instructional design in research journals: A citation analysis for the years 1980-2008. *Australasian Journal of Educational Technology, 25*(4).

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research, 18*(2), 133–145.

Pettit, R. K., McCoy, L., Kinney, M., & Schwartz, F. N. (2015). Student perceptions of gamified audience response system interactions in large group lectures and via lecture capture technology. *BMC Medical Education, 15*(1), 15–92.

Reigeluth, C. H. (1999). *Instructional-design theories and models: A new paradigm of instructional theory*. LEA.

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons, 58*(4), 411–420.

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

Salas, E., Wilson, K. A., Burke, C. S., & Priest, H. A. (2005). Using simulation-based training to improve patient safety: What does it take? *Joint Commission Journal on Quality and Patient Safety, 31*(7), 363–371.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Tan, M., & Hew, K. F. (2016). Incorporating meaningful gamification in a blended learning research methods class: Examining student learning, engagement, and affective outcomes. *Australasian Journal of Educational Technology, 32*(5), 19–34.

White, H., & Sabarwal, S. (2014). Quasi-experimental design and methods. *Methodological Briefs: Impact Evaluation, 8*, 1–16.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education., 93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.

Su, C. H. (2016). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. *Multimedia Tools and Applications, 75*(16), 10013–10036. https://doi.org/10.1007/s11042-015-2799-7