UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE – FACE
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA

GAMIFICATION EFFECTS ON LEARNING AND TRANSFER OF TRAINING

FELIPE SCHNEIDER CECHELLA

Orientadora: Prof. Dra. Gardênia da Silva Abbad

Dissertação de Mestrado em Administração

BRASÍLIA/DF
2019

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE – FACE
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA
ÁREA DE CONCENTRAÇÃO – ESTUDOS ORGANIZACIONAIS E GESTÃO DE PESSOAS

GAMIFICATION EFFECTS ON LEARNING AND TRANSFER OF TRAINING

FELIPE SCHNEIDER CECHELLA

Dissertação de Mestrado submetido ao Programa de Pós-Graduação em Administração da Universidade de Brasília como requisito para a obtenção do grau de Mestre em Administração.

Aprovado pela seguinte Comissão Examinadora:

Prof[a]. Dr[a]. Gardênia da Silva Abbad
Orientadora (Universidade de Brasília – PPGA/UnB)

Prof. Dr. Pedro Paulo Murce Meneses
Examinador Interno (Universidade de Brasília – PPGA/UnB)

Prof. Dr. Ralf Wagner
Examinador Externo (Kassel University – Germany)

Prof[a]. Dr[a]. Josivania Silva Farias
Examinadora Suplente (Universidade de Brasília – PPGA/UnB)

Brasília/DF, março de 2019.

Acknowledgment

To Lívia, the best person.

To Lili, who taught me not to have no limits in the search of my dreams.

To my family.

To professor Gardênia and all Impacto group colleagues.

An meine deutschen Freunde, die mir geholfen haben, den Ort zu finden, an dem ich hingehöre.

To God, that indicated and allowed me to live this path.

Resumo

O objetivo geral deste trabalho é avaliar os efeitos da gamificação na aprendizagem (generalização e transferência de treinamento) em eventos presenciais para gestores de uma instituição financeira brasileira. Foi utilizado um delineamento misto, qualitativo e quantitativo, com fontes primárias e secundárias de informação, e medidas repetidas de variáveis dependentes. Os estudos qualitativos compreenderam a construção de um treinamento gerencial gamificado, testes situacionais de avaliação da generalização de aprendizagem e critérios de correção das respostas aos testes. Os estudos quantitativos incluíram dois quase-experimentos. No primeiro, o delineamento incluiu três grupos Experimental 1, treinamento com gamificação; Experimental 2, treinamento sem gamificação; e o Grupo controle, sem treinamento; e medidas repetidas, com a aplicação de pré-teste e pós-teste de generalização da aprendizagem. O segundo quase-experimento, comparou estes três grupos quanto à transferência do treinamento, definida como o desempenho dos gerentes treinados nos registros de avaliação de desempenho dos membros de sua equipe em arquivos de dados da organização. Foi adotado um delineamento com os três grupos e medidas repetidas com sete medidas pré intervenção e sete medidas pós intervenção. A variável dependente é a transferência de treinamento e foi mensurada por meio de cinco taxas criadas com base em critérios. Os resultados obtidos no primeiro quase-experimento indicam que o grupo com a gamificação não mostrou melhores resultados de aprendizagem no nível de generalização quando comparado aos grupos com treinamento tradicional. No segundo quase-experimento, entretanto, o treinamento com gamificação mostrou mais resultados positivos com relação à transferência de treinamento do que os demais grupos. Esta pesquisa contribui de forma inovadora para os campos de estudo da gamificação e da aprendizagem nas organizações, pois: (1) realizou quase-experimento usando a gamificação em um treinamento para gestores de uma instituição financeira; (2) mediu a generalização da aprendizagem, utilizando testes situacionais; e (3) mediu a transferência de treinamento avaliando dados objetivos dos desempenhos dos gestores no local de trabalho, tal como registrados em fontes secundárias de dados disponíveis na organização. Propõem-se expandir as pesquisas de gamificação dentro das organizações, utilizando amostras maiores e avaliando seus impactos na melhoria dos resultados de negócios.

Palavras-chaves: Gamificação, TD&E, Aprendizagem, Generalização, Transferência de treinamento

Abstract

Gamification has emerged as a strategy that promises to help organizations promote behavior change in individuals. The main objective of this work is to evaluate the effects of gamification on learning (generalization and transfer of training) in two classroom events for managers of a Brazilian financial institution. This research method includes qualitative and quantitative approach, with primary and secondary sources of information, and repeated measures of dependent variables. Qualitative studies include the construction of the training events, situational tests to evaluate the generalization and criteria to correct the responses to tests and results on transfer of training. Quantitative studies included two quasi-experiments. In the first, the design included three groups Experimental 1, training with gamification; Experimental 2, training with expositions and discussions; and the Control Group, without training; and repeated measures, with the application of pretest and posttest of generalization. The second quasi-experiment compared these three groups on the transfer of training, measured by rates of trained managers performances, data obtained from files recording their team members' evaluation; the repeated measures happened in seven pre-intervention measures and seven post-intervention. The results obtained in the first quasi-experiment indicate that the group with gamification did not show better learning results at the level of generalization when compared to the groups with traditional training. In the second quasi-experiment, the training with gamification showed more positive results regarding the transfer of training than the other groups, although only in a few of the hypotheses. This research contributes in an innovative way to the gamification study's field and learning in organizations, because: (1) it performed quasi-experiment using gamification in a training for managers of a financial institution; (2) measured the generalization of learning, using situational performance tests; and (3) measured the transfer of training by evaluating objective data on managers' performances in the workplace. As a research agenda, it is proposed to expand the gamification research within organizations using larger samples and evaluating impacts on improving business results.

Keywords: Gamification, T&D, Learning, Generalization, Transfer of training

Summary

List of Figures

List of Tables

1 Introduction

In 2017, the global gamification market value was USD 2.17 billion (Mordor Intelligence, 2018). The growth in mobile devices supports the growing recognition of gamification as a behavioral change method, inducing innovation, productivity, and engagement. Gamification is consolidating as a commercial and cultural success and has attracted researchers and companies' attention (Seaborn & Fels, 2015). Besides its popularization as a topic of interest in various fields such as marketing and education, gamification is yet a recent phenomenon as a study subject (Robson, Plangger, Kietzmann, McCarthy and Pitt, 2015).

There is no consensus on gamification's concept. Some researchers understand it necessarily linked to technology, directly associated with video games (Hamari, Koivisto, & Sarsa, 2014), while others understand it as only the use of game elements as opposed to full games (Deterding, 2012). For some authors, gamification is directly associated with the use of game elements in learning contexts (Dominguez et al., 2013).

The most frequently used definition conceptualizes gamification simply as the use of game elements in non-game contexts (Deterding, Dixon, & Khaled, 2011). Non-game contexts are seen as any context other than entertainment (Giannetto, Chao, & Fontana, 2013). Even so, gamification is not necessarily linked to serious contexts, but it does require improvement in some human behavior (Seaborn & Fels, 2015). Gamification has also been commonly confused with the use of simulators, business games, and game theory (Deterding, 2012).

A more detailed definition affirms that gamification is the mechanics, aesthetics and game-based thinking's utilization to engage and motivate people's action, promoting learning to solve problems. Further, a game is a system where players engage in an abstract challenge, defined by rules, interactivity, and feedback, generating quantifiable results allied with affective reactions (Kapp, 2012).

According to Robson et al. (2015), there are three main reasons why gamification has gained the corporate world's attention. Firstly is the excellent video game industry growth and evolution over the last twenty years. A second, the consolidation of social media and mobile devices connected to the Internet. Finally, organizations are continually seeking new ways to promote learning and how to influence individual's behavior. Deterding (2012) explains that to a proper gamification implementation, it is first crucial to understand the organization's strategic objectives, identifying activities that involve businesses core values, developing a more

profound individuals' profile' understanding and what motivates them to engage with the organization's objectives.

McGonigal (2009) explains that when excluding all the differences between genres and technological complexity, all games share four defining characteristics: objective, rules, a feedback system, and voluntary participation. Kapp (2012) points out that thinking of a game as a system signifies the combination and integration of game elements. Furthermore, a single element is not capable of creating an immersive and engaging gamification environment, but rather the relationship between the elements.

One of the main reasons of using gamification is the promise to motivate individuals to perform routine activities with greater engagement (Dominguez et al., 2013, Pettit, McCoy, Kinney, & Schwartz, 2015; Tan & Hew, 2016). This capacity has attracted educators' interest in the quest to involve learners in an exciting and active process (Hew, Huang, Chu, & Chiu, 2016; Ibanez, Di-Serio, & Delgado-Kloos, 2014).

Games' use in the educational process is known to bring many benefits (De-Marcos, Dominguez, Saenz-de-Navarrete and Pagés, 2014). Nonetheless, the design and development of complete learning games are generally costly, which has opened space for the use of gamification as an active learning methodology with rapid implementation (Ibanez et al., 2014; Landers, 2014).

In this study, learning refers to changes that occur in individuals' behavior resulting from their interaction with the context (Abbad & Borges-Andrade, 2004). One of the organizational strategies to promote learning and behavioral change in their employees is the use of the training and development system (T&D). Organizations systematize their learning solutions through corporate education (Abbad, Borges-Ferreira, & Nogueira, 2006a). Corporate universities promote events and programs to stimulate the development of knowledge, skills, and attitudes (KSAs) that prepare employees for challenges that do not yet exist, to perform new job functions or to improve performance in their current role. The T&D departments are structured to develop all stages of an instructional solution, with its four sub-systems: (1) training needs assessment, (2) instructional design, (3) training execution and (4) assessment (Vargas & Abbad, 2006).

The instructional design consists of the stage where teaching-learning situations are built to facilitate and support learning. Four stages compose it: definition of instructional objectives; sequencing objectives and contents; selection or creation of instructional strategies; and definition of learning assessment criteria (Abbad, Nogueira, & Walter, 2006b). Instructional strategies are all operations, events or learning situations, created or selected during

instructional design, to facilitate all learning stages: acquisition, retention, generalization, and transfer of training (Abbad et al. al., 2006b). Accordingly, gamification can be applied as an instructional strategy, integrated into the instructional design (Alves, 2014).

T&D assessment stage is a process that includes data collection, used for later decision-making. At the individual level, the immediate results of the T&D system are the participants' opinions or satisfaction with the experience (reaction assessment), and the KSAs improvement or acquisition (learning). Generalization is learning performance in activities somewhat different from those adopted in training, the tasks are in some respect distinct from those adopted throughout training – with a variation of the stimuli, situations or cases, though not situated in the workplace (Pilati & Abbad, 2005).

The assessment may exceed the judgment or system's verification of effectiveness, serving as a medium for knowledge construction to the theoretical approaches adopted in the instructional design (Borges-Andrade, 2006). The learning assessment can be constructed and applied in different ways to identify how much the individual's behavior has changed. This process should be part of the instructional design stage (Abbad et al., 2006a). Transfer of training can be defined, for measurement purposes, as the effective application in the work context of KSAs acquired or developed during T&D experiences (Pilati & Abbad, 2005).

This study has academic and practical contributions, seeking to fill theoretical and methodological gaps in gamification literature. Most gamification initiatives are not generating the expected positive outcomes because of the inappropriate understanding of how to design an experience able to promote behavioral change and performance improvement (Robson et al., 2015). More empirical studies are necessary to verify the extent of gamification effects on participants and their performances. Also, gamification research should be anchored in experimental designs (Seaborn & Fels, 2014).

Specifically, organizations need further investigation on how to promote behavioral change, testing novel instructional strategies, focusing on the transfer of training increase (Pilati & Borges-Andrade, 2006). Corporate universities also want to verify beyond the T&D functioning and the results it generates, searching for possibilities to intervene, providing feedback to improve the system and increase the organizational benefits (Borges-Andrade, Abbad, & Mourão, 2012).

Gamification studies in learning contexts, in line with the literature review, have been done, almost exclusively, in environments outside organizations. This study intends to research the organizational environment, investigating managers in a training event, a context not found in the literature review. Concerning learning measures, done with a quasi-experimental design

intervention, this research is also innovative. It measures learning at generalization level with the use of situational performance tests, and transfer of training with variables originated from secondary data taken directly from the organization, reflecting managers real performances at their workplace. This research uniquely addresses gamification and learning topics way compared to the literature found.

Therefore, the research question that leads to the necessity of this study is: what effects does the use of gamification as an instructional strategy produce on learning in generalization and transfer of training?

Thus, the main objective of this study is to evaluate the gamification effects on learning in generalization and transfer of training, in classroom events for managers in a Brazilian financial institution. In order to achieve the main objective, there are the following specific objectives required:

i. Create two equivalent instructional designs, one with the use of gamification as an instructional strategy and another without its use.
ii. Create two equivalent situational learning performance tests to be applied before and after the classroom training event.
iii. Create assessment criteria to measure learning at generalization level.
iv. Create assessment criteria to measure learning at the transfer of training levels.

These are the following contents approached in the next sections: (1) a theoretical framework on T&D learning assessment and instructional design; (2) literature review on gamification in learning context; (3) designing gamification based on Mechanics, Dynamics and Emotions (MDE) model; (4) methodological procedures; (5) qualitative and quantitative results; (6) discussion; and (7) final considerations.

2 Theoretical framework

The next steps of this research are intended to bring the theoretical foundations that guide learning studies in organizations. The theoretical basis comes from national and international sources (books and literature review articles), and the sequence approached will be: (1) Learning in T&D; (2) learning assessment at generalization and transfer of training levels; (3) instructional design theories; and (4) how to design instruction.

2.1 Learning in T&D systems

The training industry makes substantial investments to improve their employees' performances in order to obtain positive effects on organizational results. The investments in 2018, in the US alone, was USD 87.6 billion (Training Magazine, 2018). Further, it is not only organizations that have expectations regarding the training results in performance at work. Today, workers value feedback, opportunities to develop skills, challenging tasks that contribute to satisfying personal ambitions, while also contributes to the organizational goals' achievements (Noe, Clarke, & Klein, 2014).

The organizations' T&D system identify the conditions vital for individuals to learn or improve KSAs required by their work functions (Abbad & Borges-Andrade, 2004). Learning in organizations can occur either in a natural or induced way. T&D solutions structure and design formal learning processes, aiming at facilitating and maximizing the effectiveness of KSAs' acquisition, retention, generalization, and transfer. The core intention is to facilitate learning, and consequently, accelerate behavioral change (Abbad et al., 2006a). For these researchers, it is important the instructional strategies' diversification, to motivate individuals during the training experience.

Training is the systematic acquisition of skills, rules, concepts or attitudes that must result in a performance improvement. The term instruction, course, event or program is used to refer to the learning environment. Instructor, educator, learner, participant is used to refer to the individuals engaged in the teaching-learning process (Goldstein & Ford, 2002). The T&D four stages of creating a training solution are: (1) training needs assessment, (2) instructional design, (3) training execution and (4) assessment (Vargas & Abbad, 2006). The only stage that will not be explored in details in this study, because it is not related to the research goals is the needs assessment. Although, it is crucial to bring the primary information about this first step on creating a training solution.

The T&D need assessment is the process that provides the information needed to design an instructional program. One result of this phase is the understanding of learner's needs before training. This information allows to establish and prioritize the instructional objectives in the next phase of T&D solutions' development –instructional design. Identifying gaps in KSAs is particularly central to needs assessment. Instructors and employees are sources to provide perspectives and details of which are the expected behaviors to be developed through training (Antes, 2014).

Learning concept is not one-dimensional and refers not only to the acquisition and retention but also to generalization, and transfer of KSAs acquired or developed in a training event (Abbad et al., 2006a). Learning is the participants' demonstration performing a task at the training event's end, and instructional objectives characterize these performances – descriptions of observable actions necessary to perform the tasks (Pilati & Abbad, 2005). Many learning situations are intermediate behaviors related to cognitive processes not relevant in the work environment. However, others are relevant and terminal, exhibited by the training participants when performing their work functions. The learning measurement in organizations should involve at least the assessment of the KSAs' at the generalization level. Instead, the measures are typically restricted to the acquisition and retention of intermediate KSAs, incompatible with nature, degree, and complexity of the tasks performed in the work environment (Abbad et al., 2006a).

The learning primary psychological process is the KSAs acquisition developed during the T&D event and constitutes the first stage of learning. Secondly, retention involves the KSAs' storage in short-term memory, subsequently transferred to long-term memory. Pilati and Abbad (2005) argue that in experimental studies or training situations, generalization is measured through the participant's performance in activities or final tests somewhat different from those adopted in training (generalization). Posttests learning assessment is examples of generalization measures when the tasks are in some respect distinct from those adopted throughout training. Altogether, there is a variation of the stimuli, situations, cases, and questions or answers, as well as the performances required from the participants (Pilati & Abbad, 2005). In this research, this is the definition used and measure that occurred.

T&D programs can have long-term effects at two or three levels: job behavior – change in participants' use of acquired KSAs, usually named transfer of training; organization – change occurred in the unit or team in which a T&D event participant works; and final value – change in production or services or other social and economic benefits provided by the organization (Borges-Andrade, 2006). This section demonstrated the importance of T&D as a learning system to organizations. Next, it is explained how T&D systems assess the efficacy of their products. Unique attention granted to learning assessment, at generalization and transfer of training levels.

2.2 T&D learning assessment

T&D assessment is a process that collects data to use later for decision-making. A summative assessment refers to designing, obtaining and analyzing information, providing decision's subsidies to the adoption or rejection of an entire program or a single T&D event (Borges-Andrade, 2006).

The assessment may also test theoretical principles that were constructed throughout the T&D event or program's instructional design. The theoretical approach used may be a reason to select a program for evaluation. In this situation, the assessment should transcend its characteristic of judging the achievement or not of the objectives or the verification of effectiveness, but also build knowledge about the theoretical approach used (Borges-Andrade, 2006).

The immediate results of a T&D system can be assessed at two levels: reaction – opinions or participants' satisfaction on various aspects of the T&D program or event; and learning – participants' acquisition or development of competencies (KSAs) indicated in the instructional objectives (Borges-Andrade, 2006).

Reaction assessment is an essential first step in evaluating the effectiveness of a T&D event because it indicates whether participants were satisfied or not with the training (Abbad, Zerbini, & Borges-Ferreira, 2012). Abbad (1999) affirms that an instructional event is expected to produce favorable reactions in participants. Reaction assessment should be collected at the end of the instructional event (Abbad et al., 2012).

Alternatively, the learning assessment can be constructed and applied in a variety of ways, trying to ascertain how participants' performance has changed, considering instructional objectives defined for the T&D solution. An item or question to assess learning should measure precisely what is described in the instructional objective, belong to the same domain and level of difficulty (Queiroga, Andrade, Borges-Ferreira, Nogueira, & Abbad, 2012).

Learning is assessed according to expected results. The generalization occurs in the training environment and is more complex than merely assessing acquisition and retention because it evaluates individuals' KSAs in tests containing tasks different in some aspects (cases or situations) from those worked during training (Pilati & Abbad, 2005). The details are explored in the next section.

2.2.1 Generalization

It is widely known that to achieve instructional objectives, there must be an alignment between instruction and learning assessment (Gulikers, Bastiaens, & Kirschner, 2004). This process needs to be based on instructional objectives focused on the learning and development of KSAs. The purpose of assessing is to verify the acquisition of high order mental processes, not only factual knowledge and concepts comprehension. An assessment with authenticity positively influence learning outcomes and participants' motivation (Herrington & Herrington, 1998). Authenticity, however, is only vaguely described as a dimension of assessment.

Two important reasons for using competencies (KSAs) based assessments are: (1) construct validity and (2) consequential validity – impact on learning outcomes. Construct validity is related to whether the measurement is evaluating what it was supposed to measure. In competency assessment, this means that the task should appropriately reflect what needs to be assessed in terms of KSAs, with the content of assessment involving authentic tasks – representing real-life problems. Basing the assessments on authenticity criteria, construct validity is higher when compared with objective or traditional tests assessments (Gielen, Dochy, & Dierick, 2003).

The crucial aspect of authentic performance assessment is the tasks' fidelity degree to the conditions under which performance will generally occur. Authentic assessment requires learners to use the same skills or combinations of KSAs, using the same criteria they need in workplace situations. In real life, employees usually know the required criteria for their performances. Therefore, authentic assessments must follow this same logic. Learners need to know in advance the quality standards of the products that are expected from them (Gulikers, Bastiaens, & Kirschner, 2004).

Gaps continue to exist regarding data collection and analysis strategies in T&D learning assessments. The learning measurement in T&D should involve the assessment of the KSAs' generalization learned in training – the demonstration of what was learned in situations different from those experienced in the event. The assessment cannot be restricted to assess the acquisition and retention of KSAs, initial phases of storage and maintenance of learning (Queiroga et al., 2012). Learning measures at the generalization level need to be formulated with items that adequately assess the mastery degree of what was taught in the instructional event and depends on expected performances' clear specification. For a learning assessment item to measure the performance described properly, it must belong to the same level of complexity. For more complex ordering KSAs, it is advisable that the items have situational

performance tests format, monographs or reports, problem-solving or new work's solutions creation in the form of open questions (Queiroga et al., 2012).

Typically, learning criteria are measures of training outcomes, though they are not working performance measures. They are typically operationalized using paper and pencil tests or situational performance tests. Behavioral criteria are measures of actual performance in the work environment. They are measures that show and help the identification of training performance effects assessing individuals in their current job function (Arthur, Bennett, Edens, & Bell, 2003).

Although learning and behavioral criteria are conceptually linked, researchers have had limited success in empirically demonstrating this relationship, is due to the reason that behavioral criteria are susceptible to environmental variables that can influence the transfer of training – use KSAs in the work environment (Colquitt, LePine, & Noe, 2000).

Learning assessment can also be situated in the workplace, where training participants perform the KSAs learned or improved in instructional events. The next section approach training of transfer, a learning measure that investigates the individuals' application of KSAs in their job functions.

2.2.2 Transfer of training

Transfer of training is the individual's ability to apply what was learned in situations other than those offered in the instructional event (Abbad et al., 2006a). It is a demonstration of the KSAs learned in situations other than those adopted in the training event and encompasses all the other stages of learning – acquisition, retention, and generalization. This concept is very close to the concept of generalization, but it is placed, situated and measured in the work environment (Abbad et al., 2006a).

The degree to which the trainees apply KSAs acquired in training in workplaces is another transfer of training definition (Wexley & Latham, 1981). In order for acquired KSAs to be transferable, training content must be learned and retained (Kirkpatrick, 1967). Summarizing, transfer of training definition stems from the existence of two conditions: (1) learning generalization for work – extension in which the KSAs are applied in contexts, people or situations different from those trained; and (2) maintenance of trained skills – the extent to which changes resulting from a learning experience persist over time (Baldwin & Ford, 1988; Blume, Ford, Baldwin, & Huang, 2010).

Systematic literature reviews claim that the transfer of training construct is one of the oldest research topics of organizational and industrial psychology (Bell, Tannenbaum, Ford, Noe, & Kraiger, 2017). The questions about the transfer of training remain relevant, and there is academic and organizational demand for more shreds of evidence that indicate the design and execution of effective training initiatives. There is a paradox in organizations because if, on the one hand, results from training investments at the organizational level are viewed positively, on the other, results from individual training events are still viewed with skepticism (Baldwin, Ford, & Blume, 2017).

The transfer of training questions has been the subject of several studies that explored factors affecting organizational outcomes. Researches seek to find strategies to increase the probability that KSAs will be applied in the most diverse contexts and tasks (Ford & Kraiger, 1995). Transfer measures can be taken immediately after training or after a certain period, and it is expected that the relationships are stronger the closer the measures are from the training experience, both in the physical and temporal context (Barnett & Ceci, 2002). The transfer of training has been measured as the amount of use of the KSAs trained, and also as the effectiveness of its application. Studies measuring the effectiveness aspect are believed to be more consistent than those that measure the amount of use. The last interest of researchers and organizations is usually in the useful application of the training KSAs and not in its simple use, which may or may not result in positive outcomes (Blume et al., 2010).

Early research on transfer of training asserted that its effects are maximized when there are identical stimuli and response elements in training and transfer contexts (Thorndike, & Woodworth, 1901). This behavioral view advocated that teaching through general principles facilitates transfer when trainees are taught in general rules and theoretical principles underlying the content being trained, not just applicable KSAs (McGehee & Thayer, 1961). The variability of relevant stimuli is another premise to maximize the positive transfer of training (Ellis, 1965).

Moreover, there is a longing for studies on instructional methods and strategies that explicitly target the transfer of training, not just immediate learning as a dependent variable (Yelon, Ford, & Anderson, 2014). Baldwin and Ford (1988) organized a model, widely adopted by scholars in the field, in which the transfer of training is related to the following dimensions: characteristics of the trainees, training design, and work environment. Although the training design has received plenty attention from researchers, there is still a lack of the fundamental element when it comes to instructional design and transfer of training – the relevance of the instructional objectives or training program goals concerning the explicit transfer. These objectives refer to the expected transfer of training results, which should define the choice of

assessment criteria (Blume et al., 2010). It is vital to use different designs for the initial learning processes (acquisition and retention) and transfer of training, applying design elements that can effectively increase its outcomes (Ford, Baldwin, & Prasad, 2018). Concluding, it is crucial to optimize the transfer process, renewing research that investigates the principles of learning, now with an explicit focus on transfer and not only on learning at initial levels (Baldwin, Ford, & Blume, 2017).

So far, there was a theoretical framework talking about T&D structure and how organizations assess learning at the generalization and transfer of training levels. The explanations were essential before moving on to instructional design approaches. Instructional design is a previous step to assessment in T&D solutions development. Although, instructional design fundaments are be enlightened in the following section because of it closely connected with gamification (the central theme of this research) that is applied as an instructional strategy in the selected training.

2.3 Instructional design theories

Learning theories are responsible for understanding how the basic individual learning processes occur and how they influence the instructional design, enabling the elaboration of conditions and situations to promote learning (Abbad et al., 2006b). Learning theories help the understanding of why an instructional design works and serves as the basis for the construction of instructional theories (Reigeluth, 1999).

Instructional theories exist to understand how individual differences interact with instruction and with contexts to produce learning outcomes. They are instructional approaches linking content, procedures, and methods. Reigeluth (1999) affirms that instructional theories have their importance because they are easier to apply to solve learning gaps, helping to improve the teaching-learning situations' design.

Instructional design theories aim to guide the choices and have a prescriptive characteristic – their goal is to facilitate instructional objectives' achievement. Alternatively, learning theories provide the rationale for why these prescriptions are useful, thus having a descriptive characteristic (Abbad et al., 2006b; Reigeluth, 1999).

Instructional design theories help T&D professionals to improve the design of teaching-learning situations. Changes occurred in society, and teaching techniques have impacted the instructional design theories. Professionals should acquire and maintain increasingly complex

KSAs for problem solving, teamwork, self-management, and career planning and learning to learn (Abbad et al., 2006b).

Instructional design theories are constituted by two components: (1) methods to facilitate learning and human development; (2) situations that these methods should be applied (Reigeluth, 1999). The methods are going to take into account the conditions under which instruction will take place, the nature of the KSAs to be developed, the profile of the individuals, the characteristics of the environment and the constraints imposed on the development of the instruction. Instructional design should also indicate expected learning outcomes in terms of effectiveness and efficiency, the degree of satisfaction with the experience, and long-term effects – e.g., impacts on individual's performance in the job (Abbad et al., 2006b).

In summary, an instructional design theory should be: (1) oriented toward creating and achieving objectives, (2) prescriptive – indicating methods to achieve expected objectives, (3) situational – influenced by instructional conditions and expected outcomes, and (4) flexible - working with the idea that methods are probabilistic and nondeterministic.

After visualizing the theories that underlie instructional design application, the next section moves to a practical approach, showing steps required to design instructions. It addresses the instructional objectives writing and also the creation and selection of instructional strategies.

2.4 Designing instruction

Traditional ways of designing instruction mostly understood learners as passive agents acquiring only KSAs that should be appropriate perform work functions facilitating a direct transfer of training (Noe, Clarke, & Klein, 2014). One of the most widely known approaches to design training is the Instructional System Design (ISD), a systematic way to design learning solutions (Gagne, 1992). The ISD model usability is recognized, but still, it needs to be reviewed and adapted to emerging approaches centering learning on trainees (usually involving technology), placing educators in a background mediating position. The increase of informal and self-directed learning did not remove the relevance of formal training events. Many important learning issues yet need to be investigated, especially about the instructor’s role and the type of KSAs best learned online, in the classroom, or hybrid approaches (Noe, Clarke, & Klein, 2014).

The instructional design can be divided into six stages: (1) writing instructional objectives; (2) choose the training delivery method; (3) establish the sequence of

objectives/contents; (4) select or create instructional strategies; (5) define learning assessment criteria; and (6) testing the instructional design (Abbad, Zerbini, Carvalho, & Meneses, 2006c). The next sections address the theoretical aspects of the instructional objectives' creation and the instructional strategies' development or selection.

2.4.1 Defining instructional objectives

A result of the training needs assessment, and also the instructional design's first stage is the specification of the instructional objectives – originated in the needs assessing showing lack of KSAs that need to be trained. KSAs can be classified into three broad categories: cognitive, interpersonal, and psychomotor. Cognitive tasks are related to thinking, generating ideas, understanding and solving problems. Interpersonal are those associated with interacting with others in a workgroup, or with internal or external clients. The types of KSAs and how they will impact the effectiveness of training should be considered – there must be a connection between training delivery, method and task being trained (Arthur, Bennett, Edens, & Bell, 2003).

An instructional objective should represent performance, and therefore, the verb must have limited meaning, being possible to be measured at the assessment. The performance must be composed by a verb, representing the action, and an object that undergoes the performance. Two other components integrate the structure of an instructional objective: condition – situation or environment where the performance occurs; and criteria – performance's standard or desired quality (Abbad et al., 2006c).

Instructional objectives should explicitly inform the participants' the expected learning outcomes. They provide the basis for designing the learning environment and the assessment (Goldstein & Ford, 2002). The writing of the instructional objectives is going to specify to the learners what is expected from them at the end of the training experience. These objectives should be composed of actions represented by measurable verbs such as describe, identify, apply, analyze and formulate. Each sentence built must be connected with some KSA. After identifying the objectives, the designers plan the sequence of objectives and content to facilitate learning. At this stage are the decisions about instructional strategies and the activities format (Goldstein & Ford, 2002). The ISD model emphasizes that each instructional strategy, going from readings to practical activities, must work in an integrated way.

Bloom's taxonomy systematizes the acquisition of KSAs in three domains: cognitive, psychomotor and attitudinal. Table 1 summarizes the six levels of the cognitive domain (Bloom, 1956).

Table 1
Bloom's taxonomy - Cognitive domain

Category	**Definition**	**Verbs examples**
Knowledge	Recall data or information	Identify, list, and reproduce.
Comprehension	State a problem in one's own words	Distinguish, translate, and associate.
Application	Apply a concept in a new situation	Calculate, operate, and manipulate.
Analysis	Separate into parts, understanding its organization structure	Discriminate, infers, and deconstruct.
Synthesis	Create a new meaning or structure from diverse elements.	Design, rewrite and modify.
Evaluation	Judge ideas or material's value	Conclude, justify, and criticizes

Source: Adapted from Bloom (1956)

2.4.2 Instructional strategies' choice or creation

Instructional strategies are all operations, events, or learning situations created or chosen in instruction design aiming to facilitate the learning acquisition, retention, generalization, and transfer. Instructional strategies are operations necessary to produce instructional outcomes (Borges-Andrade, 1982). Fontanive (1982) recommends that designers should observe the type of performance involved in the instruction before the creation or choice of these strategies. It is essential to observe the motivational advantages and disadvantages of each available strategy, the level of complexity of expected behavior, and the cost involved in choosing. The author emphasizes the importance of procedures diversification, aiming to maintain motivation levels in the experience.

The choice of the best instructional strategies must be based on the instructional objectives' nature. To illustrate, if the objective requires interaction between participants to

demonstrate a certain verb present in the objective, the activity must, in some way, promote contact between participants. Dramatization, behavioral modeling, and role-playing are examples of interactive strategies (Abbad et al., 2006c). Developing KSAs that involve teamwork demands learning KSAs with actions as cooperating, synchronizing and sharing experiences. Simulating problem-solving and team decision making may be appropriate strategies for this type of competence. Rather, if the aim is to understand concepts, a dialogued presentation should be a sufficient strategy. Abbad et al. (2006c) cite how strategies, procedures, techniques or instructional approaches can be adopted in the development of T&D solutions: (1) use of team cooperative projects; (2) games with competitive challenges; (3) real situations or processes' simulation; and (4) behavioral modeling based on the Social Learning Theory (Bandura, 1977).

Because there is no single, proven, most effective strategy for creating training, designers must continually seek new methods to facilitate learning. With the ongoing technological advances in methods and learning theories, more options are available when choosing strategies to train – gamification is one example. Information delivery needs to be differentiated from training, which is focused on the job application of this information (Ittner & Douds, 1997). The training method is a set of systematic procedures, activities or techniques designed to impact the participants' KSAs and have direct utility in improving their work performance. Thirteen central methods were identified for instruction: case study, game-based learning, supervised internships, role rotation, role mirroring, reading, programmed instruction, role modeling, role-playing, simulation, stimulus-based training, and team training (Martin, Kolomitro, & Lan, 2014). The authors affirm that other methods are mere extension or subcategories of these mentioned.

The learning mode is the communication form by which content is given to learners: learning by doing, by watching or by listening. When trainees acquire or develop KSAs performing a task, they are using learning by doing mode. This approach is aligned with the educational philosophy of experiential learning (Kolb & Kolb, 2005).

Some other aspects were identified by researchers to help design training events (Martin, Kolomitro, & Lan, 2014):

1. Environment: location in which the training takes place.
2. Presence: the participant is necessitating the presence of an educator or another source – e.g., computer.
3. Proximity: distance between training and participants –e.g., face-to-face or online.

4. Interaction: level of interaction between training and the participants – also participants between themselves.
5. Costs: identification of the most significant expenditures associated with each particular method of training delivery – also considering, in addition to the initial expenses, ongoing expenses throughout the process.
6. Time demand: time required of the trainees. In this context, it is essential to consider how duration can impact trainer, trainees and the organization.

At the end of this section, essential topics were discussed such as the main theoretical foundations on T&D, including the concepts of learning in organizations, the different levels of expected results in T&D, aspects about learning assessment at the generalization and transfer of training levels, and finally the fundamentals to create an instructional design.

The next section shows the results of the literature review on the central theme of this study – gamification. It was crucial to approach the topics discussed so far since gamification is part of this research within a context of T&D. In this research, gamification is an instructional strategy, so that it is later possible to measure its effects on generalization and transfer of training.

3 Literature review

In order to obtain a systematized view of the scientific production on gamification in learning contexts, a search was done based on the findings of other literature reviews on the subject (Dichev & Dicheva, 2017; Hamari et al., 2014; Seaborn & Fels, 2015). The research was done in all databases available in Brazilian Capes' periodical portal (periodicos.capes.gov.br). The keywords used for the research were **gamif*** for **gamification** and its variations of uses (gamified, gamify, and others) and **learning** or **training** or **education** (the search was made to obtain the results in two languages, English and Portuguese). Only peer-reviewed, article type documents were selected. The count of articles found in Portuguese was 32 studies and in the English language was 3,486 studies. After receiving the results, a criterion was applied for all publications restricting to articles in journals with Qualis Capes[1] A1, A2, B1 or B2, or impact factor[2] greater than two, when there was no Qualis Capes rating.

[1] Qualis is the set of procedures used by Capes to stratify the quality of the intellectual production of graduate programs. Retrieved on February 9th, 2018. website: http://www.capes.gov.br
[2] The rating used was accessed in https://www.scimagojr.com/

These criteria were adopted because most of the articles analyzed at a first moment were of areas related to Computer Science, and did not have adequate educational theoretical basis to support this research.

Additional exclusion criteria were used by removing from the analysis articles that addressed gamification in contexts where learning was not the focus of the research (health improvement studies). We also excluded articles in which the gamification presented results related to learning though did not occur in formal learning environments (e.g., research related to fitness improvements).

For this research, were included only empirical articles that applied gamification designs collecting participants' data – 42 in total. Besides, literature reviews and theoretical articles were analyzed whose main objective was to develop gamification models for one or more specific game elements, totalizing 33 articles. The theoretical articles were analyzed so that it could have enough theoretical and empirical background to apply the gamification as an instructional design strategy. After the analysis of the 75 studies, some cross-references were considered essential and were included in this study to conceptualize specific game elements in gamification design.

The next section describes the findings from the 42 empirical articles analysis. Afterward, theoretical aspects on gamification were elaborated, with a proposal to organize the game elements eligible to be applied in a financial institution managers' training event, study context of this research.

3.1 Gamification in the learning context

Active learning methodologies are a type of strategy that introduces practical activities aiming higher student engagement (Prince, 2004). Students who participate in interactive activities learn concepts better, and hold them longer, applying them more effectively in other contexts, when compared to students who experience passive instructional designs (Freeman et al., 2014). Gamification can be applied in learning contexts in two ways: (1) partially gamified; and (2) fully gamified. The first consists of applying game elements to a single component of a course (e.g., a unit or the assessment), while the second comprises the application of gamification throughout the entire experience, completely changing the instructional design. Learning from a fully gamified experience does not focus on changing one component, but on creating a new learning experience where most instructional strategies need to be redesigned (Barata, Gama, & Gonçalves, 2016).

In the attempt of approaching gamification from learning theories, there are more common elements with the Behaviorist Learning Theory (Skinner, 1938), that uses positive reinforcements, small step-by-step tasks and contingent feedback as strategies (Ding, Guan, & Yu, 2017). When considering that gamification has the potential to use specific digital games' elements, it approaches to Connectivism (Siemens, 2013). Two characteristics can be highlighted by placing gamification close to it: (1) learning networks with community's reinforcements; and (2) the capability of selecting learning content is as important as the knowledge acquired. Gamification may use assessment mechanisms based on community reinforcement. In this model, the participants themselves make the assessment, similar to what happens in multiplayer online digital games. Gamification is also capable of offering diversified learning paths since the emphasis is on small achievements. Finally, gamification emphasizes the visual learning dimension, especially through visual feedback mechanisms, allowing participants to observe their progress while the experience is occurring (Ding et al., 2017).

In the next section are the results of empirical articles analyzed. This evaluation was essential to determine the methodological approach chosen for this study.

3.2 Empirical articles main goals

In this section, the main research objectives found in the empirical articles are going to be presented. Some of the research founded were interested in investigating specific game elements. The points' systems were studied in its simplest form, numerical values (Attali & Arieli-Attali, 2015), but also in a more sophisticated way – represented by virtual currencies (Filsecker & Hickey, 2014). Leaderboards comparing participants' performances were too studied in some articles found (Christy & Fox, 2014; Landers & Landers, 2015; Nebel, Beege, Schneider, & Rey, 2016).

In terms of variables and constructs, gamification research in learning context has tried to discover influences in individual's learning performances – measuring learning itself or influencing variables that could have moderation or mediation effects. Many scholars have shown an interest in investigating motivation and engagement (Alcivar & Abad, 2016; Dominguez et al., 2013; Filsecker & Hickey, 2014; Hanus & Fox, 2015; Ibanez et al., 2014, Pettit et al., 2015, Tan & Hew, 2016). Others sought to observe psychology constructs as cognitive load (Su & Chen, 2015), attention status (Auvinen, Hakulinen, & Malmi, 2015), personality trait, learning styles (Buckley & Doyle, 2017) and self-efficacy (Adukaite et al., 2017, Tan & Hew, 2016).

It is essential to explore further the engagement concept, a construct so often measured in gamification research, and this is done under two perspectives. The first comes from the context of web applications since most gamification studies are set in technological contexts. User engagement, in this setting, is the emotional, cognitive and behavioral relation existing between a user and a resource (Attfield et al., 2011). Secondly, the educational view defines engagement as a construct encompassing: (1) participation or effort – behavioral aspect; (2) interest – affective aspect; and (3) psychological involvement – cognitive aspect (Fredricks, Blumenfeld, & Paris, 2004). Both perspectives are aligned with the KSAs definition adopted in the T&D area and are appropriate to our further discussions.

Regarding the learning assessment complexity level (Bloom, 1956) 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, Park, Liebert , & Lau, 2015), although they were minor compared to lower complexity levels.

The different participants' skill levels were investigated, seeking to find its impacts on gamification outcomes (Hamari et al., 2016; Santhanam, Liu, Shen, & Santhanam, 2016; Wang, Chen, & Chan, 2016). Most of these studies included competition – and the impacts depending on learners' KSAs levels (Hanus & Fox, 2015).

After presenting some of the goals, variables, and constructs found in the literature review, the next section summarizes the main findings on the investigated gamification empirical articles.

3.3 Gamification' studies results

In this section, the results obtained in the empirical research analyzed are discussed. Firstly, studies indicated that gamification may be effective in increasing engagement levels demonstrated through interest, measured by self-perception or data collection reporting the player's number of interactions with the learning content (Ambrosio & Garofalo, 2016; Dias, 2017; Hamari et al., 2016; Hew et al., 2016; Huang & Hew, 2018; Ibanez, et al., 2014; Landers & Landers, 2015; Mekler, Brühlmann, Tuch, & Opwis, 2015; Paiva et al. 2016; Pettit, et al., 2015; Sun & Hsieh, 2018; Tan & Hew, 2016; Tenorio et al., 2016; Yildirim, 2017).

When comparing gamification with other learning strategies, such as the use of social networks, one study obtained negative results in the users' participation in the gamified environment (De-Marcos et al., 2014). One possible explanation for the difference in these results may be associated with the challenges' difficulty level proposed by gamification design.

The engagement can be higher when there is a balance between skill and challenge levels, always maintaining the difficulty of the challenge as the player evolves in the experience (Hamari et al., 2016; Su & Chen, 2015; Tan & Hew, 2016). Studies that measured satisfaction obtained positive results, indicating a preference for learning experiences using gamification (Alcivar & Abad, 2016; Armstrong & Landers, 2017; Ding et al., 2017; El Tantawi et al., 2016; Graafland et al., 2014).

Two studies went in the same direction revealing that the use of gamification generated better learning performance when the focus was on practical concepts' applications. On the contrary, these same researches revealed that gamification had not yielded positive results in terms of factual or conceptual learning – even in written assessments (De-Marcos et al., 2014, Dominguez et al., 2013). On the other hand, two articles found positive effects associated with theoretical knowledge acquisition (Filsecker & Hickey, 2014; Ibanez et al., 2014).

Regarding the most common progression mechanics present in the analyzed studies, points, badges and leaderboards (PBL), it is noticed that the research does not have consensual results. The points system as a reward tool did not show significant differences as to the quality of responses, though it reflected on players' speed to solve tasks (Attali & Arieli-Attali, 2015, Su, 2015). In turn, badge systems, more elaborate feedback and recognition mechanism, due to the more specific distribution criteria, presented positive outcomes in engagement levels (Davis & Singh, 2015; Ibanez et al., 2014).

Leaderboards are one of the game elements more often used the empirical articles analyzed. In one specific study, there was empirical support pointing out the causal relationship between leaderboards' use and time spent performing tasks (Landers & Landers, 2015). Conversely, negative effects were also found when using leaderboards, worsening individuals' motivation and reflecting negatively on learning performance (Christy & Fox, 2014; Hanus & Fox, 2015).

The competition was able to generate positive engagement, but only for profiles of individuals with a preference for this type of dynamics (Pettit et al., 2015). The competition was considered challenging and motivating only when the effort was rewarded, not only the skill level (Landers & Landers, 2015).

For some scholars, social interaction functionalities can be fundamental to create sustainable and engaging gamification (Koivisto & Hamari, 2014). Researchers applied social networks to create small worlds, with cooperative structures similar to those found in real-world networks and multiplayer video games. Social gamification may be able to create more cohesive

and purpose-driven learning communities, providing relevant information about individuals' behavior when acting as a team (De-Marcos et al., 2016; Pettit et al., 2015).

Due to their degree of sophistication, some gamification designs are still rarely used in empirical research. Content randomization systems, bringing up the surprise, uncertainty, and curiosity, are still barely present in gamification designs (Pettit et al., 2015). Autonomy and freedom of choice, are also rarely seen in empirical studies (Nebel et al., 2016). The customization of paths, combined with recommendation systems, which could increase interactions associated with collaborative learning, were present in only one study (Paiva et al., 2016). The novelty effect was verified in some studies. Over time, there was a decrease in satisfaction, participation, perceived fun and usability (El Tantawi et al., 2016; Hanus & Fox, 2015; Koivisto & Hamari, 2014).

In many studies, gamification did not show positive learning performance outcomes but was able to increase satisfaction, engagement, and motivation. Accordingly, gamification should not be introduced isolated in the learning environments, but balanced and integrated, complementary to others learning strategies (Buckley & Doyle, 2017, De-Marcos et al., 2016 and Yildirim, 2017). Table 2 summarizes the results explained in this section.

Table 2
Empirical articles' results

Effect	**Construct**	**Authors**
Positive	Learning performance	(Alcivar & Abad, 2016; Auvinen et al., 2015; De-Marcos, et al., 2014; De-Marcos, et al., 2016; Dominguez et al., 2013; El Tantawi et al., 2016; Filsecker & Hickey, 2014; Ge, 2018; Hamari et al., 2016; Ibanez, Di-Serio, & Delgado-Kloos, 2014; Smith, 2017)
	Satisfaction	(Alcivar & Abad, 2016; Armstrong & Landers, 2017; Ding et al., 2017; El Tantawi et al., 2016; Graafland et al., 2014)
	Engagement	(Ambrosio & Garofalo, 2016; Dias, 2017; Hamari et al., 2016; Hew et al., 2016; Huang & Hew, 2018; Ibanez, et al., 2014; Landers & Landers, 2015; Mekler, Brühlmann, Tuch, & Opwis, 2015; Paiva et al. 2016; Pettit, et al., 2015; Sun & Hsieh, 2018; Tan & Hew, 2016; Tenorio et al., 2016; Yildirim, 2017)

Effect	Construct	Authors
	Motivation	(Nebel et al., 2016; Tan & Hew, 2016; Sun & Hsieh, 2018)
Negative	Learning performance	(De-Marcos et al., 2014; Dominguez et al., 2013; Hanus & Fox, 2015)
	Engagement	(Dominguez et al., 2013)
Negative	Motivation	(Hanus & Fox, 2015)
	Learning performance	(Attali & Arieli-Attali, 2015; Hew et al., 2016; Nebel et al.2016; Prestopnik et al., 2017)
No effect	Engagement	(Filsecker & Hickey, 2014; Monterrat, Lavoué, & George, 2017)
	Motivation	(Filsecker & Hickey, 2014; Landers, Bauer, & Callan, 2015)

Source: Author's elaboration

After exploring the empirical article's results, it is essential to evaluate the research gaps that remain. In the next session is pointed out the limitations and research agenda advised by the gamification in learning contexts research.

3.4 Gamification's studies limitations and research agenda

Based on the empirical articles' analysis, it was possible to perceive potential in applying gamification promote learning, as well as weaknesses that still need to be further explored empirically. Firstly, it is vital to highlight the difficulty in isolating game element's effects when implementing complex designs with multiple elements (Attali & Arieli-Attali, 2015; De Marcos et al., 2016; Santhanam et al., 2016).

There is also the need to assess more complex cognitive levels since most of the research found measured learning using multiple choice items. Indeed, this can be due to the designers' difficulty to create and have access to environments where complex knowledge is developed (Attali & Arieli-Attali, 2015, Lin et al., 2015, Tan & Hew, 2016).

Some methodological issues need to be addressed as the instruments' lack of semantic validation by judges, self-report bias to assess learning performance and restricted research settings – frequently with students in schools and universities (Buckley & Doyle, 2017; Christy & Fox, 2014; Graafland et al.; Wang et al., 2004).

Researchers' challenges include the use of assessments for more complex cognitive levels, using tools more appropriate to what they are intended to measure (Landers & Landers, 2015; Tan & Hew, 2016). Research is also to investigate gamification effects' in longitudinal designs, applying other sources of data – speech and video recordings (Christy & Fox, 2014; Hamari et al., 2016; Yamabe & Nakajima, 2013). Scholars and organizations should be interested in investigating how gamification can influence workers behavioral change, measuring with situational performance assessments (Graafland et al., 2014).

Gamification designs should provide a better understanding of how to use reward systems. The choice of rewards mechanisms deserves further study in gamification since extrinsic rewards is a topic of debate, especially when related to motivation and the promotion of competition among participants (Dominguez et al., 2013; Filsecker & Hickey, 2014; Hanus & Fox, 2015). Whenever possible, gamification designs should test more cooperative dynamics (Hanus & Fox, 2015).

About players' profiles, there is a lack on research interested to discover individual's preferences and repertoires prior to the gamified experience, especially those who have negative attitudes and perceptions regarding gamification (Buckley & Doyle, 2017; Christy & Fox, 2014; El Tantawi et al., 2016; Filsecker & Hickey, 2014; Ibanez et al., 2014; Koivisto & Hamari, 2014).

Finally, it is necessary to investigate the strategies of commercial digital games in order to identify possible applications in the learning environment, bringing the success elements of this industry to this field that is continuously focusing on develop new strategies to engage, motivate, entertain and to solve learning problems in the most diverse environments and areas (Prestopnik et al., 2017; Su, 2015)

Next section describes game designs adopted in the empirical research investigated. This section is important as a preparation for the next steps of this study, which addresses the game elements organization genuinely. Therefore, it is possible to design instruction to achieve the proposed goals using a quasi-experiment.

3.5 Empirical gamification designs

Distinct gamification designs were used in the empirical studies analyzed. This section shows some of the details of the designs used by scholars. Gamification has often been applied in question-and-answer games, especially with the use of points systems. Frequently, environments are built to increase the immersion level of participants in the experience

(Adukaite et al., 2017, Attali & Arieli-Attali, 2015, Christy & Fox, 2014; Prestopnik, Crowston, & Wang, 2017; Santhanam et al., 2016). Technological instruments are supporting gamification, were applied as a strategy to promote collective participation. In such cases, response time was also a criterion to reward good performance (Graafland et al., 2014; Pettit et al., 2015).

Points, badges and leaderboards, known in the literature as the PBL elements, were the most frequent design element used in the studies found and analyzed. Points systems are the most simple feedback mechanism, though it can also be applied in a more sophisticated way, transforming points in coins or medals that can be further exchanged. Some researchers created specific currencies to reward good behavior in the classroom, and others transformed points in experience points (XP) – a strategy present in commercial video games (Barata et al., 2016; Buckley & Doyle, 2017; Dominguez et al., 2013; Hanus & Fox, 2015). Badges are a more elaborated form of feedback, given when a specific or particular task is achieved. Leaderboards were also used when the objective was to promote competition between participants (De-Marcos, Garcia-Lopez, & Garcia-Cabot, 2016; Davis & Singh, 2015; Hanus & Fox, 2015; Tenorio et al., 2016).

Immersion was a strategy adopted to bring the participants closer to the application context of the KSAs worked during the experience. Virtual characters were also present, in this same context of approximation with reality, establishing dialogues with the participants of the gamified experience (Ambrosio & Garofalo, 2016; Filsecker & Hickey, 2014; Lin et al., 2015). Another element used in designs is the avatars (players' symbolic representation), especially allied with role-playing games strategies (Su, 2015; Wang et al., 2016).

As already discussed, many researchers associate gamification necessarily with digital contexts. Many studies were conducted on online platforms – virtual learning environments (Barata et al., 2016; De-Marcos et al., 2014, 2016. El Tantawi et al. 2016, Ibanez et al., 2014). Although most of the research was developed on adapted platforms, one study explored learners' interactions in a native gamified platform (Paiva, Bittencourt, Tenorio, Jaques, & Isotani, 2016). This type of digital learning environment allows scholars to collect an impressive amount of data, with more possibilities to measure participants' behaviors. Gamification research is typically conducted using quantitative data. Mixed or hybrid learning contexts, combining classroom and digital platforms, are more suitable to investigate gamification outcomes qualitatively, exploring more complex knowledge, skills, and attitudes (Barata et al., 2016; Filsecker & Hickey, 2014; Yildirim, 2017).

Next section is a synthetic organization using Mechanics, Dynamics, and Emotions (MDE) gamification model (Robson et al., 2015). The effort made was to fit game elements found in the literature in MDE definitions. This theoretical section is a result literature review and aims to help this research to build the instructional design, applying gamification as an instructional strategy.

4 Designing gamification

The next theoretical sections propose to deepen the fundaments of gamification design. When the literature defines that gamification is the use of game elements, it does not specify what elements these are and how they should be integrated to create an environment and experience that facilitate learning. The organization of the elements proposed is authored by the researcher, based on a gamification model, which is discussed below.

Robson et al. (2015) adapted the electronic game developers' literature (Hunicke, Leblanc, & Zubek, 2004) to define the three guiding fundaments of gamification: mechanics, dynamics and emotions (MDE). The MDE model proposed by the authors is described below and used to substantiate the identification of gamification elements in this literature review.

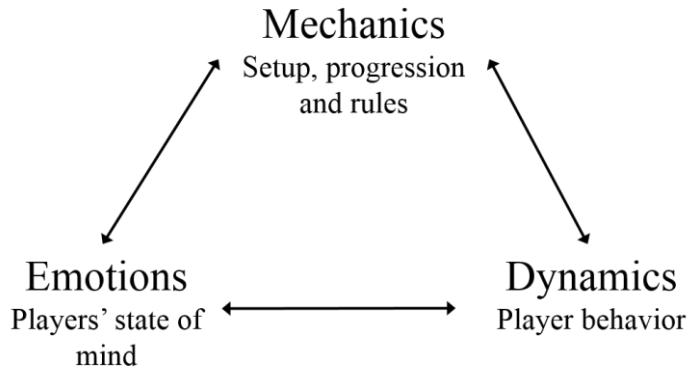

Figure 1. MDE model
Source: Robson et al. (2015).

4.1 Gamification mechanics

Mechanics are the set of decisions designers make to define the objectives, rules, structure, context, types of interaction, and the boundaries that are integrated into the gamified experience. Usually, gamification mechanics remain constant throughout the experience but may change during or after if desired. There are three different kinds of mechanics: setup,

progression, and rules (Robson et al., 2015). The next sections present each mechanic, trying to connect the game elements found in the literature and their respective definitions.

4.1.1 Setup mechanics

Robson et al. (2015) argue that the setup mechanics are composed of game elements that shape the entire gamification – defining the objects that are needed and distributed in the environment. Players interact with these elements chosen and built during the design process (Elverdam & Aarseth, 2007). The setup mechanics indicate the objects that compose the experience, as well as the characters (fictional or real), and how the participants are integrated into the other elements. Another decision to be made is about the environment, space and its characteristics, where the experience occurs – e.g., if it is going to happen in a real or a virtual environment (Robson et al., 2015). Overall, the game elements analysis allowed to classify setup mechanics into two groups: objects and environments.

Objects are the elements that players interact directly or indirectly (Elverdam & Aarseth, 2007). From this definition, three types of objects are in this category: (1) artifacts – inanimate elements created by designers or players (e.g. cards with information or coins for collection) (Elverdam & Aarseth, 2007); (2) characters – elements with human or other beings characteristics (e.g. monsters, heroes, villains) (Plass, Homer, & Kinzer, 2015); (3) players – avatars can represent gamification participants symbolically, through real or fictitious characters (Bharathi, Singh, Tucker, & Nembhard, 2016; Kapp, 2012).

Scientific research provides evidence that individuals are socially influenced by anthropomorphic automated agents, more commonly known as avatars (Kapp, 2012). The avatars' use can facilitate the application of learning content, since there is a social representation and desired behaviors' projection. If an avatar that looks like the individual performs an activity in a certain way, there can be an influence on the same activity performance in the future – behavioral changes that occur in gamification can be transferred to the workplace (Kapp, 2012).

The second setup mechanics group is the creation of gamification's environments, scenarios. In this group are all physical or virtual spaces where artifacts, characters, and players are distributed for interaction (Plass et al., 2015). According to Kapp (2012), the environments built for a learning gamification solution should resemble the real work environment in which

the expected behavior will be demonstrated. The author asserts that situating the individual in a realistic environment increases the transference likelihood to maintain learned behaviors.

Table 3
Gamification setup mechanics

Elements	Definition	Examples
Artifacts	Inanimate elements created by designers or players (Elverdam & Aarseth, 2007)	Cards, coins, dices
Characters	Elements with human or other beings characteristics (Plass et al., 2015)	Monsters, heroes, villains
Players	Automated anthropomorphic agents - players' symbolic representations (Bharathi et al., 2016; Kapp, 2012)	Avatars
Environments	Physical or virtual spaces where artifacts, characters, and players are distributed for interaction (Plass et al., 2015)	Real and virtual worlds

Source: Author's elaboration

4.1.2 Progression mechanics

They are the design decisions that create elements to dictate the experience evolution as players interact with setup mechanics (Elverdam & Aarseth, 2007). In learning context gamification, progression mechanics are especially important because they define the feedback mechanisms present in the experience. They signalize the actions crucial to move towards the goal, modeling the behaviors that will be rewarded and that must be repeated in the future. The feedback system helps participants stay on track towards the goal, avoiding possible mistakes, updating progress and informing the path already taken and what remains to be achieved. It provides the players' success signal (Barata et al., 2016).

Most of the game elements found in the literature review were associated to this section, though it is important to mention that progression mechanics hardly will be useful acting in isolation since they need to work in conjunction with the setup and rules mechanics (Robson et al., 2015). Next, the most usual progression mechanics, PBL elements, are presented.

(1) Points system. Numeric variable players can gain or lose while interacting with setup mechanics. The acquisition and loss of points have defined criteria in the rules mechanics

and measure players' performances. They may relate to the collection of artifacts or mean status in gamification. Points may be required to unlock contents and be used as an exchange item (Dicheva, Dichev, Agre, & Angelova, 2015; Hew et al., 2016).

(2) Badges. Visually represented by icons, badges reward special achievements in recognition to player's performances and progress (El Tantawi et al., 2016, Hanus & Fox, 2015, Hew et al., 2016, Kapp, 2012, Tan & Hew, 2016).

(3) Leaderboards. Usually visible to all players, exhibiting and comparing performances, are often used to promote competition (Christy & Fox, 2014, Dicheva et al., 2015, El Tantawi et al., 2016, Hanus & Fox, 2015).

Due to their use without design criteria, they are criticized for being called Pointsfications (Bogost, 2011), due to its indiscriminate use without being linked with gamification objectives.

In addition to the PBL elements, levels and progress bars are often mentioned in the literature. Levels refer to the player's domain in a given task (Tan & Hew, 2016). In learning environments, levels are essential in categorizing challenges by the level of difficulty (Hanus & Fox, 2015). Progress bars, in turn, function as an individual feedback mechanism, a graphical representation that measures how much has already been achieved and the path that still needs to be achieved, similar to rankings, but privately shown (Dicheva et al., 2015, El Tantawi et al., 2016). Table 4 presents a summary of the game elements.

Table 4
Gamification progression mechanics

Element	**Definition**	**Examples**
Point system	Numerical variables players can gain or lose while interacting with setup mechanics (Dicheva et al., 2015, Hew et al., 2016)	XP (Experience points)
Badges	Iconic elements meaning special achievements (El Tantawi et al., 2016; Hanus & Fox, 2015; Hew et al., 2016; Kapp, 2012; Tan & Hew, 2016).	Medals and trophies
Leaderboards	Progression elements visible to all players exhibiting and comparing performances, usually promoting competition (Christy & Fox, 2014, Dicheva et al., 2015, El Tantawi et al., 2016, Hanus & Fox, 2015).	Grade boards

Element	Definition	Examples
Progress bars	Graphical representation measuring players' advance and how much is left to achieve (Dicheva et al., 2015, El Tantawi et al., 2016).	Life and energy bars

Source: Author's elaboration

4.1.3 Rule mechanics

These mechanics need to be close to gamification purposes (Elverdam & Aarseth, 2007). They define permissions and restrictions (e.g., time and space) that limit players' actions, creating situations that direct interactions (Robson et al., 2015).

Permission rules allow players to practice actions to achieve their goals. Goals here are objectives defined to participants inside the gamification environment, not other goals such as strategic (what organizations expect when using gamification) or learning (what is expected from learners in terms of knowledge acquired). Designers may use permission rules intended to allow players to make inadequate actions, encouraging a trial and error process to promote learning (Robson et al., 2015).

Alternatively, there must be restriction rules, impositions limiting players' actions. The primary role of these rules is to allow players only to perform the actions listed in the permission rules. Restrictions prevent players from performing undersigned behaviors that may push them away from gamification goals. Restriction rules can also encourage players to act in line with what was planned, being forced to act as designers intended (Robson et al., 2015).

The use of uncertainty rules, similar to random mechanisms present in chance games, can create randomization to maintain the players' engagement level, leading to the emergence of feelings related uncertainty and surprise (Robson et al., 2015).

The rules, as mechanics elaborate by designers, are connected with the other two mechanics presented previously: (1) structure rules: indicate the restrictions and permissions associated to players' interactions with objects and environments; and (2) progression rules: define the criteria for advancing towards goals. Rule mechanics are linked to objectives and designed to promote specific behaviors required to unlock subsequent challenges, previously unavailable (Robson et al., 2015).

Usually, rules prevent players from taking easy or obvious paths towards goals. Consequently, rules choice or creation is vital as a stimulus to players' strategy development –

promoting creative thinking. Rules are essential to gamification as they encourage players to understand them to achieve mastery in the experience (Cain & Piascik, 2015).

The mechanics presented in this section represent the fundamental basis for developing gamification since they define the use of the main game elements. Notwithstanding, develop gamification mechanics not enough to create an experience that will promote behavioral change through learning. The MDE model is composed of dynamics and emotions, manifest as consequences the interaction with the mechanics, in the form of players' behaviors.

4.2 Gamification dynamics and emotions

The MDE model presents in separate sections dynamic and emotions in gamification (Robson et al., 2015). Assuming that both are consequences, although different, of players' interactions with the gamified experience, here in this section they are conceptualized together. It is a common sense that dynamics and emotions cannot be fully predicted throughout the design. Some behaviors flourish as participants' strategic actions and thinking, arising unexpectedly during the experience – e.g., the bluff, a common strategy in some card games (Camerer, 2003). Designers' challenge is trying to predict the dynamics and emotions that emerge as behaviors, for only then develop the gamification mechanics aligned with the intended goals. Anticipate the behaviors' presence in the experience, is a design task that should be done by writing what is expected.

Dynamics manifest in players' contact with gamification structure (including other players) and can manifest themselves as an individual or collective behaviors (Robson et al., 2015). Kapp (2012) argues that in researching the types of player-preferred interactions, he realized that designers repeatedly made the mistake of thinking of gamification development in terms of content rather than in terms of interactions that lead to behaviors.

When designing mechanics with team missions, dynamics such as cooperation can emerge. On the other hand, the use of mechanics that allow social comparison (e.g., leaderboards) often promote competition between participants (Hunicke et al., 2004). Kapp (2012) exemplifies as activities that promote social interactions negotiation and leadership tasks. According to McGonigal (2009), multiplayer online games can promote the creation of communities, uniting people through a common interest, creating socialization capacities, even when individuals do not have common characteristics between them. Individuals who participate in social experiences are more likely to contribute and collaborate in projects

because they are willing to meet obstacles voluntarily. They are also recognized for their digital skills and rapid learning of new interactive interfaces.

Instead, emotions are a set of feelings resulting from the gamification dynamics (Robson et al., 2015). Emotions should be positive, after a fun and emotionally engaging experience (Hunicke et al., 2004). People will only stay involved in something if they have positive feelings about it (Sweetser & Wyeth, 2005). Positive emotions can manifest in many ways, such as amusement, surprise, enchantment, and personal triumph over adversity. Designers should be aware of the possibility that emotions may manifest as behaviors resulting from negative feelings, such as disappointment, sadness, or frustration at the failure to achieve a goal or gain a reward (Robson et al., 2015).

Digital games' developers claim there are four ways to unlock players' emotions: (1) provide challenge, strategy and problem-solving opportunities (lots of fun); (2) insert elements that promote mystery, intrigue and curiosity (easy fun); (3) put players in a state of excitement or relief (altered states); and (4) promoting competition and teamwork (people's fun) (Lazzaro, 2005). Furthermore, players' types of emotions can be organized into eight categories: sensation, fantasy, narrative, challenge, companionship, discovery, expression, and submission (Ibanez et al., 2014).

Games are an effective way of structuring experiences that provoke positive emotions since the fun in the games is not in the mastery – in the learning of the game is the real mission (McGonigal, 2009). Correspondingly, Sheldon (2012) reinforces the importance of turning learning experiences into something fun and enjoyable. McGonigal (2009) argues that the longer socially interacting individuals stay, the more a positive emotion arises: pro-social emotion, a type of emotion essential for the creation of permanent social bonds. The feeling of belonging is also mentioned by the author, who describes it as the emotion that comes from participating in something greater than yourself. Kapp (2012) enriches this perspective by stating that the affective domain is one of the most significant challenges for designing in learning environments. Working with individuals' values and beliefs is something that most organizations neglect when developing an instructional solution. Kapp's (2012) study mapped research that tested games with a focus on pro-social emotions, revealing evidence that they positively influence behavior, encourage participation and role performance.

Many elements presented in the literature do not fit into the MDE model since they do not fit in provided definitions of mechanics, dynamics, and emotions. Many elements such as the use of challenge and narratives resemble design principles. Therefore, the next section details the main findings in the articles respecting this aspect of gamification.

4.3 Gamification principles

When designing setup progression and rules mechanics, designers must observe the principles to arrange game elements (Dicheva et al., 2015). In this section, these principles are presented.

4.3.1 Setup principles

Environments are one of the setup mechanics designers need to deal. One of the purposes of a proper environment design is to allow players to exercise autonomy through freedom of choice. Multiple routes to success allow access to sub goals before accomplishing larger tasks (Dicheva et al., 2015).

Narrative strategies are known for their importance to connect individuals to real life experiences, giving meaning to gamification. Techniques such as storytelling and character building can motivate players in a role-playing process. Creating a narrative context needs to be associated with gamification goals (Hanus & Fox, 2015). Alves (2014) emphasizes the importance of incorporating game narrative in learning contexts to provide relevance and meaning to the experience.

Align the story with environments, objects, and progression mechanics, creating a narrative context around tasks can increase the participants' motivation and involvement (Clark & Rossiter, 2008). Simulation strategies, searching higher levels of psychological and physical fidelity, increases immersion and engagement (El Tantawi et al., 2016).

4.3.2 Progression principles

Decisions involving progression mechanics' design are critical and may lead all gamification to fail if planned and implemented improperly. Therefore, the principles of progression discussed below are critical to an adequate design of gamification (Robson et al., 2015).

The progression mechanics are the feedback mechanisms guiding participants in their evolution toward the goal. Thus, progression should be easily visible so that it is possible to achieve mastery, improve reputation, and gain social credibility and recognition. Feedback should be immediate or in short cycles, functioning as a reward system at the exact time that behaviors are expressed, rather than in distant and long-term. Feedback mechanisms generate

cycles of engagement, which can be competitive or cooperative (Dicheva et al., 2015; Landers & Callan, 2011).

McGonigal (2009) defines that in its most basic form, the feedback system can be as simple as informing the players if the goal or result has been achieved or not. The speed of the feedback is the most significant difference between digital and non-digital gamification. In digital ones, there is an interactive loop purposely associated with goal fulfillment, where there is virtually no time lag between actions and system responses (McGonigal, 2009). Sheldon (2012) applied a feedback system in a classroom incrementally, with all participants starting from zero scores. According to the author, rapid and incremental feedback cycles stimulate strategic thinking, enabling more opportunities for success.

Progression mechanics should induce players to overcome challenges at different levels of play, motivating them to continue their quest at higher levels (Cain & Piascik, 2015). The information must be organized by difficulties or tasks that the players need to perform. In learning environments, this process is called a scaffolding instruction (Cheong, Flilippou, & Cheong, 2014), which may be useful but must accommodate conforming to the needs of each person individually, in order to maintain motivation. Some complex games can adapt difficulty progression individually, keeping players at a specific level until they have demonstrated the mastery necessary to move on, gaining access to new challenges and content (Beed, Hawkins, & Roller, 1991).

Two relevant discussions need to be made regarding the progression mechanics and its implementation principles. The first considers the use of internal/external and intrinsic/extrinsic rewards, a discussion that is mainly held under the watchful eye of Self-Determination Theory (Deci & Ryan, 2000). The internal reward is understood as the one that generates benefits only within the gamified experience, while the external reward has value outside the experience. If rewards have symbolic, intangible value, even within gamification, it is considered an intrinsic reward (for example, positive emotions in players). Whether the rewards are concrete, with tangible benefits inside or outside the gamification environment, it means that they are extrinsic. Attitudes toward intrinsic rewards are defined as intrinsic motivation, and for extrinsic rewards as extrinsic motivation (Deci & Ryan, 2000).

After an extensive review of studies, Deci and Ryan (2000) argue that the use of extrinsic rewards harms individuals' intrinsic motivation. This debate is still open. Kapp (2012) states that if using extrinsic rewards, internal or external, ensuring that it is linked to performance improvement, there will be no harmful consequences on the participants. Besides, to use external rewards, designers must be able to have the budget and other conditions to

maintain them until the end of the experience and the achievement of the planned objectives. Conquests must remain desired by the players. Otherwise, the experience may lose its importance. The distribution of external rewards needs to be maintained throughout the experience. However, internal rewards may be of unique value to players and cost ordinarily close to zero. In particular, rewards with social significance, such as badges and trophies (Robson et al., 2015).

The second discussion on the progression principles is the use of public visibility rankings – as a tool of social comparison. Some authors argue that competitive environments are beneficial for learning (Kapp, 2012; Muntean, 2011). However, leaderboards must be implemented with caution, always observing if there are differences in skill and knowledge among the players and if they are in an equal position to compete to achieve the goals. If used, leaderboards should be visible and clearly show the effort required to achieve the proposed goals (Nebel et al., 2016). Some studies have found strong adverse effects on the use of them, worsening individuals' motivation and reflecting in poorer learning performance – meeting the propositions of Self-Determination Theory (Christy & Fox, 2014, Hanus & Fox, 2015). It should be considered that competition could generate positive engagement, but only for profiles with a preference for this type of dynamics – generally the simple fact of competing values more than the actual accomplishment of tasks (Pettit et al., 2015). Competition challenges can motivate when the effort is the main element that marks accomplishments (Landers & Landers, 2015). In summary, it is necessary to observe the participants' profiles and their respective levels of ability to implement a competitive environment that generates positive results.

4.3.3 Rules principles

The rules are permissions and restrictions related to the interaction of players with the setup and progression mechanics. Thus, designers need to be aware of the principles that guide them, in order to make gamification more efficient. Goals participants need to achieve can be presented as missions or challenges (Cain & Piascik, 2015). Rules must be clear, specific and immediate or close (scalable). The objectives should also be clear and concrete, selected through actionable tasks (available) and with increasing complexity (Dicheva et al., 2015). In this context, Kapp (2012) defines the goal as a specific result that players must seek to achieve and emphasizes that adequate gamification design leaves no doubt, whether or not the goal has been achieved since there should be no ambiguity.

Rules determine another critical principle to be observed during design: the approach to error. Gamification should allow participants to restart the tasks and try again through several attempts, leaving them free to take risks and make mistakes that can be retrievable. In learning environments, the error must have a different approach from that used in traditional learning environments (Sheldon, 2012). The author reveals that, just as in games, the ideal is that gamification always allows a new chance to solve problems. A process of trial and error is possible when people think about their behaviors. To encourage error as a strategy designers need to insert task repetition and feedback cycles to allow the desired performance improvement. Freedom to make mistakes allows participants to experiment and act fearlessly, increasing their involvement and engagement, and consequently their relationship to gamification (Hanus & Fox, 2015; Lee & Hammer, 2011).

Gamification should be carefully designed not only to promote engagement but also to prevent, detect and discourage dishonest behavior that may arise – e.g., trying to break established rules (Ibanez et al., 2014). Finally, rules and objectives should be designed in a growing complexity, allowing for a personalized experience to encourage autonomy, with adaptive difficulties and challenges tailored to players' abilities, providing more difficult missions as the ability grows (Dicheva et al. al., 2015).

5 Methods

This section describes the methodological strategies for achieving the main objective of evaluate the gamification effects on learning in generalization and transfer of training, in classroom events for managers in a Brazilian financial institution.

The methodological approach of this research is mixed since it includes a combination of qualitative and quantitative approaches, supported in a more consistent dataset (Creswel, 2010). This research is composed of six studies, four of which are descriptive and exploratory in nature and two quasi-experiments of an explanatory nature. About the composition of quasi-experiment I and II, there are three groups, Experimental Group I (with gamification as an instructional strategy), Experimental Group II (without gamification) and Control Group (without training) (Creswel, 2010). Therefore, it is research with a microanalysis level – investigations are made on individual’s behavior. The choice of the sample of participants in the research is non-probabilistic and for convenience, since individuals were invited according to the target audience of the selected training, and this is a quasi-experimental design because it d was not be possible to assign participants randomly. Within the population of managers of

the financial institution were withdrawn the sample of managers who participated in the training events.

Pilati and Borges-Andrade (2006) point out that it is fundamental that research in T&D uses experimental and quasi-experimental designs because, even with reduced samples, they contribute to the development of increasingly effective procedures for the T&D systems. Precautions were made to use this methodological design conforming to the aspects Shadish, Cook, and Campbell (2002) affirm as threats to study's internal validity. In the discussion session, Table 5 will be revisited, explaining how this research treated each of the threats to internal validity. Some further precautions were token respecting the internal validity once was not the same instructor conducting the training events.

Table 5
Threats to internal validity

Reason	**Definition**
Ambiguous Temporal Precedence	Lack of clarity about which variable occurred first.
Selection	At the start of an experiment, the average person receiving one experimental condition already differs from the average person receiving another condition. This difference might account for any result
History	Events occurring concurrently with treatment could cause the observed effect.
Maturation	Naturally occurring changes over time could be confused with a treatment effect.
Regression	When units are selected for their extreme scores, they will often have less extreme scores on other variables, which can be confused with a treatment effect.
Attrition	Loss of respondents to treatment or measurement can produce artificial effects if that loss is systematically correlated with conditions.
Testing	Exposure to a test can affect scores on subsequent exposures to that test, which can be confused with a treatment effect.
Instrumentation	The nature of a measure may change over time or conditions in a way that can be confused with a treatment effect. For example, a spring might become weaker over time and easier to push, artificially increasing reaction times.

Reason	Definition
Additive and Interactive Effects of Threats to Internal Validity	The impact of a threat can be added to that of another threat or may depend on the level of another threat.

Source: Adapted from Shadish, Cook, and Campbell (2002)

This research used a sequential exploratory strategy (Creswel, 2010). This mixed method approach involves a first phase of qualitative data collection and analysis followed by a second phase of quantitative data collection and analysis. The second phase is developed on the results of the first qualitative phase. For each of the four qualitative studies (required for the quasi-experiments realization), it was necessary to create products (intermediate results) that subsidized the subsequent studies – especially the quantitative ones. For this reason, the methods section of each qualitative study will also present these products as results of this sequential research.

The sequential exploratory strategy is often chosen when the development of instruments is necessary – available instruments do not meet the proposed objectives. Typically, it is divided into three stages: (1) qualitative data collection and analysis occur, (2) data collected and analyzed are used for instrument development, and (3) the instrument is subsequently applied in a population sample (Creswel, 2010).

The phases were placed in sequence to facilitate the understanding of the specificities of this research (Figure 2). Here, they are explained briefly and explored in more detail in specific sections that follow.

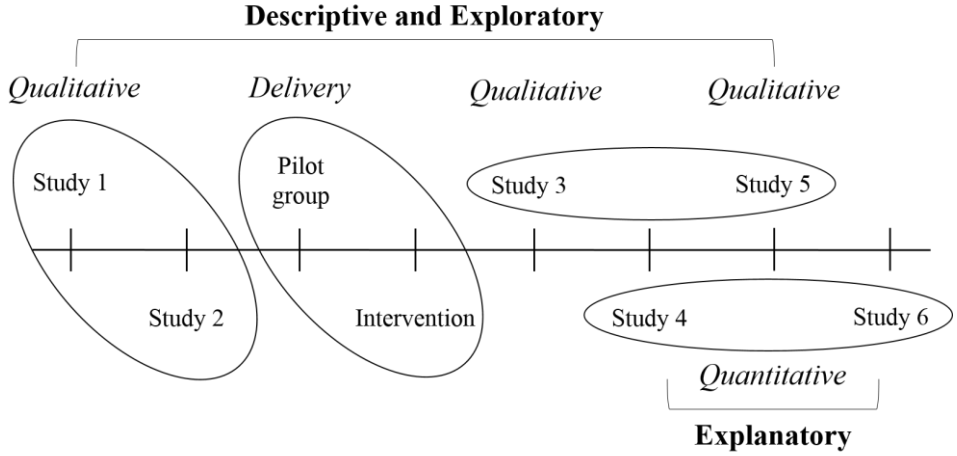

Figure 2. Research sequential exploratory strategy
Source: Author's elaboration

Each of the stages in the figure is briefly explained below:

- **Study 1** – Intervention development: Qualitative approach based on the theoretical framework, literature review and documental analysis to achieve the specific objective (i) – Create two equivalent instructional designs, one with the use of gamification as an instructional strategy and another with traditional strategies used to contrast with gamification – group discussions and dialogues exhibitions. The trainings are equivalent because they have the same instructional objectives. Only instructional strategies differ and were manipulated. Control Group was used to control alternative external explanations. Differences in the pretest were statistically monitored – the groups were not randomly assigned to the three quasi-experimental conditions.
- **Study 2** – Situational performance tests development: Qualitative approach based on the instructional objectives to achieve a specific objective (ii) – Create two equivalent situational performance tests to measure generalization, and to be applied before and after the classroom training event.
- **Pilot group**: Training event conducted with the aim of (1) test to adjust the design applying gamification as an instructional strategy and to (2) collect data to obtain evidence of validity of the situational performance tests, and create the generalization assessment criteria.
- **Intervention** – Two training events delivered with the aim of collecting data from situational performance tests, in three groups: Experimental 1, Experimental 2, and Control Group.
- **Study 3** – Learning assessment criteria development: Qualitative approach based on data collected from pilot group participants to achieve a specific objective (iii) – Create assessment criteria for generalization. Criteria for assessing learning (pretest and posttest), defined by generalization measures (situations different from those presented by training).
- **Study 4** – Quasi-experiment I: Quantitative approach using primary data, with repeated measures (two periods), to check the proposed hypotheses for generalization.
- **Study 5 –** Learning assessment criteria development: Qualitative approach based on data collected from pilot group participants to achieve a specific objective (iv) – Create assessment criteria for the transfer of training, defined by objective measures of performance of the training events' participants.

- **Study 6** – Quasi-experiment II: Quantitative approach using secondary data, with a longitudinal design, to check the proposed hypotheses for transfer of training.

The assumptions and definitions of quasi-experiment variables and their fundamental aspects of design: details on intervention groups, the definition of measures, and operational definition of variables will be explained in detail in the sections that follow. Next, the context of this research is described, bringing aspects of the organization studied and the training chosen.

5.1 Research Context

The research was conducted in an organization with a consolidated corporate university with structure, investments, and practices that facilitated the realization of the quasi-experiment. Another justification for the organization's choice was the feasibility of data registered and collected in the corporate systems. It is a large Brazilian financial institution with national and international economic importance. Currently, the organization is a leader in several economic indicators and has business units in almost all Brazilian cities.

The organization's T&D system (currently called corporate university) makes investments of around USD 30 million annually (data from 2017) for the development of internal (planned and developed within the organization), external (contracted in the market) actions, as well as in the granting of scholarships for the improvement of its employees (undergraduate, graduate and languages). The FI corporate university has classrooms in all Brazilian states, to carry out classroom training events, and a learning platform for distance training, accessible to all employees (approximately 90 thousand people). The organization has a T&D evaluation system based on the Integrated Workforce Impact Assessment (Abbad, 1999) supporting decision making and giving feedback to training events and programs.

This research happened in a classroom course internally developed by the organization. The course selected for the quasi-experiment is called Performance Appraisal (PA) Course. The classroom training has 16 hours long, happening in two consecutive days of 8 hours each. The training target audience is around 30 thousand employees who hold management positions in the FI. Through several actions, the organization seeks to develop these managers in topics related to management and human resources. Currently, the organization has focused on the development of leaders focused on topics such as strategic thinking, cultivating innovation, and management by purpose and talent development.

The process of Performance Appraisal (PA) in the FI is based on competency management model. All managers have crucial tasks associated with the PA model. The first task concerns the plan of an expected performance agreement for the evaluation cycle (six months duration). The second is to perform the monitoring of their team members by issuing verbal feedback and written annotations. The third is the issue of concepts for the members of their team. The training has existed for more than a decade since the PA model is implemented in the FI. Since the performance agreement execution and the emission of concepts are particular aspects carried out respectively at the beginning and the end of the evaluation cycle, annotations and verbal feedback are the most critical aspects of the PA model. The annotations are recorded in a corporate system and refer to performances observed by these managers. They record at any time, for any member of their team. The annotations base the concepts' emission at the end of the evaluation cycle.

The following sections detail the steps of the sequential exploratory method adopted in this research.

5.2 Study 1 – Intervention development

This section describes the procedures vital to achieving the goals regarding the creation of the intervention used in this quasi-experimental design. This study uses a qualitative approach based on the theoretical framework, literature review and documental analysis.

The first step in developing the instructional designs applied in this quasi-experimental research was the analysis of the selected training event's (Performance Appraisal Course). The FI already had an instructional material of the chosen training, though the latest version available was outdated and not being used for eight months. In addition to the mentioned version, there were instructional materials of four versions of the same training used for more than ten years in the FI – all versions kept the main learning objective only with minor redaction changes. The own researcher implemented this analytical task for approximately one month.

To achieve the quasi-experiment objective, it was necessary to create two distinct instructional trainings applying gamification as an instructional strategy in only one of them. After the latest version instructional design analysis, there was a conclusion that it would not be possible to execute the quasi-experiment inserting gamification into the existing activities to later remove them, creating the training without gamification as an instructional strategy. The pre-existing design would only allow the insertion of simple elements like points or leaderboard. As it was an instructional event of only 16 hours duration, if the gamification

variable were not manipulated in order to give sensitivity to it, would be more likely to have no impact at all on learning outcomes, especially at the transfer for training. Consequently, the strategy adopted was to reconfigure all instructional strategies using gamification in those identified during the workshop to build the instructional design.

A week-long workshop was conducted with this study's researcher and two FI staff members: a specialist in the training subject (an employee of the strategic unit responsible for FI's PA strategy) and an instructional designer (employee of FI corporate university, instructor and designer of the latest version of the training). Throughout this week all the instructional objectives (IO), general and specific, of the Performance Appraisal Course were reviewed. After this work, it was rewritten the main learning objective to "evaluate performances using performance agreement to plan, giving contingent feedbacks (verbal and written) to monitor, and scores to end the assessment cycle". This rewriting was crucial because the instructional objective had redaction problems, positioning the action in the work environment.

Following the instructional design sequence, all specific objectives were rewritten, and the activities ordered. The rewriting of all specific objectives was necessary because there were writing problems and the way they were, could not be adopted in the quasi-experiment. The cognitive domain of Bloom's taxonomy (1956) was used to accomplish this task. It was decided to order activities based on the sequence of PA procedures in work performance in the FI (this was already the sequence in previous outdated versions). Among the activities of the training, the realizations of the pretest and posttest during the events were already foreseen. It was also decided to use mobile devices (tablets) in the practical activities, for creation and sharing of products. The result of this workshop is in Table 6.

Table 6
Specific instructional objectives after revision

IO	**Activity name**	**Instructional Objective (IO)**	**Complexity level**
IO1	Tablet navigation tutorial	Explore the tablet's features to finish the assessment tests.	Application
IO2	Pretest	Perform tasks related to PA based on previous knowledge.	Synthesis
IO3	Introduction	Identify the main objective, contents, and rules of the course.	Knowledge
IO4	Concepts review	Identify the PA model used in the FI characteristics.	Knowledge

IO	Activity name	Instructional Objective (IO)	Complexity level
IO5	PA model	Identify the PA model concepts used in the FI	Knowledge
IO6	PA planning stage	Relate context fragments to the planning process in FI.	Comprehension
IO7	Theoretical exposition (planning stage)	Identify the performance agreement importance to planning and other PA stages.	Comprehension
IO8	Practical activity (planning stage)	Write performance agreements from fragmented information, considering PA planning stage concepts.	Synthesis
IO9	PA planning stage (assessment)	Evaluate the quality of the performance agreements, according to the knowledge acquired about the PA planning stage.	Evaluation
IO10	PA monitoring stage	Relate performance description fragments with the PA monitoring stage fundaments	Comprehension
IO11	Theoretical exposition (verbal feedbacks)	Identify the importance of continuous verbal feedback in the PA process.	Comprehension
IO12	Practical activity (verbal feedbacks)	Provide verbal feedbacks to simulated situations.	Application
IO13	Theoretical exposition (annotation)	Identify the importance of continuous monitoring, with constant annotations in the PA process.	Comprehension
IO14	Practical activity (annotations)	Write annotations for simulated situations and performances, according to the PA model.	Synthesis
IO15	Annotations (assessment)	Evaluate the annotations produced by the participants, according to PA fundaments.	Evaluation
IO16	Theoretical exposition (closing stage)	Relate the scores' emission to the planning and monitoring stages according to the PA model.	Comprehension

IO	Activity name	Instructional Objective (IO)	Complexity level
IO17	Practical activity (closing stage)	Give scores to simulated situations, relating to FI competencies and annotation types.	Application
IO18	Final review	Recap the contents learned during the course.	Comprehension
IO19	Posttest	Perform tasks related to PA based on the knowledge gained during the course.	Synthesis
IO20	Pretest and posttest comparison	Compare the perceived performance in the pretests and posttests.	Comprehension

Source: Author's elaboration

The next step in instructional design was the creation/selection of the instructional strategies. It was decided that the workshop components would perform this step using the gamification that would later be removed from the activities in the course design called traditional – without the use of gamification, replacing it by longer discussions and expositions. The instruction without gamification's use needed to be appropriate for the study, all instructional objectives were identical to those of the gamified training event. The only aspect maneuvered experimentally was the instructional strategy, maintaining all the other instructional design aspects constant, in order not to influence the internal validity of the quasi-experiment. Table 7 visually explains the maintenance of the objectives but already shows the activities where instructional strategies differ from one event to another.

As the use of game elements required more extensive instruction prior to the activities, and more complex procedures in the execution and finalization of the activities, the remaining times of each class activity without gamification were used with elongation of group discussions and dialogued exhibitions. In average, participants of the not gamified group (traditional) had around 20 minutes more per activity (in the activities with gamification’s use) to discuss concepts and the outcomes from products presented in the practical activities. It represents around 2 hours more to discuss the course content with the instructor or between the participants.

Table 7
Instructional designs

IO	Activity name	Instructional Objective	Complexity level	Instructional strategy	
				Traditional	**Gamified**
IO1	Tablet navigation tutorial	Explore the tablet's features to finish the assessment tests.	Application	Instructed navigation	
IO2	Pretest	Perform tasks related to PA based on previous knowledge.	Synthesis	Learning assessment	
IO3	Introduction	Identify the main objective, contents, and rules of the course.	Knowledge	Dialogued exposure	
IO4	Concepts review	Identify the PA model used in the FI characteristics.	Knowledge	Group discussion	Competitive quiz
IO5	PA model	Identify the PA model concepts used in the FI	Knowledge	Dialogued exposure	
IO6	PA planning stage	Relate context fragments to the planning process in FI.	Comprehension	Group discussion	Board activity
IO7	Theoretical exposition (planning stage)	Identify the performance agreement importance to planning and other PA stages.	Comprehension	Dialogued exposure	
IO8	Practical activity (planning stage)	Write performance agreements from fragmented information, considering PA planning stage concepts.	Synthesis	Practical activity	

IO	Activity name	Instructional Objective	Complexity level	Instructional strategy	
				Traditional	Gamified
IO9	PA planning stage (assessment)	Evaluate the quality of the performance agreements, according to the knowledge acquired about the PA planning stage.	Evaluation	Group discussion and evaluation with qualitative feedback	Group discussion and evaluation with quantitative feedback
IO10	PA follow-up stage	Relate performance description fragments with the PA monitoring stage fundaments	Comprehension	Group discussion	Board activity
IO11	Theoretical exposition (verbal feedbacks)	Identify the importance of continuous verbal feedback in the PA process.	Comprehension	Dialogued exposure	
IO12	Practical activity (verbal feedbacks)	Provide verbal feedbacks to simulated situations.	Application	Practical activity with qualitative feedback	Practical activity with quantitative feedback
IO13	Theoretical exposition (annotation)	Identify the importance of continuous monitoring, with constant annotations in the PA process.	Comprehension	Dialogued exposure	
IO14	Practical activity (annotations)	Write annotations for simulated situations and performances, according to the PA model.	Synthesis	Practical activity	
IO15	Annotations (assessment)	Evaluate the annotations produced by the participants, according to PA fundaments.	Evaluation	Group discussion and evaluation with qualitative feedback	Group discussion and evaluation with quantitative feedback

IO	Activity name	Instructional Objective	Complexity level	Instructional strategy	
				Traditional	Gamified
IO16	Theoretical exposition (closing stage)	Relate the scores' emission to the planning and monitoring stages according to the PA model.	Comprehension	Dialogued exposure	
IO17	Practical activity (closing stage)	Give scores to simulated situations, relating to FI competencies and annotation types.	Application	Practical activity	
IO18	Final review	Recap the contents learned during the course.	Comprehension	Group discussion	Competitive quiz
IO19	Posttest	Perform tasks related to PA based on the knowledge gained during the course.	Synthesis	Learning assessment	
IO20	Pretest and posttest comparison	Compare the perceived performance in the pretests and posttests.	Comprehension	Group discussion	

Source: Author's elaboration

5.2.1 Gamification as an instructional strategy

In this section, is presented the result of gamification use as an instructional strategy. As gamification is the main subject if this research, it deserves a special specification of how it was implemented, and how the game elements were combined seeking to produce effects on generalization and transfer of training.

Study 1 was the development of two instructional designs applied in the intervention classes of this research. The anatomy of the two instructional designs created was already shown, exhibiting the instructional objectives and the instructional strategies that differed in the use or not of gamification. The instructional materials in their entirety, will not be shown in this research, because it contains identification of the studied FI, and contents were considered confidential. The report will focus on the instructional results of the use of gamification – the main theme of this research.

Initially, the game elements that were used throughout the event (without specific link with a particular activity) will be explained, and later the game elements that were adopted in specific activities will be discussed – the second group had a pattern for their use. These same specific activities were those where the instructional strategy was different in the events. Before specifying each game element applied, Table 8 shows the game element's distribution in the training structure ("IO" means instructional objective). Gamification was applied differently from a game-based approach (use of games as activities inside a training solution), something already mentioned as essential to avoid the theoretical shadowing that exists between game-based learning and gamification (Landers, 2014).

5.2.1.1 Elements used throughout the entire event

(a) Avatars, Progress bars, and upgrade points

These elements (joined in the form of a physical card – see Figure 3) were delivered for participants at the training event's beginning. The **avatars** were adopted with the purpose of creating a bond between the participants and the symbolic representations chosen by them. The avatars had information that needed to be filled by the participants: name, age and time in the FI. In total, 37 models of avatars were offered, representing different genres, ages, and ethnicities. Each avatar had three empty **progress bars** that were completed by the participants during the event, due to the performance obtained in the practical activities. In order to complete the progress bars, participants should purchase **upgrade points**, which were distributed according to the performance achieved in the practical activities.

Table 8
Game elements use per activity

Game element	IO1	IO2	IO3	IO4	IO5	IO6	IO7	IO8	IO9	IO10	IO11	IO12	IO13	IO14	IO15	IO16	IO17	IO18	IO19	IO20
Answer points																				
Avatars & Progress bars																				
Boards & Dices																				
Info cards																				
Team points																				
Upgrade points																				
Leaderboard																				
Digital feedback system																				
Team avatars																				

Source: Author's elaboration

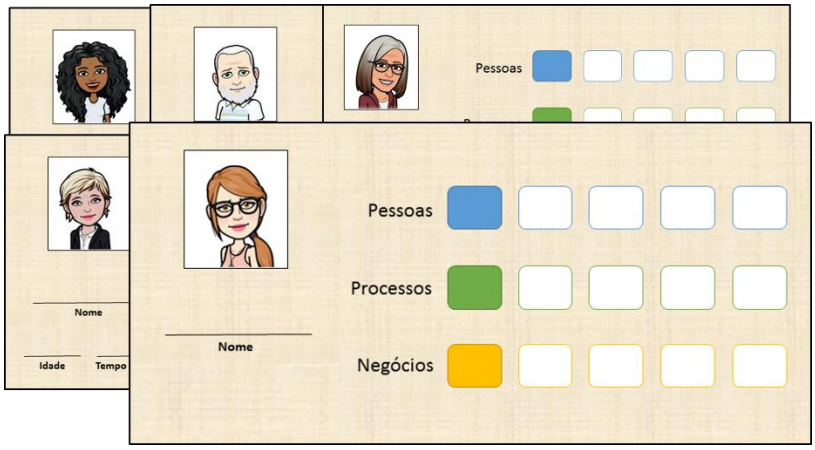

Figure 3. Avatars and their progress bars
Source: Author's elaboration

(b) Team points, Leaderboard and Progress bars

Working in the same logic as the upgrade points, **team points** were earned as a performance reward in practical activities. Each participant could acquire a different number of points in the same activity, valuing, even more, the excellence of performances. The team points were symbolically associated with the FI goal system, known to all participants. The team points were converted to the end of the tasks where they were won in points that went directly to the **leaderboard** that compared teams' performances. This leaderboard was updated in two ways: A physical (always visible to all participants); and digital (in some activities the updating of the points was done in a digital file and then transcribed to the physical mural). The digital version of the leaderboard had the evolution represented by a **progress bar.**

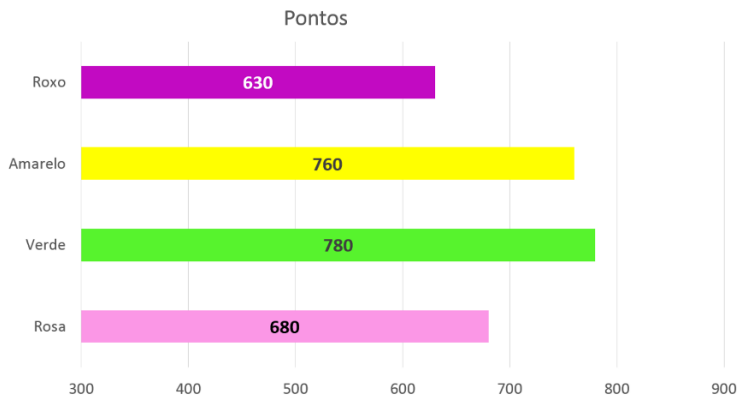

Figure 4. Digital leaderboard
Source: Author's elaboration

5.2.1.2 Elements used in specific activities

(c) Boards, Dices and Info cards

These elements were adopted in two specific activities, though with the same intention: randomize the access to the information contained in the **info cards**. Two types of info cards were built: (1) cards describing simulated behaviors; and (2) cards with contextual situations from the FI's units. The **boards** created were of simple use (advance to the end), with space destined for each team. The random factor was a result of **dices** rolling (applied electronically via tablets app). The reason to use was to reward participants through the uncertainty of the results randomly. A precaution was taken so that even groups with worse performance had enough info cards to execute the practical course activities, that was the main reason for the use of these game elements since the cards conquered by the participants were applied later in the PA practical activities. In the event without the use of gamification, the information was delivered evenly to the participants, without the aspect of randomization.

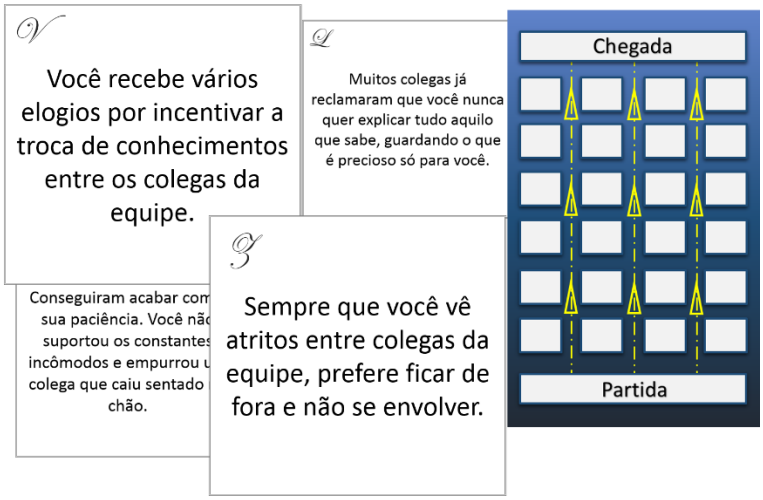

Figure 5. Info cards and board
Source: Author's elaboration

(d) Digital feedback system

In all practical activities, where PA performances were demonstrated, participants should present products (individually or as a team). Because it was a training for managers, whose theme was performance evaluation, a digital performance assessment tool as used in all

practical activities, and this tool, available on each participant's tablet, allowed the participants to give scores for each performance presented throughout the course. The score was derived from the evaluations that were carried out by the participants themselves. A 7-point Likert scale was adopted similarly to that used by the participants in the execution of their PA tasks: (1) Did not express what was required; (2) Express much lower than expected; (3) Express moderately below expectations; (4) Expressed slightly below expectations; (5) Express as expected; (6) Express slightly above expectations; and (7) Express much higher than expected.

The best-evaluated performances were rewarded with team points and upgrade points conforming to criteria established and presented previously to the participants. In the event without the use of gamification, the evaluation of the products of the practical activities was assessed only qualitatively by the instructor and the other participants through group discussions.

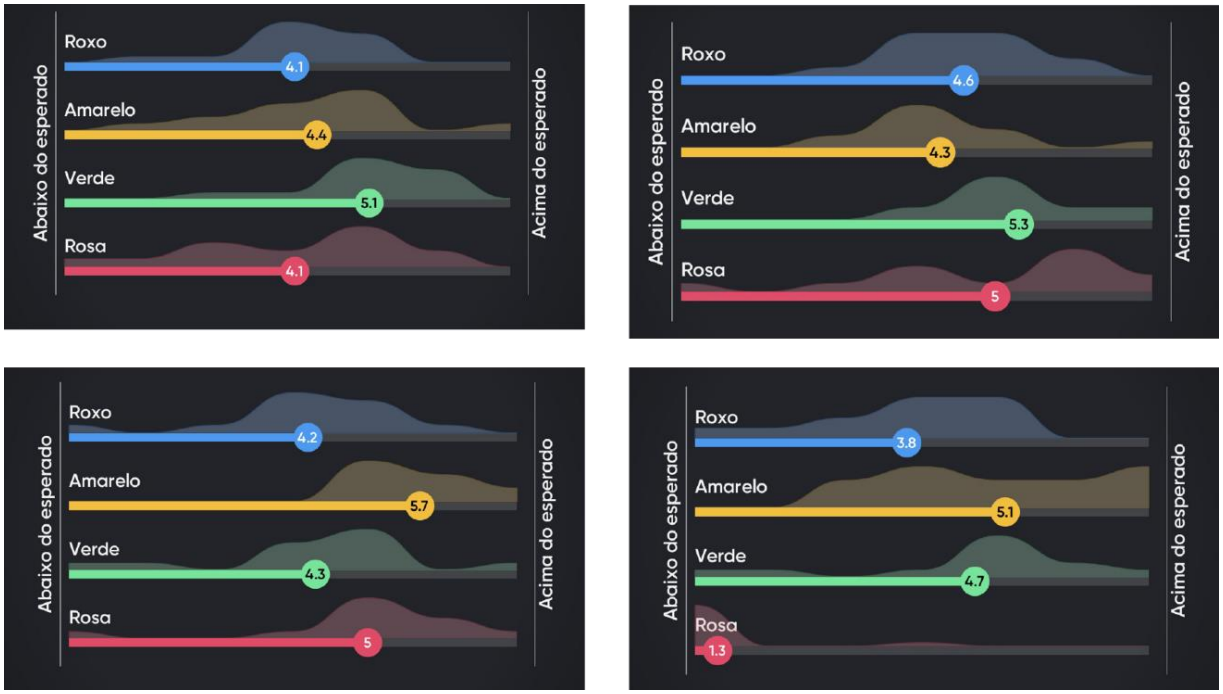

Figure 6. Digital feedback system
Author's elaboration

(e) Answer points and super avatars

During the instructional event, two content review activities were conducted. The initial review activity aimed at activating the participants' memory of their prior PA knowledge. For this activity was used as an application on mobile devices that gave response points for

participants for the answer's accuracy and speed. The top four players were given the power to assemble their teams and choose their avatars first. The final content review activity was also the final activity of the competition between teams (competing through team points shown in the leaderboard). This activity was also a question-and-answer competition but carried out in a way that group participants should collectively decide on the answer to be shown (via a physical colored card). Before the activity started, a super avatar was equipped for each team. This game element had two attributes: (1) XP: accumulated score of team points displayed in the leaderboard before this activity's beginning – acquired by the teams throughout the whole event; and (2) Combo: the sum of all upgrades points acquired by team members throughout the entire event. At every correct answer, the groups punctuated the equivalent of the Combo of their super avatar, adding this value to the XP – starting score. This element of gamification was used to finalize the competition, giving an opportunity for the teams that were not in the lead to remain motivated to seek the victory. Also, the sum of the avatars' upgrades and the answers' joint decision reinforced the participants' team spirit. In the event without gamification, the review of contents happened in the form of group discussion, with the instructor asking the question and the participant who wanted to respond to raising the hand and giving the answer.

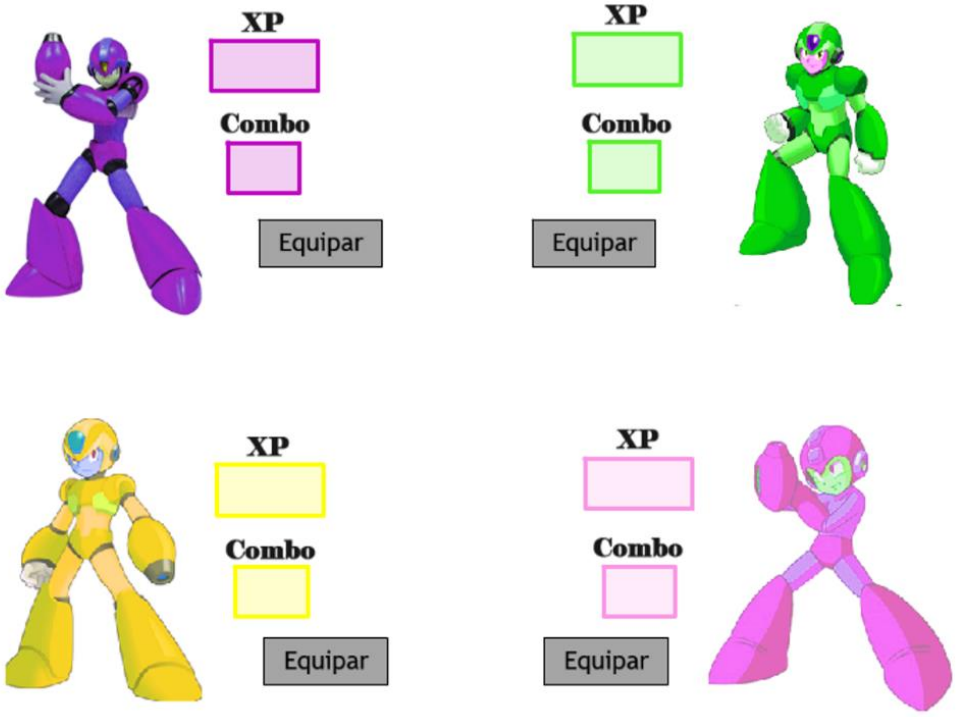

Figure 7. Unequipped super avatars
Source: Author's elaboration

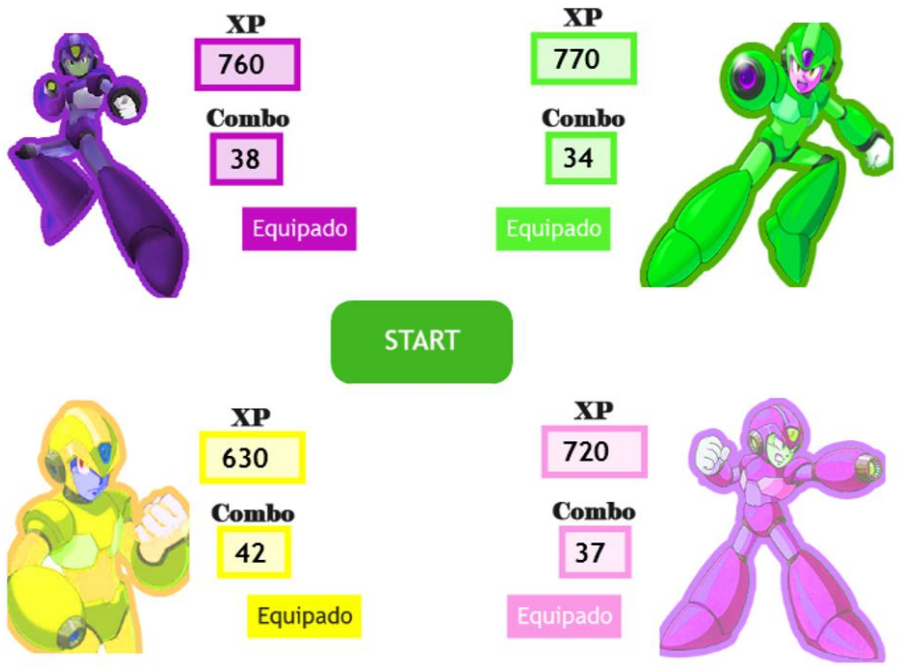

Figure 8. Equipped super avatars
Source: Author's elaboration

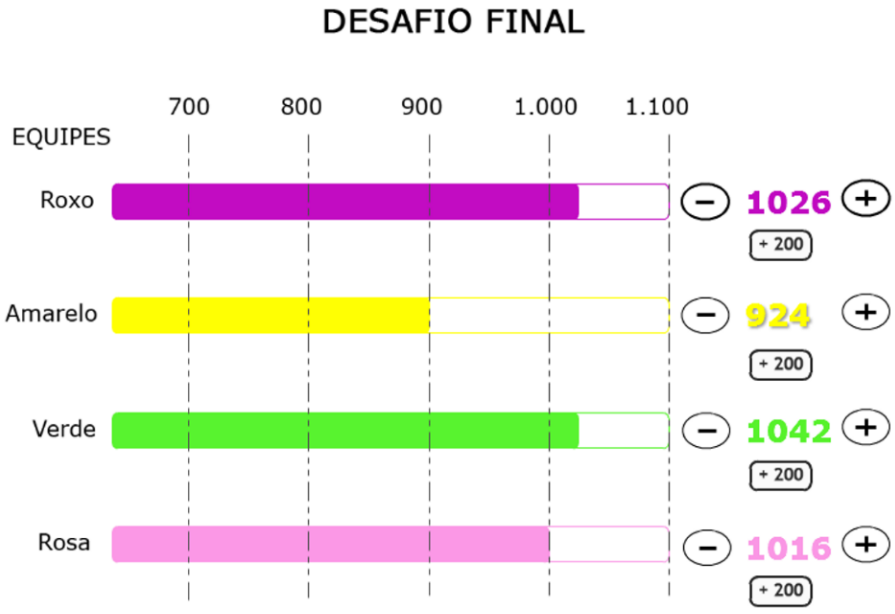

Figure 9. Digital leaderboard after the final challenge
Source: Author's elaboration

5.3 Study 2 – Situational performance tests development

Study 2 also uses qualitative approach sequentially based on the instructional objectives built in Study 1. To assess generalization was required the creation of tests similar to the work performances shown by the managers' participants of the training events part of the quasi-experiment. In this section is reported the methodological procedures to develop the situational performance tests.

To execute the instruments construction phase to assess generalization was conducted a workshop with the length of four days, composed of professionals indicated by the FI, all different from those who had participated in the group for the development of the instructional design. In addition to the presence of the researcher responsible for the study, the group had three other participants: an instructor indicated by a manager of the board responsible for the PA subject in the FI (same instructor that conducted the pilot group and Experimental Group 2), and two business unit managers who excelled in the performance appraisal of their teams – both were part of the training target audience.

As the purpose of this research was to measure generalization and transfer of training, only specific objectives with the highest levels of the cognitive domain were selected to create the learning assessment questions (Bloom, 1956).

The chosen objectives involved tasks such as understanding and analyzing scenarios, writing, and synthesizing ideas, and evaluating PA products. Therefore, the use of multiple choice questions would be possible for intermediate instructional objectives (at the knowledge and comprehension levels), but not for the terminal ones – most important to achieve the proposed research objectives. Thus, the creation of situational performance tests with open answers became vital to search for the attainment of the proposed objectives (measure instructional objectives achievement from synthesis and evaluation levels). The tests were designed to simulate the sequence of PA situations that the participants perform when in their workplace functions. Table 9 shows the theoretical instructional anatomy during tests creation. The activities explaining and reviewing concepts were prerequisite for the subsequent practical activities – all tasks executed in those activities had previous explanations of the concepts and examples crucial to perform them. The initials IO represents “instructional objectives”. The “Assessed” column lists the higher level objectives, while the “Requisite” column lists the lower levels containing necessary knowledge to achieve the tasks required. The questions also delivered to the participants some information regarding situations and specifications that allowed them (allied with the conceptual knowledge acquired) to answer what was being asked.

Table 9
Generalization assessment structure

Question	Part	Assessed IO	Requisite IO	Given information	Task required	Question type
1	1	IO8, IO9	IO4, IO5, IO6, IO7	Unit context, describing problems related to human resources, processes and business in the FI Performance agreement with inappropriate writing	Evaluate the given performance agreement, editing the text to improve its quality.	Open answer
2 / 3	1	IO10	IO4, IO5	Fictional character's profile with performances descriptions Competencies options Annotation's type options	Select one competence and the annotation's type.	Multiple choice answer
	2	IO14	IO4, IO5, IO10, IO11, IO13	Fictional character's profile with performances descriptions	Write an annotation to the fictional character	Open answer
	3	IO17	IO4, IO5, IO16	Seven points Likert scale used by the FI in the PA.	Give a score to the fictional character in the selected competence	Multiple choice answer

Source: Author's elaboration

In order to verify the behavioral change in the quasi-experiment participants, it was necessary to create two tests, one to be applied before and another to be applied after the training events (except for the Control Group, in which the two different tests were applied sequentially). To verify clarity in the instructions, tests' equivalence and to seek to construct validity's evidence, two pilot tests were implemented with following adjustments after each one. The first happened with two managers belonging to the training target audience – who had not participated in any previous phase of the research. They undertook a first pilot test to verify the instructions for usability of the mobile devices, and also the instructions' clarity for accomplishing the tasks required in the questions. Adjustments were made after this first pilot test. Secondly, the tests were submitted to two specialists from the Board responsible for the PA strategy in the FI (who had not participated in the tests' construction) to verify the questions' balance and if the questions were adequately measuring what they were supposed to measure. The final instrument applied to all participants (pilot group, Experimental Group 1, Experimental Group 2, and Control Group) is in its entirety in Annex 1.

5.4 Pilot group training event

This section describes the first step in the execution of this research. A training event was conducted with the aim of testing to adjust the design applying gamification as an instructional strategy and collecting data from situational performance tests. This pilot group tested only the instructional design with the use of gamification as an instructional strategy. As the instructional design without gamification only removed the game elements, substituting them for activities as dialogical exposition and group discussions (techniques known and with less conduction complexity), this design was not tested.

This group had 19 participants, the average age was 39.53 years, 68.4% were in business unit positions, and 52.6% were females. The FI nominated the invited instructor as one of the best specialists in the PA theme. After the pilot group was done, the last adjustments were made in the instructional materials for use in the quasi-experiment.

Next, of this execution section, the instructional designs and the situational performance tests were applied in the intervention training events (pretest and posttest were also applied for the Control Group).

5.5 Intervention training events and Control Group

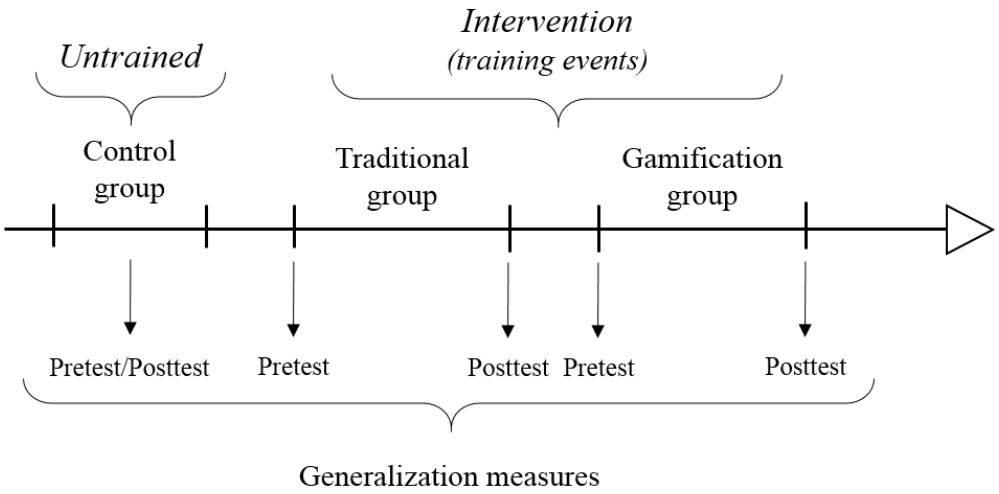

Figure 10. Situational performance tests application.
Source: Author's elaboration

At first, the structure of the intervention executed in two experimental groups training classes are explained, one with the instructional design with gamification and another without its use (Experimental Group 1, Experimental Group 2). In both, data were collected through built-up performance situational tests (called pretests and posttests). Next, the choice and application of situational performance tests in the Control Group (untrained) are described.

Two sequential classes of training were performed. The first, the Experimental Group 2 (traditional instructional design, without gamification) (N = 19) had 68.4% males and 31.6% females, 57.9% business unit managers and 42.1% support unit managers. The participants' age was between 28 and 61 years (M = 45.3, SD = 8.36), management experience between 3 and 19 years (M = 9.10, SD = 4.50), and experience in the FI between 8 and 35 years (M = 16.21, SD = 9.85).

The second class was executed for the Experimental Group 1 (N = 19), that had 52.6% males and 47.4% females, 63.2% business unit managers and 36.8% support unit managers. The participants' age was between 31 and 60 years (M = 45.31, SD = 8.81), management experience between 1 and 23 years (M = 9.05, SD = 7.05), and experience in the FI between 5 and 32 years (M = 14.94, SD = 7.65).

For both events pretests and posttests were conducted in the same two moments: (1) before the beginning of the classroom event; and (2) at the end of the classroom event. The first moment occurred before the experiences' beginning, to assess the participants' entry levels –

previous knowledge. The second moment, collected data at the instructional events' end, precisely to verify if there was any change in the participants learning performance.

The Control Group was composed of managers accessible in their work environments. The own researcher moved to the units of the participants and applied pretest and posttest, sequentially at the same time. These participants were taken to an isolated and quiet environment for conducting the tests – similar to the one obtained in the classroom. The explanations prior to the tests were provided with the same parameters applied in the classroom. The time available to perform the tests was also identical to the time available for the participants of the trained groups. Control Group (N = 15) had 46.7% males and 53.3% females, 66.7% business unit managers and 33.3% support unit managers. The participants' age was between 37 and 53 years (M = 45.60, SD = 5.35), management experience between 1 and 20 years (M = 10.80, SD = 5.25), and experience in the FI between 5 and 33 years (M = 16.93, SD = 8.81).

Therefore, with the union of the three groups (Experimental Group 1, Experimental Group 2 and Control Group), the sample studied through the intervention was a total of 53 participants, designated for convenience in the three groups: Experimental Group 1 (use of gamification) (N = 19), Experimental Group 2 (no gamification) (N = 19) and Control Group (N = 15).

5.6 Study 3 – Generalization assessment (Criteria development)

This methodological section is part of this research of mixed methods declared as a sequential exploratory strategy (Creswel, 2010). The qualitative approach used was based on data collected from pilot group participants. The data collected were from primary sources, pretests and posttests applied in the classroom. The procedures for developing the criteria for assessing the situational performance tests – learning at the generalization (quasi-experiment I) are described.

In the pilot group, the situational performance tests were applied before and after the event. Further, they were analytically investigated by the own researcher in order to establish criteria to qualify the responses given by the participants. In total, all the answers of the 19 pretests and 19 posttests of the pilot class participants were evaluated. Also, it is important to mention that the researcher was present in the pilot group and all the other classroom events. It was not possible to use external judges to assess the pretests and posttests results. This work was done by the own researcher in total work time of 40 hours. The files containing each test

were not identified (it was used as a random number for each participant). The order was also random and blind to the researcher.

For the multiple-choice items, the correct answers were already set (some items allowed interpretation with more than one correct possibility). For the questions with open answers, it was necessary to create a feedback sheet (Annex 2) indicating the performances that would be crucial for the participants to receive a score in each question evaluated. For each learning objective, underlying criteria were identified and written. These criteria were not explicit in the training instructional design, though they manifested in two ways: corrections of practical activities conducted by the instructor throughout the pilot event class; and in the analytical process carried out with the responses of the pilot class participants.

Items assessing the highest levels of the cognitive domain were written tasks. Question 1, item c, required the correction of the given performance agreement (observing the criteria in the assessment sheet). For the annotations tasks (part 2 of questions 2 and 3) it was written golden scores, with the expected response, observing the criteria created for assessing learning – also present in the annexed feedback sheet.

The identified criteria were transformed into questions – the entire written structure was maintained the same. This process was made to facilitate the learning assessments' correction. With the exception of question 1, item c (where there should be a counting of action plans describing solution to contextual problem presented), the other questions would only allow two answers' types: (1) yes - the participant was considered to have correctly answered the question, receiving the score equivalent to it, and (2) no - it was considered that the participant did not answer the question correctly and did not receive the score.

Table 10 shows the relation between the instructional objective (developed in study 1) and the criteria (developed in this study), and also the number of points given for each correct answer. In order to avoid any sharp imbalances between the score of each question and its respective items (though observing the complexity level required in each of them), each correct answer received 4 or 5 points (the maximum possible overall score in pretest and posttest was 100 points). In question 1, the two first items (“a” and “b”) were worth 4 points, because the participant should only make reports in the text, while the item that required writing an action plan was worth 5 points (for each action plan written). In questions 2 and 3 (identical), the multiple choice items associated to learning objectives at the comprehension level (Q21 / Q31 and Q23 / Q33) were worth 4 points and items Q22e / Q32e and Q22f / Q32f referring to tasks of less complexity in relation to the other items of the same part of question 2 – also reports in the text. In total, the test has 19 items.

Table 10

Generalization assessment criteria

Question	Instructional Objective	Items	Underlying criteria in question form	Points
1	Evaluate the quality of the performance agreements produced, according to the knowledge acquired about the PA planning stage.	Q1a	Did the participant report if the performance agreement was built with the team's participation?	4
		Q1b	Did the participant report how the performance agreement will be monitored?	4
		Q1c	Did the participant write action plans to improve the performance of the fictional unit, including, excluding or altering the given performance agreement?	20 (max)
2 / 3	Relate performance description fragments with the PA follow-up stage fundaments	Q21 / Q31	Did the participant properly select a competence and the type to write the annotation to the fictitious character, observing the performance description?	4
	Write annotations for simulated situations and behaviors, according to the PA model.	Q22a / Q32a	Did the participant describe behavior or situation appropriate to the scenario given, annotation's type and the competence selected?	5
		Q22b / Q32b	Did the participant describe enhancement guidance appropriate to the scenario given and the competence selected indicating path with action to improve performance?	5
		Q22c / Q32c	Did the participant report if there was dialogue (verbal feedback) with the fictional character prior to the annotation?	5

Question	Instructional Objective	Items	Underlying criteria in question form	Points
2 / 3	Write annotations for simulated situations and behaviors, according to the PA model.	Q22d / Q32d	Did the participant report how the fictional character will be monitored?	5
		Q22e / Q32e	Did the participant report being available to assist the performance improvement of the fictional character?	4
		Q22f / Q32f	Did the participant use direct speech in the annotation, directing it to the fictional character?	4
	Give scores to simulated situations, relating to FI competencies and annotation types.	Q23 / Q33	Did the participant give an adequate score for the selected competence, considering that the behavior description was maintained throughout the assessment cycle?	4

Source: Author's elaboration

The data collection to develop the criteria was performed in the pilot group. After applying the pretests and posttests to the experimental groups (Experimental Group 1, Experimental Group 2 and Control Group), the correction of the tests was conducted by the own researcher. All the pre-tests and post-tests had no identification of the subjects (only a random code) and the final scores obtained were placed in a spreadsheet, for later statistical analysis (Study 4). The own researcher the correction of the tests using the correction sheet (Annex 3).

5.7 Study 4 – Quasi-experiment I

Study 4 is a quantitative approach study using primary data, with repeated measures (two-time points), to check the proposed hypotheses for generalization. First, there is a description of the variables of this study, indicating independent and dependent variables. The independent variable categorical variable with three levels, represented by each group (Experimental Group 1, Experimental Group 2 and Control Group). There is one dependent variable in this study – learning at generalization (measured in two different times, before and after the training, with two different but equivalent instrument, pretest, and posttest). These variables were measured with a continuous way, through pretest score (Spre) and posttest score (Spro). Figure 11 shows the composition of this study.

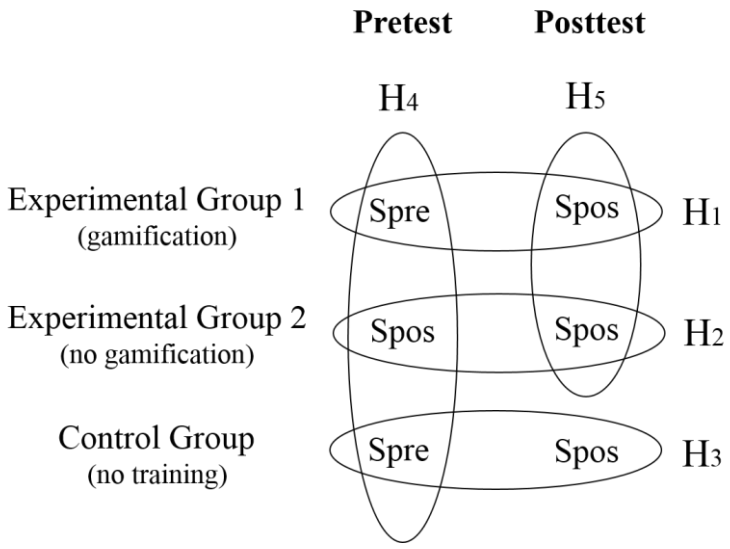

Figure 11. Study 4 – groups and hypotheses
Source: Author's elaboration

5.7.1 Within groups' hypotheses and analysis procedures

Below are seen as hypotheses of this study to the within groups analysis. Subsequently, are presented the statistical procedures made to check each of the hypotheses formulated and declared.

H_1: There will be an increase statistically significant between the participants' generalization from the pretest to the posttest for the Experimental Group 1.

H_2: There will be an increase statistically significant between the participants' generalization from the pretest to the posttest for the Experimental Group 2.

H_3: There will be no statistically significant difference between the pretest and posttest generalization for the Control Group.

Initially, it will be presented the descriptive statistic from the variables, including the values of the Shapiro-Wilk's test of normality for all variables and for each factor – a common assumption for all the parametric tests intended to be adopted in this study.

To test the within groups hypotheses presented, the paired-samples t-test are used to determine whether the mean difference between paired observations is statistically significantly different from zero. The participants are either the same individuals (from each of the groups) that were tested at two-time points. The paired-samples t-test is also referred to as the dependent t-test or repeated measures t-test. In order to run a paired-samples t-test, four assumptions need to be considered. The first two relate to the study's design choice and data nature (this research meets these requirements), while the second two relate to the paired-samples t-test itself that are going to be checked together with the test's execution:

- **Assumption 1:** There is one dependent variable measured at a continuous level, having the same participant being measured on two occasions.
- **Assumption 2:** There is one independent categorical variable.

5.7.2 Between groups' hypotheses and analysis procedures

Below are seen as hypotheses of this study to the between groups analysis. Further, are presented the statistical procedures made to check each of the hypotheses formulated and declared.

> H_4: There will be no statistically significant difference between the three groups (Experimental Group 1, Experimental Group 2 and Control Group) in generalization assessed in the pretest.
>
> H_5: There will be statistically significant difference between the generalization assessed in the posttest, with higher scores for participants of the Experimental Group 1 (Alcivar & Abad, 2016; Auvinen et al., 2015; De-Marcos et al., 2014; De-Marcos et al., 2016; Dominguez et al., 2013; El Tantawi et al., 2016; Filsecker & Hickey, 2014; Ge, 2018; Hamari et al., 2016; Ibanez et al., 2014; Smith, 2017).

To investigate H_4 is going to be used the one-way analysis of variance (ANOVA) to determine whether there are any statistically significant differences between the means of the three independent groups. In order to run the one-way ANOVA, six assumptions need to be considered. The first three assumptions are related to the study design's choice and the measurements made (this research meets these assumptions), while the second three assumptions relate to how the data fits the one-way ANOVA model and will be checked during the testing procedures. The first three assumptions are:

- **Assumption 1:** There is one dependent variable measured at a continuous level.
- **Assumption 2:** There is one independent variable consisting of three or more categorical, independent groups.
- **Assumption 3:** There is independence of observations – no relationship between the observations in each group of the independent variable or between the groups themselves. Independent groups are groups where there is no relationship between the participants in any of the groups – different participants in each group.

To investigate H_5 it is done independent-samples t-test (or independent t-test, for short) to compare the means between two unrelated groups on the same continuous, dependent variable. The initial assumptions are the same as presented for the one-way ANOVA only that

here there are only to two categorical groups, instead of three – reason to not use ANOVA (again this research meets these assumptions).

In this research, it was not possible to measure generalization in the pretest and posttest for the Control Group, with the same time lag (2 days interval) of the measure done for the other two groups (experienced the training in the classroom). Therefore, the statistical test (paired t-test) is being done to check if the pre and post tests were equivalent instruments – one of the requirements for assessing generalization, because the tests did not have the same questions, but measured the same learning objectives. Although the Control Group was a strategy for the quasi-experiment internal validity, it is plausible to assume that in the case of this research the maturation effect would not be able to show considerable differences, because of the short time lag of the intervention – around 34 hours long between pretest and posttest measures.

5.8 Study 5- Transfer of training assessment

In the same way, as for the creation of the assessment criteria to measure learning at generalization, this step used data (in this case secondary) of the participants of the pilot group. The dada was from secondary sources, records in the corporate system of the FI' PA tasks. As it is sequential research, the development of criteria to assess the transfer of training level was also based criteria adopted to assess generalization. This study demonstrates the methodological procedures to construct the criteria that made possible to assess learning at the transfer of training level (quasi-experiment II). The tasks executed by managers regarding PA are recorded in text format. Further, it is also discussed the rates' creation (dependent variables) that will be analyzed statistically in study 6. The specificity of these tasks is also be explained below.

5.8.1 Assessment criteria identification and selection

Transfer assessment was done through an analytical process of secondary data provided by the FI. All the institution's managers insert these data referring to tasks associated with the PA model. FI has a system where the PA task information is entered. All managers are required to perform this task. The performance agreement is mandatory, and the registration is given at the beginning of the evaluation semester (which is six months in length.) The registration of

annotations is not obligatory, though highly recommended. At the end of the semester, managers assign concepts based on a Likert scale for all their team members.

The three tasks performed by managers in the workplace initially interested in this study as they were directly related to the instructional objectives worked in the instructional events and could be provided by the FI through data reports:

(1) **Team performance agreement:** This task is performed once every six months, at the beginning of the PA cycle, and is the guiding tool for the entire PA evaluation process.

(2) **Writing annotations:** This task can be performed at any time by the managers. The annotations can be of two types, recognition or enhancement, and must be associated with professional competence, among the competencies available to selection.

(3) **Scores' emission:** At the end of each PA cycle, managers give scores to their team members, supporting their decisions in what was agreed in the performance agreement and the annotations given during the PA cycle.

In order to decide about using the secondary data mentioned above, the pilot group participants' had their performances evaluated. These same performances were the basis for decision-making concerning the criteria developed for assessing the transfer of training.

After the data analysis conducted by the own researcher, a decision was made that the performance agreements and the participants' scores emission would not be assessed. This decision is justified by the fact that performance agreements are tools that bring specific aspects of the contextual reality of each unit of the FI. Therefore, their aspects could not be checked by the researcher to measure quality and whether if the transfer has occurred or not. The decision to not use scores' emissions was based on the specificity of the KSAs evaluated and that any change of behavior in the emission of the concepts would be insufficient to conclude whether it represented the transfer of training positively or negatively (improvement or not in their performances).

The analysis of some aspects of the managers' behavior was considered to develop the assessments criteria. It was noticed that some aspects could influence the transfer of training measures and were mapped and considered for purposes of this study: (1) the number of team members (more opportunities to apply the KSA acquired) and (2) concentration of certain behaviors in a specific period of the PA cycle (e.g. higher annotations' concentration in the PA cycle final month).

After analyzing these peculiarities, it was decided to evaluate the data in a total period of 14 months – seven months before and seven months after the instructional event's end.

Figure 12 reports the evaluated period and the PA cycle change points.

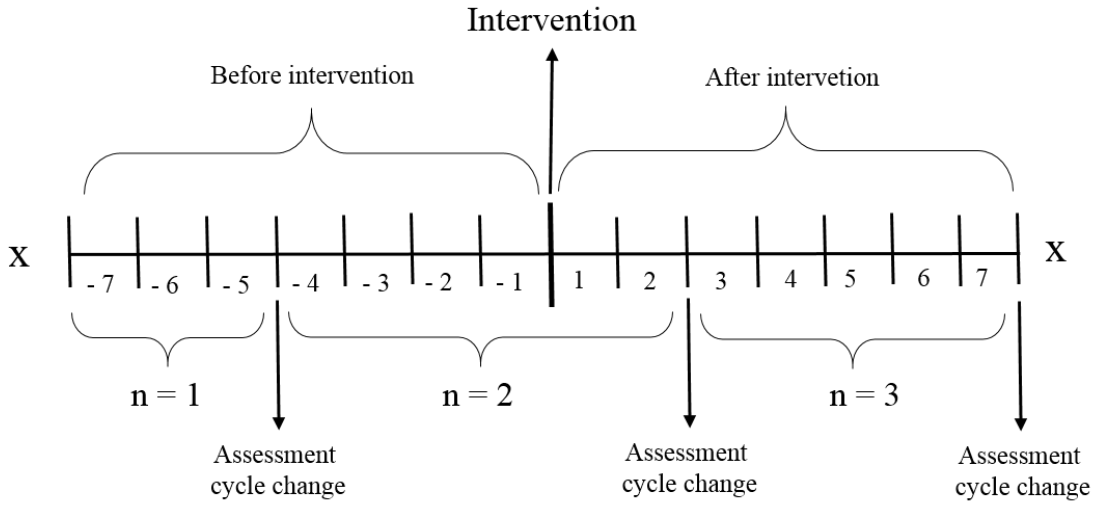

Figure 12. Transfer of training periods and cycles
Source: Author's elaboration.

The PA cycle changing times were marked, since with at each new cycle there may be a change in the manager's team members (individuals and size). The value that measures the team's size is represented as "NrE", and it was measured in the three cycles covered by the selected data analysis period (14 months in total).

The data concerning the annotations issued by all the participants' pilot group were collected from the corporate system of FI where they are registered. The criteria for evaluation were constructed based on the data received and the criteria used to evaluate the generalization. All sums were placed on a spreadsheet, without identification of subjects. In this step, there is no assignment of scores, that there is only a count of the number of times that each behavior, measured by the criteria adopted, manifested itself in each moment. These sums (numerical values) are not the dependent variables of this study. The next step (rate creation) will define the variables dependent on the study 5.

The results of the pilot group data analysis generated the following set of sums (further they will be transformed in the dependent variables of this study). Table 11 presents the codes, the operational definition and the reasons for identifying each of the sums.

Table 11
Transfer of training criteria

Sum	**Description**	**Reason to be investigated**
∑NrE	Team members ' sum evaluated in each PA cycle.	Influences the possibility of practicing the knowledge acquired in the instructional event

Sum	Description	Reason to be investigated
∑TtAn	Total annotation's sum	Annotations' quantity measure the will to execute this task
∑RpAn	Repeated annotations' sum	Below-expected performance that massifies the annotations' writing, damaging the PA process
∑AnQa	Annotations' sum in which behavior or situation has been properly described to the type of annotation and the selected competence	Related to generalization variables Q22a / Q32a
∑AnQb	Annotations' sum in which enhancement guidance has been properly described to the selected competency, indicating path with action to improve performance.	Related to generalization variables Q22b / Q32b
∑AnQc	Annotations' sum in which there was a report of a dialogue (verbal feedback) before the annotation.	Related to generalization variables Q22c / Q32c
∑AnQd	Annotations' sum in which the performance monitoring was reported.	Related to generalization variables Q22d / Q32d
∑AnQe	Annotations' sum in which the manager has made himself/herself available to assist in the performance improvement.	Related to generalization variables Q22e / Q32e
∑AnQf	Annotations' sum in which there was the use of direct speech	Related to generalization variables Q22f / Q32f

Source: Author's elaboration

The annotations' tasks shown to be more sensitive to the training effects during the pilot group data analysis, since they reflected a performance that occurs with no date and quantity defined, and may have more representativeness in terms of behavioral change and transfer of training. All sums were grouped in each period selected for the study (seven months before and seven months after the instructional event), and measured in the 14 months of the studied period.

The sum of total annotations issued by the participants (ΣTtAn) is of interest in this study, although it is the only task that can manifest a change in behavior' quantity. The repeated annotations' sum (ΣRpAn), is the only variable chosen that represents performance below what

the FI expects from its managers (making the same annotations for individuals with different performances).

The other annotations' sums identified and selected were based on the same criteria adopted in the assessment of the generalization effect (ΣAnQa, ΣAnQb, ΣAnQc, ΣAnQd, ΣAnQe, and ΣAnQf). The behavioral change was identified through the participant's demonstration of these aspects in their annotations text. This measurement, as well as the others regarding the issue of annotations, consider the total period of data analysis, separating behaviors executed in the months before and after the intervention – training event.

All this qualitative data analysis was made by the own researcher. First, trained evaluators external to the FI would not be able to interpret all the peculiarities and expressions used by the participants in their texts (the researcher was immersed for months in the theme and was also present in all classes done). Second, the FI did not have specialists available to do this assessment, due to the high amount of data analyzed and the work hours required for it. After finishing the sums' definition process, the annotations of the three groups participants, Experimental Group 1, Experimental Group 2 and Control Group (N = 53) were evaluated. All the sums were taken and placed on a spreadsheet, without identifying the participants. The values were measured month by month, for each of the 14 months that comprised the period of analysis. The reading of each annotation, in total, 1,104, were done in around one hundred hours of work by the own researcher

The next section explains how rates were constructed to represent the transfer of training performance. These rates are study 6 dependent variables.

5.8.2 Rates' creation

After finishing the organization of the sums made in the previous study, it was noticed that the transfer of training occurred differently depending on each participant. As the participant's behavioral change manifested distinctly for each sum, it was necessary the creation of rates to measure the quantity and quality of the behavioral change (transfer of training variables), based on the analysis of the annotations collected in the FI's PA system.

In this rate creation stage, all the sums were already organized in a spreadsheet by period and individual. The rates' creation was a mathematical calculation, for three groups' participants' sums (Experimental Group 1, Experimental Group 2 and Control Group). The rates are the dependent variables of Study 6. The details of the calculations of each rate are described below.

Each one of these rates was developed to represents one aspect of the behavioral change – better or worse performance of the three groups participants, and are presented below.

(a) Total annotations per team member's rate – An: This rate measures criteria related to annotation's quantity. Its measurement is crucial because it may represent a greater willingness of the quasi-experiment participants to apply the KSAs acquired in their work functions. The number of team members' sum (∑NrE) in the PA model is a factor that should be considered because it represents a greater possibility of practicing the KSA acquired in the training event. The condition that relates the month (represented by the variable "x") and the PA cycle (represented by the variable "n") when the analyzed behavior happened, is the same used for the other rates built that are presented below.

$$An(x,n) = \frac{\sum TAn(x)}{\sum NrE\ (n)}$$

$$\text{if} \begin{cases} -7 < x \leq -5;\ n = 1 \\ -4 \leq x \leq 2;\ n = 2 \\ 3 \leq x \leq 7;\ n = 3 \end{cases}$$

- "x" is the month of the measure, from "-7" to "+7" (seventh month before to seventh month after the intervention).
- "n" is the PA cycle identificator, from 1 to 3 (the three cycles covered by the total period of time).

**the same conditions were applied in the next rates' presented.*

(b) Replicated annotation per team member's rate – Rp: This rate represents inappropriate managers' behavior when executing their PA tasks. Replicated annotations broaden the approach to talent development, one of the premises of the FI's PA. Each employee should have their behavior observed and receive annotations to recognize performances (when equal to or higher than expected by the FI) or annotations to improve performances (when below to what is expected by FI).

$$Rp(x,n) = \frac{\sum RpAn(x)}{\sum NrE\ (n)}$$

(c) Behavior description quality per team member's rate – Bd: This rate comes from a field (mandatory) where the situation or behavior observed in the PA process must be written. The description should be sufficient so that it can be related to the competence and the annotation type chosen. This text is mandatory in both recognition and enhancement annotations.

$$Bd(x,n) = \frac{\sum AnQa(x)}{\sum NrE\ (n)}$$

(d) Enhancement guidance quality per team member's rate – Ge: This rate comes from a field (mandatory only for enhancement annotations) that must be filled with guidance text to improve the performance of their team members. This text should contain a path to be followed and an indication of action that must be taken to improve performance.

$$Ge(x,n) = \frac{\sum AnQb(x)}{\sum NrE\ (n)}$$

(e) Management commitment per team member's rate – Mc: This rate is a grouping of four different criteria that emerged from the researcher's analytical work in evaluating the managers' annotations. They were grouped because they are not related to fields (mandatory or not) that need to be filled when performing PA tasks. They are textual aspects that arise and show the commitment of the manager (who writes the annotation) with the formation of their team member. The four aspects of the rate are: (1) dialogue report (verbal feedback) should occur before the annotation – fact that approximates the manager and the team member; (2) indication of how the monitoring process will be carried out – shows the manager's commitment to verify the improvement in the team member performance; (3) availability to assist in the development of KSA – shows explicitly the manager's commitment with assisting the team member development; and (4) use of direct discourse – impersonal language (directed to the system or to the FI) shows the manager's distance from the development process as opposed to the direct discourse, which has a strength of approximation between the manager and the team member (e.g. call by the first name of who is receiving the annotation).

$$Mc(x,n) = \frac{\sum AnQc(x) + \sum AnQd(x) + \sum AnQe(x) + \sum AnQf(x)}{\sum NrE\ (n)}$$

5.9 Study 6 – Quasi-experiment II

Study 6 is a quantitative approach study using secondary data, with a longitudinal design, to check the effects of gamification on the transfer of training. This study is the final investigation vital to reach the main objective research proposed. Transfer of training is a construct that can be measured in several ways. In this research, it was evaluated with the analysis of annotations written by FI's managers, which generated a set of rates representing aspects of their behaviors in the workplace.

First, there is a description of the dependent and independent variables of this study. The independent variable is categorical with three levels, represented by each group (Experimental Group 1, Experimental Group 2 and Control Group), the same of quasi-experiment I. There are five dependent variables in this study – rates measuring aspects of transfer of training (measured in fourteen different times, seven before and seven after the intervention): Ta, Rp, Bd, Ge, and Mc (their constitutive definition and calculation were shown in the previous section).

During the annotations' data analysis and the creation of each rate, it was already observed that the data would probably not meet the assumption of normality. The descriptive statistic will show if this assumption is confirmed. The behavior measures were distinct from one another – some individuals changed the number of behaviors while others had no change at all with zero observations. Therefore it will be indicated in this section the decision to execute non-parametrical statistical tests (that only will be applied after tests to confirm the data non-normality).

5.9.1 Within groups' hypotheses and analysis procedures

The main purpose of this phase is to determine whether there are any statistically significant differences between the repeated measures in each rate (Ta, Rp, Bd, Ge and Mc) and in all the three groups (Experimental Group 1, Experimental Group 2 and Control Group).

As the measurements occurred in a longitudinal design (14 months) in relation of a training event of 16 hours duration, it was decided to fractionate the time for the statistical

analyzes in smaller equidistant periods from the intervention. Thus, it will be possible to investigate the different transfer of training outcomes within each group, at distinct periods – closer or more distant from the intervention. Figure 6 shows the division into smaller intervals made, to carry out the statistical analyses.

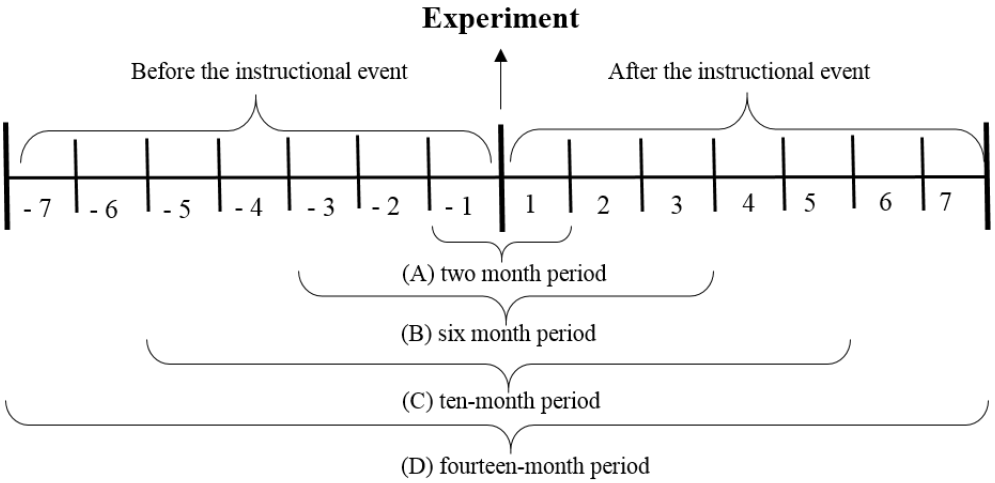

Figure 13. Study 5 – Intervals division.
Source: Author's elaboration

The interest is finding statistically significant differences between periods before and after the intervention (it is the training events effects the aim of this investigation). In total for the rates' comparisons within groups, the within group investigation has a total of 60 hypotheses, which are grouped by rates, according to the interval the analysis belongs.

Set 1 – Interval A (2 months)

H_1 / H_4 / H_7 / H_{10} / H_{13}: There will be statistically significant difference between paired observations (interval A) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 1 participants.

H_2 / H_5 / H_8 / H_{11} / H_{14}: There will be statistically significant difference between paired observations (interval A) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 2 participants.

H_3 / H_6 / H_9 / H_{12} / H_{15}: There will **no** be statistically significant difference between paired observations (interval A) in Ta / Rp / Bd / Ge / Mc for the Control Group.

Set 2 – Interval B (6 months)

H_{16} / H_{19} / H_{22} / H_{25} / H_{28}: There will be statistically significant differences between the distributions of repeated measures (interval B) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 1 participants.

H_{17} / H_{20} / H_{23} / H_{26} / H_{29}: There will be statistically significant differences between the distributions of repeated measures (interval B) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 2 participants.

H_{18} / H_{21} / H_{24} / H_{27} / H_{30}: There will be **no** statistically significant differences between the distributions of repeated measures (interval B) in Ta / Rp / Bd / Ge / Mc for the Control Group participants.

Set 3 – Interval C (10 months)

H_{31} / H_{34} / H_{37} / H_{40} / H_{43}: There will be statistically significant differences between the distributions of repeated measures (interval C) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 1 participants.

H_{32} / H_{35} / H_{38} / H_{41} / H_{44}: There will be statistically significant differences between the distributions of repeated measures (interval C) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 2 participants.

H_{33} / H_{36} / H_{39} / H_{42} / H_{45}: There will be **no** statistically significant differences between the distributions of repeated measures (interval C) in Ta / Rp / Bd / Ge / Mc for the Control Group participants.

Set 4 – Interval D (14 months)

H_{46} / H_{49} / H_{52} / H_{55} / H_{58}: There will be statistically significant differences between the distributions of repeated measures (interval D) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 1 participants.

H_{47} / H_{50} / H_{53} / H_{56} / H_{59}: There will be statistically significant differences between the distributions of repeated measures (interval D) in Ta / Rp / Bd / Ge / Mc for the Experimental Group 2 participants.

H_{48} / H_{51} / H_{54} / H_{57} / H_{60}: There will be **no** statistically significant differences between the distributions of repeated measures (interval D) in Ta / Rp / Bd / Ge / Mc for the Control Group participants.

All hypotheses that expect statistically significant differences of training events (Experimental Group 1, Experimental Group 2), and all the hypotheses that expect **no** statistically significant differences in the Control Groups are based on research reviews that affirms the positive effects of T&D systems on positive behavioral changes (Bell et al., 2017).

All the hypotheses situated in interval A (from H_1 to H_{15}) will be analyzed using the non-parametrical Signed test – used to determine whether there is a median difference between paired observations. This test is adopted when the distribution of differences between paired observations is neither normal nor symmetrical (this is a plausible consideration for the same reason justified above related to normality). The participants are the same individuals tested on two occasions the same dependent variable (transfer of training measures through the rates built). There are two basic assumptions for this test (this research meet these assumptions):

- **Assumption 1:** There is one dependent variable measured at a continuous level.
- **Assumption 2:** There is one independent variable consisting of two categorical, related groups or matched pairs.

In the study design for interval A, the goal is to determine if there are changes in the rates between two-time points (the first month before and the first month after the intervention). In between these two time points, there was the intervention.

For all the hypotheses formulated for the intervals B, C and D (from H_{16} to H_{60}), it will be used the Friedman test – a non-parametric test adopted to determine whether there are any statistically significant differences between the distributions of three or more related groups. The groups are related as they contain the same cases (e.g., participants) in each group, and each group represents a repeated measurement on the same dependent variable. This test is also used if the assumption of normality is markedly violated.

There are basic requirements to use the Friedman test. The first two are related to the study design's choice, while the other three reflect the data's nature. The first and second assumptions are:

- **Assumption 1:** There must be one dependent variable that is measured at the continuous or ordinal level.
- **Assumption 2:** There must be one independent variable consisting of three or more categorical, related groups or matched cases. Related groups indicate that the groups are not independent.

This study's design meets assumptions 1 and 2, so it is required to explain the null and alternative hypothesis for the Friedman test's execution. The null hypothesis for a Friedman test is: H_0: the distribution of rates in each group are the same. Complementary, the alternative hypothesis is: H_A: at least two of the groups' distributions differ (possibly for location; e.g., median). While the Friedman test allows determining whether there is an overall effect of the independent variable on the dependent variable, it does not inform which of the groups differ from each other. In order to know where any differences lie, it is necessary to do a post hoc test. Although there are exceptions, most post hoc tests consider all possible variations of group comparisons.

A Friedman test is often adopted for types of study design that want to determine if there are differences between three or more time points. This study wants to investigate how the variables (measured through the rates created) change over time in the same participants and compare three or more time points. In this study, between the time points, there was the intervention in two of the three groups, the training event. In the case of this repeated measure test, there is an interest to observe the same variable at all times points.

5.8.2 Between groups' hypotheses and analysis procedures

The main purpose of this phase is to determine if the median of the rates created (Ta, Rp, Bd, Ge and Mc) of at least one group is different from the median of another group (Experimental Group 1, Experimental Group 2 and Control Group) in the fourteen months measured in this study. Figure 14 illustrates the between the groups analysis that will be executed.

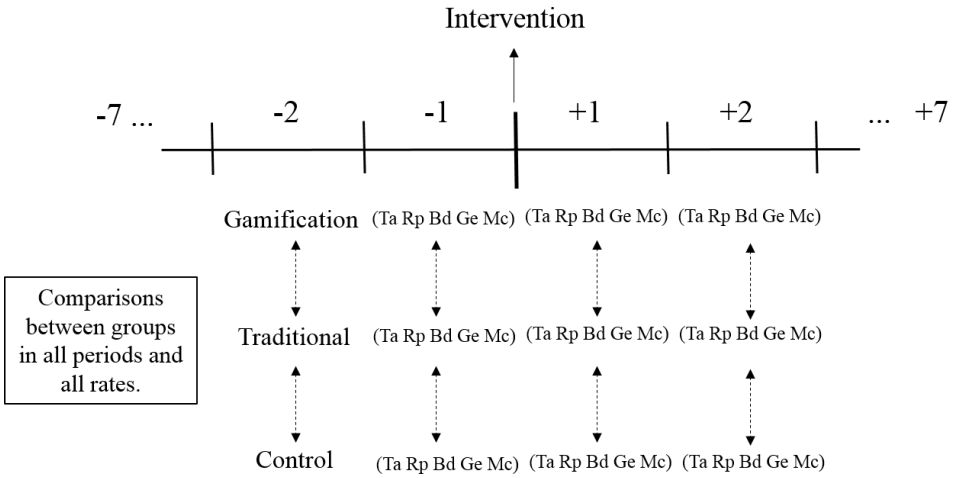

Figure 14. Study 5 – Between groups' analysis
Source: Author's elaboration

The research was focused on finding no statistically significant difference between neither groups nor rates in any period before the intervention. The following hypotheses refers to this assumption.

> H_{61} / H_{62} / H_{63} / H_{64} / H_{65}: There will be **no** statistically significant differences between the three groups (Experimental Group 1, Experimental Group 2 and Control Group) in the rates Ta / Rp / Bd / Ge / Mc, before the intervention.

For comparisons between groups in the periods after the intervention, was necessary to formulate hypotheses considering three different factors: the period (month of comparison – seven after the intervention), the comparisons between groups (Experimental Group 1 with Experimental Group 2; Experimental Group 1 with Control Group; and Experimental Group 2 with Control Group); and the five rates created (Ta, Rp, Bd, Ge and Mc). The easiest logic to organize these hypotheses was through a table creation. As tables have only two dimensions, it was decided to repeat the factor of smaller number (the comparison between the groups). This rational resulted in Tables 12, 13, and 14 – which contains all hypotheses between groups for this longitudinal study.

The hypotheses in Table 12 complete the sentence: Participants in Experimental Group 1 will present statistically significant differences compared to participants in Experimental Group 2, in the rate **<row>** in the **<period in the column>**.

Example (row 1, column 1), H_{66}: Participants in Experimental Group 1 will present statistically significant differences compared to participants in Experimental Group 2, in the rate Ta in the first month after the intervention.

Table 12
Hypotheses for comparisons between Exp. Groups 1 and 2

Variable	**Periods**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Ta	H_{66}	H_{67}	H_{68}	H_{69}	H_{70}	H_{71}	H_{72}
Rp	H_{73}	H_{74}	H_{75}	H_{76}	H_{77}	H_{78}	H_{79}
Bd	H_{80}	H_{81}	H_{82}	H_{83}	H_{84}	H_{85}	H_{86}
Ge	H_{87}	H_{88}	H_{89}	H_{90}	H_{91}	H_{92}	H_{93}
Mc	H_{94}	H_{95}	H_{96}	H_{97}	H_{98}	H_{99}	H_{100}

Source: Author's elaboration

The hypotheses in Table 13 complete the sentence: Participants in Experimental Group 1 will present statistically significant differences compared to participants in Control Group, in the rate **<row>** in the **<period in the column>**.

Table 13
Hypotheses for comparisons between Exp. Groups 1 and Control Group

Variable	**Periods**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Ta	H_{101}	H_{102}	H_{103}	H_{104}	H_{105}	H_{106}	H_{107}
Rp	H_{108}	H_{109}	H_{110}	H_{111}	H_{112}	H_{113}	H_{114}
Bd	H_{115}	H_{116}	H_{117}	H_{118}	H_{119}	H_{120}	H_{121}
Ge	H_{122}	H_{123}	H_{124}	H_{125}	H_{126}	H_{127}	H_{128}
Mc	H_{129}	H_{130}	H_{131}	H_{132}	H_{133}	H_{134}	H_{135}

Source: Author's elaboration

The hypotheses in Table 14 complete the sentence: Participants in Experimental Group 2 will present statistically significant differences compared to participants in Control Group, in the rate **<row>** in the **<period in the column>**.

Table 14
Hypotheses for comparisons between Exp. Groups 2 and Control Group

Variable	**Periods**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Ta	H_{136}	H_{137}	H_{138}	H_{139}	H_{140}	H_{141}	H_{142}
Rp	H_{143}	H_{144}	H_{145}	H_{146}	H_{147}	H_{148}	H_{149}
Bd	H_{150}	H_{151}	H_{152}	H_{153}	H_{154}	H_{155}	H_{156}
Ge	H_{157}	H_{158}	H_{159}	H_{160}	H_{161}	H_{162}	H_{163}
Mc	H_{164}	H_{165}	H_{166}	H_{167}	H_{168}	H_{169}	H_{170}

Source: Author's elaboration

Additionally, the expected better rates differences Experimental Group 1 in relation to Experimental Group 2 are based on the analysis of the empirical articles on gamification which indicate positive results for learning and engagement (Alcivar & Abad, 2016; Ambrosio & Garofalo, 2016; Auvinen et al., 2015; Dias, 2017; De-Marcos, et al., 2014; De-Marcos, et al., 2016; Dominguez et al., 2013; El Tantawi et al., 2016; Filsecker & Hickey, 2014; Ge, 2018; Hamari et al., 2016; Hew et al., 2016; Huang & Hew, 2018; Ibanez, Di-Serio, & Delgado-Kloos, 2014; Landers & Landers, 2015; Mekler, Brühlmann, Tuch, & Opwis, 2015; Paiva et al. 2016; Pettit, et al., 2015; Smith, 2017; Sun & Hsieh, 2018; Tan & Hew, 2016; Tenorio et al., 2016; Yildirim, 2017). The hypotheses that expect better for the Experimental Groups 1 and 2 in comparison to the Control Group are based in reviews on efficacy of training events in the transfer of training (Bell et al., 2017).

In order to check the formulated hypotheses (from H_{61} to H_{170}) it will be used the Kruskal-Wallis H test, a rank-based nonparametric test to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.

The following assumptions must be met to use this test. The first three relate to the study design's choice, while the fourth reflects the data's nature (this research meet all these assumptions):

- **Assumption 1:** There must be one dependent variable, measured at the continuous level.
- **Assumption 2:** There must be one independent variable, consisting of three or more categorical, independent groups.
- **Assumption 3:** There must be independence of observations – no relationship between the observations in each group of the independent variable or between the groups themselves.

There is another assumption to determine whether the distribution of scores for each group of the independent variable have the same shape or different shape. This assumption will be tested together with the statistical analysis.

Kruskal-Wallis H test is statistically significant (i.e., $p < .05$) only indicates that the median of at least one group is different from the median of another group. To discover which group(s) are different from which other groups are necessary to run a post hoc test. In the case of the Kruskal-Wallis H test, it will be done and interpreted pairwise comparisons using Dunn's (1964) procedure with Bonferroni adjustment.

6 Results

6.1 Study 4 – Generalization assessment

Here are presented the results of this study that measure participants' generalization. The independent variable categorical variable with three levels, represented by each group (Experimental Group 1, Experimental Group 2 and Control Group). There is one dependent variable in this study – learning at generalization (measured in two different times, before and after the training, with two different but equivalent instrument, pretest, and posttest).

6.1.1 Within groups' findings

Initially the results of the comparisons within groups will be presented, in a repeated measures design, with a pretest and posttest application for the generalization measure. The hypotheses follows.

H_1: There will be an increase statistically significant between the participants' generalization from the pretest to the posttest for the Experimental Group 1.

H_2: There will be an increase statistically significant between the participants' generalization from the pretest to the posttest for the Experimental Group 2.

H_3: There will be no statistically significant difference between the pretest and posttest generalization for the Control Group.

The descriptive statistics showed no outliers as also detected through histograms visualization. The assumption of normality was not violated, as assessed by Shapiro-Wilk's test (Table 15) and also by graphical analysis of the boxplots.

Table 15
Tests of Normality

Variable	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Spre	0.098	53	.200*	0,977	53	.405
Spos	0.083	53	.200*	0,984	53	.690

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The paired-samples t-test was adopted to determine whether there was a statistically significant mean difference between the participants' scores in pretest and posttest for all three groups. Concerning H_1, participants of the gamified group had better performance in the post test (42.32 ± 12.8) as opposed to the pretest (30.00 ± 13.1), a statistically significant increase of 12.30 (95% CI, -20.62 to -4.01), $t(18) = -3.11$, $p = .006$, $d = .71$. Results regarding H_2 also showed that the participants of the Experimental Group 2 obtained higher results in the posttest (38.16 ± 12.5) compared to the pretest (25.16 ± 12.4), a statistically significant increase of 13 (95% CI, -20.80 to -5.19), $t(18) = -3.49$, $p = .003$, $d = .80$. On the contrary of the first and second result shown, the participants of the Control Group did not show any statistical significant difference between results in the posttest (27.27 ± 7.1) and the pretest (25.60 ± 12.2), $t(14) = -0.58$, $p = .571$, $d = .14$.

Table 16
Descriptive Statistic and paired t-test results

	Pretest		Posttest			95% CI for Mean			
Group	M	SD	M	SD	n	Difference	*p*	t	df
Gamification	30.00	13.1	42.32	12.8	19	-20.62, -4.01	.006*	-3.11	18
Traditional	25.16	12.4	38.16	12.5	19	-20.80, -5.19	.003*	-3.49	18
Control	25.60	12.2	27.27	7.1	15	-7.82, 4.49	.571	-0.58	14

* $p < .05$

The within groups analysis results demonstrated that Experimental Group 1 and Experimental Group 2 had mean differences statistically significant different from zero. The comparison between pretest and posttest scores indicated that the two trainings were effective, as they resulted in generalization of learning. The result for the Control Group showed mean difference between pretest and posttest was statistically equal to zero. As both the pretest and posttest were applied in the Control Group at the same time, without temporal lag, this result shows that the items are equivalent in complexity. Therefore, we can accept all three hypothesis – H_1 , H_2, and H_3.

6.1.2 Between group's findings

Straightaway, the results comparing between groups will be presented. These are hypotheses formulated.

> H_4: There will be no statistically significant difference between the three groups (Experimental Group 1, Experimental Group 2 and Control Group) in generalization assessed in the pretest.
>
> H_5: There will be statistically significant difference between the generalization assessed in the posttest, with higher scores for participants of the Experimental Group 1 (Alcivar & Abad, 2016; Auvinen et al., 2015; De-Marcos et al., 2014; De-Marcos et al., 2016; Dominguez et al., 2013; El Tantawi et al., 2016; Filsecker & Hickey, 2014; Ge, 2018; Hamari et al., 2016; Ibanez et al., 2014; Smith, 2017).

A one-way ANOVA was conducted to determine if the score in the pretest was different for the three groups, Experimental Group 1 (N = 19), Experimental Group 2 (N = 19) and Control Group (N = 15). There were no outliers, as assessed by boxplot. Shapiro-Wilk's test

had already shown the data normality, and there was the homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .983). The higher score happened in the Experimental Group 1 (M = 30, SD = 13.2), followed by the Control Group (M = 25.6, SD = 12.3), finally by the Experimental Group 2 (M = 25.2, SD = 12.5), though the differences between the groups' scores in the pretest were not statistically significant, $F(2, 50) = 0.825$, p = .444 (Table 17). As the groups' means were not statistically significant different, the proposed hypothesis (H_4) is accepted.

Table 17
One-way ANOVA results comparing pretests

Scores in pretest	Sum of Squares	*df*	Mean Square	*F*	*p*
Between Groups	264.85	2	132.42	0.825	.444
Within Groups	8026.12	50	160.52		
Total	8290.98	52			

There were 19 participants in the Experimental Group 1 and also 19 in the Experimental Group 2. An independent-samples t-test was run to determine if there were differences in the learning performance measured through the posttest scores for these two quasi-experimental groups that experienced the training event, in order to check H_5. After a data inspection via boxplot, it was found one outlier in the Experimental Group 1 (outlier identified in the posttest for the analysis between these two groups). The decision made was to keep the outlier in the analysis, after running independent-samples t-tests with and without it. The variable was normally distributed, as assessed by Shapiro-Wilk's test, and there was the homogeneity of variances, as assessed by Levene's test for equality of variances (p = .752). Table 15 shows the results for the independent t-test. The scores in the posttest were higher for the gamified group participants (M = 42.63, SD = 12.139) when compared to the Experimental Group 2 (M = 38.16, SD = 12.518), though the result that was not statistically significant difference, M = 4.47, $t(36) = 1.118$, p = .271. As the groups' means were not statistically significant different, the proposed hypothesis (H_5) is rejected.

Table 18
Independent t-test – Scores in the posttest

	Groups						95% CI for			
	Gamification			Traditional			Mean			
	M	SD	n	M	SD	n	Difference	t	d	*p*
Spos	42.63	12.3	19	38.16	12.5	19	-3.63, 12.5	1.118	36	.752

* p < .05.

Table 19 summarizes the hypotheses tested in study 4.

Table 19
Findings summary – Study 4

Hypotheses	**Statistical test**	**Result**
H_1	The paired samples *t*-test	Accepted
H_2	The paired samples *t*-test	Accepted
H_3	A paired samples *t*-test	Accepted
H_4	One way ANOVA	Accepted
H_5	Independent *t*-test	Rejected

Source: Author's elaboration

6.2 Study 6 – Transfer of training assessment

In this section, the results of study 6 will be presented. This study had a quantitative approach using secondary data, with a longitudinal design, to check the effects of gamification on the transfer of training. In this research, it was evaluated with the analysis of annotations written by FI's managers, which generated a set of rates representing aspects of their behaviors in the workplace. Relating to the variables, the independent was a categorical variable with three levels, represented by each group (Experimental Group 1, Experimental Group 2 and Control Group), the same of quasi-experiment I. The dependents were the five rates measuring aspects of transfer of training, in fourteen different times, seven before and seven after the intervention – Ta, Rp, Bd, Ge, and Mc.

6.2.1 Within groups' findings

As explained in the method section, the analysis within groups was performed in different intervals of the total period analyzed in this study. The principal objective of this phase is to determine whether there are any statistically significant differences between the repeated measures in each rate (Ta, Rp, Bd, Ge and Mc) and in all the three groups (Experimental Group 1, Experimental Group 2 and Control Group). The hypotheses were grouped according to the interval division. Interval A covered a period of two months (one before and one after the intervention), the interval B a period of six months (three before and three after), interval C a period of ten months and lastly interval D the entire 14 months period. The results are presented respecting this order.

(1) Interval A – Signed test (two-month period)

The Signed test was used because the data was neither normal nor symmetrical as visualized by histograms. The statistical significance of this test was calculated using binomial distribution – an approximate *p-value* obtained by the normal approximation to the binomial. The exact test, based on the binomial distribution, is implemented for this data set because there are less than 25 cases.

The results for the Ta came from the sign test with continuity correction, comparing differences in the rate for the two measures – the first month before intervention and the first month after the intervention. The Experimental Group 1 and Experimental Group 2 showed a statistically significant increase in Ta comparing before and after the quasi-experiment, $p = .002$ and $p = .012$, respectively. Control Group did not show any statistically significant difference, $p = .727$.

For Rp rate, all the three groups presented no statistically significant difference. Instead, Bd had a statistically significant rate increase for the Experimental Group 1, $p = .006$, though Experimental Group 2 and Control Group showed no increase, $p = .070$ and $p = 1.000$, respectively. For the Ge rate was observed the same result as for Rp, no statistically significant difference for all three groups. The rate Mc had a result similar to Bd, when only the Experimental Group 1 had a statistically significant difference, $p = .004$, while Experimental Group 2 and Control Group showed no increase, $p = .125$ and $p = 1.000$, respectively.

(2) Interval B – six-month period

The results of the analysis reported are from a period of six months – three before and three after the intervention. All the rates were assessed using the Friedman test. The results follow.

For the Ta rate, the test showed that the differences were not statistically significant for the Control Group, $\chi 2(5) = 5.000$, $p = .416$. For the Experimental Group 2, initially it was showed significant result ($\chi 2(5) = 12.609$, $p = .027$), but after the Bonferroni correction ($p < .0005$ level) no significant difference was obtained in the pairwise comparison.

Concerning the Experimental Group 1, the Friedman test result was statistically significant. Pairwise comparisons were executed with a Bonferroni correction for multiple comparisons. The difference in Ra for this six-month period was statistically significantly different ($\chi 2(5) = 32.684$, $p < .001$). The results are graphically reinforced by Figure 15, showing the periods with a statistically significant difference between rates' mean ranks in each period.

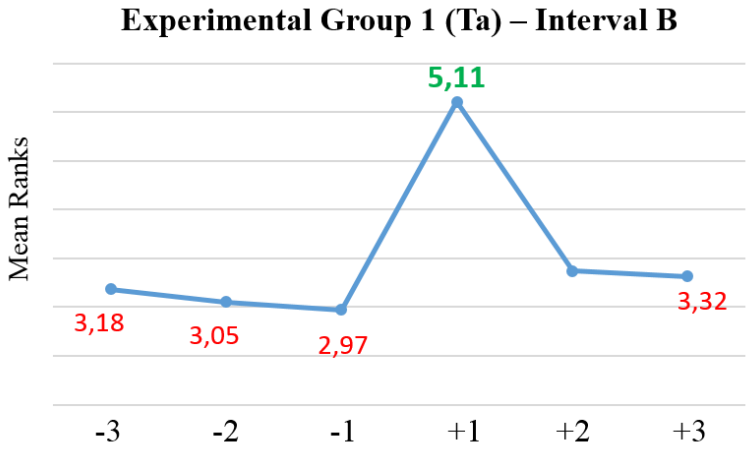

Figure 15. Total annotations rate in interval B – Exp. Group 1
Source: Author's elaboration

Friedman test showed no statistically significant differences for the Control Group and Experimental Group 2 for the Rp rate, $\chi 2(5) = 10.154$, $p = .071$, and $\chi 2(5) = 6.854$, $p = .232$, respectively. For the Experimental Group 1, initially it was showed significant result ($\chi 2(5) = 13.257$, $p = .021$), but after correction ($p < .0005$ level) there was no significant difference between the mean ranks.

For the Bd rate in this period, the test showed no statistically significant differences for the Control Group, $\chi 2(5) = 4.219$, $p = .518$. For Experimental Group 1 and Experimental Group 2, the difference was initially significant, though after the correction, no statistically significant, $\chi 2(5) = 26.246$, $p < .0005$, and $\chi 2(5) = 12.044$, $p = .034$, respectively.

Friedman test showed differences statistically significant for the Ge rate only before the corrections for the Experimental Group 1 and Experimental Group 2, $\chi 2(5) = 12.414$, $p = .030$, and $\chi 2(5) = 12.868$, $p = .025$, respectively. Again, for the control, no differences statistically significant were found, $\chi 2(5) = 8.200$, $p = .146$.

The Experimental Group 1, for the Mc rate in this six-month period Friedman test, was initially significant, but later no differences were considered after Bonferroni correction, $\chi 2(5) = 26.224$, $p < .0005$. For traditional and Control Groups, no difference was found, $\chi 2(5) = 10.950$, $p = .052$, and $\chi 2(5) = 4.000$, $p = .549$, respectively.

(3) Interval C – ten-month period

Friedman tests' results are reported from this ten-month period (five before and five after the intervention) for all groups in all rates created.

The Experimental Group 1, for the Ta rate in this ten-month period Friedman test, showed statistically significant differences between distributions, $\chi 2(9) = 40.051$, $p < .0005$. The significant differences are highlighted in Figure 16. Experimental Group 2 firstly had a significant difference result ($\chi 2(9) = 23.768$, $p = .005$), but after the adjustment and analyzed with the pairwise comparison considered no significant. Control Group again, had no significant differences, $\chi 2(9) = 8.348$, $p = .499$.

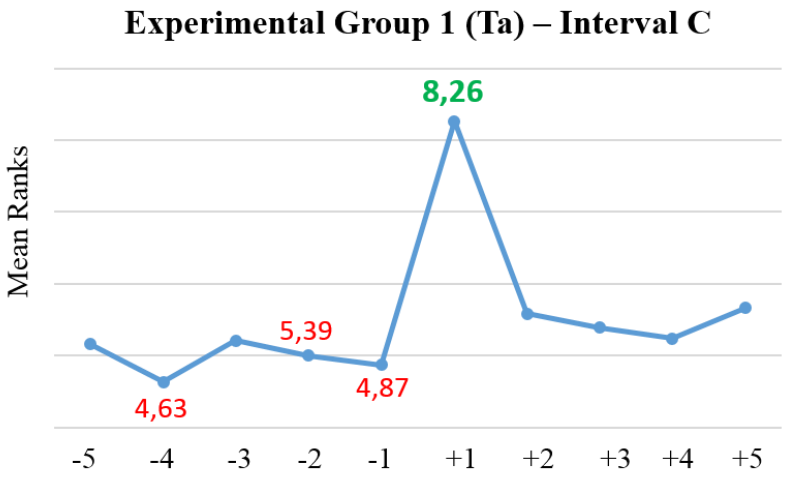

Figure 16. Total annotations rate in interval C – Exp. Group 1
Source: Author's elaboration

For the Rp rate, the test showed that the differences were not statistically significant for the gamification and Control Group, $\chi2(9) = 16.296$, $p = .061$ and $\chi2(9) = 9.304$, $p = .410$, respectively. For the Experimental Group 2, initially there was a significant result in terms of mean differences ($\chi2(9) = 17.525$, $p = .041$), but after the Bonferroni correction ($p < .0005$ level) no significant difference was shown. Experimental Group 1 was the only one showing an initial significant difference in Bd ($\chi2(9) = 32.737$, $p < .0005$), though only before Bonferroni correction and pairwise analysis. Traditional and Control Group had no statistically significant differences in Bd rate mean ranks between any of the ten-month analyzed, $\chi2(9) = 16.526$, $p = .057$ and $\chi2(9) = 8.653$, $p = .470$, respectively.

The Friedman tests demonstrated, for the Ge rate, no statistically significant differences for Experimental Group 1 and Control Group, $\chi2(9) = 15.123$, $p = .088$ and $\chi2(9) = 14.422$, $p = .108$, respectively. Even though Experimental Group 2 had an initial significant difference between the measures in Ge for this interval, later no significant result was found, $\chi2(9) = 18.802$, $p = .027$.

Mc showed only initial significant result for Experimental Group 1 ($\chi2(9) = 38.277$, $p < .0005$) with no significance after correction, and no difference statistically significant for traditional and control, $\chi2(9) = 16.047$, $p = .066$ and $\chi2(9) = 6.368$, $p = .703$, respectively.

(4) Interval D – fourteen-month period

A Friedman test was run to determine if there were differences in the Ta rate throughout a fourteen-month period, seven months before the training events and seven after, for the three

groups of this study. The differences were not statistically significant for the Control Group, $\chi2(13) = 20.438$, $p = .085$. For the Experimental Group 2, the Friedman test showed significant difference $p = .007$. Although, the pairwise analysis revealed no statistically significant differences with a Bonferroni correction for multiple comparisons ($\chi2(13) = 28.702$, $p = .007$) when statistical significance was accepted at the $p < .0005$ level.

Pairwise comparisons were implemented with a Bonferroni correction for multiple comparisons for Experimental Group 1. Results obtained after the correction show yet statistically significantly differences $\chi2(13) = 47.412$, $p < .001$. Figure 17 highlight the periods with statistical difference identified by this test.

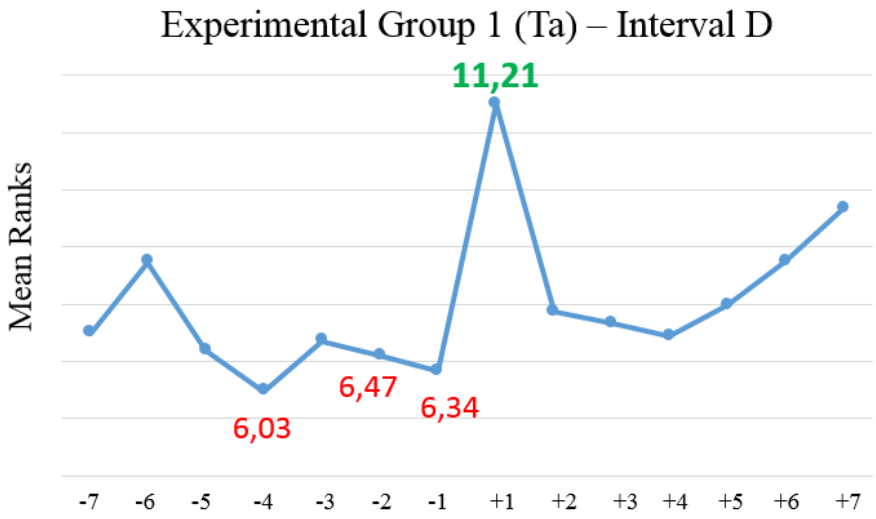

Figure 17. Total annotations rate in interval D – Exp. Group 1
Source: Author's elaboration

The Friedman test for Rp rate showed that the differences were not statistically significant for all three groups: the Experimental Group 1, $\chi2(13) = 22.241$, $p = .052$, Experimental Group 2, $\chi2(13) = 20.487$, $p = .084$ and Control Group, $\chi2(13) = 16.814$, $p = .208$.

The Bd rate for the fourteen-month period demonstrated there were not statistically significant for the traditional $\chi2(13) = 19.636$, $p = .105$ and Control Group, $\chi2(13) = 22.155$, $p = .053$. For the Experimental Group 1, the Friedman test showed a significant difference, $\chi2(13) = 42.269$, $p < .0005$. However, the pairwise analysis revealed no statistically significant differences with a Bonferroni correction for multiple comparisons, when statistical significance was accepted at the $p < .0005$ level.

The Friedman test ran to determine differences in Ge rate showed not statistically significant for all the groups: Experimental Group 1 $\chi 2(13) = 21.525$, $p = .063$; Experimental Group 2, $\chi 2(13) = 21.463$, $p = .064$ and Control Group, $\chi 2(13) = 16.996$, $p = .199$.

Finally, referring to Mc rate the differences were not statistically significant for the Experimental Group 2, $\chi 2(13) = 19.500$, $p = .108$ and Control Group, $\chi 2(13) = 12.028$, $p = .525$. For the Experimental Group 1, after the Bonferroni correction for multiple comparisons ($p < .0005$ level), there was no statistically significant differences, although firstly there was a significant result, $\chi 2(13) = 46.509$, $p < .0005$.

The two tables below summarize the findings for this phase of study 6, which sought to compare the results within groups (see Table 20 and Table 21).

Table 20
Friedman and Signed tests synthesis*

Group / Rate / Variable			Interval			
			A	B	C	D*
Gamification	Ta	Test Statistic	47.412	40.051	32.684	13.000
		Asymp. Sig.	< .001	< .001	< .001	.002
	Rp	Test Statistic	22.241	16.296	13.257	4.000
		Asymp. Sig.	.052	.061	.021	.375
	Bd	Test Statistic	42.269	32.737	26.246	11.000
		Asymp. Sig.	< .001	< .001	< .001	.006
	Ge	Test Statistic	21.525	15.123	12.414	5.000
		Asymp. Sig.	.063	.088	.030	.062
	Mc	Test Statistic	46.509	38.277	26.224	9.000
		Asymp. Sig.	< .001	< .001	< .001	.004
Traditional	Ta	Test Statistic	28.702	23.768	12.609	10.000
		Asymp. Sig.	.007	.005	.027	.012
	Rp	Test Statistic	20.487	17.525	10.154	5.000
		Asymp. Sig.	.084	.041	.071	.062
	Bd	Test Statistic	19.636	16.526	12.044	7.000
		Asymp. Sig.	.105	.057	.034	.070
	Ge	Test Statistic	21.463	18.802	12.868	5.000

Group / Rate / Variable			Interval			
			A	B	C	D*
		Asymp. Sig.	.064	.027	.025	.219
	Mc	Test Statistic	19.500	16.047	10.950	4.000
		Asymp. Sig.	.108	.066	.052	.125
	Ta	Test Statistic	20.438	8.348	5.000	5.000
		Asymp. Sig.	.085	.499	.416	.727
	Rp	Test Statistic	16.814	9.304	6.854	3.000
		Asymp. Sig.	.208	.410	.232	.250
Control	Bd	Test Statistic	22.155	8.653	4.219	3.000
		Asymp. Sig.	.053	.470	.518	1.000
Control	Ge	Test Statistic	16.996	14.422	8.200	3.000
		Asymp. Sig.	.199	.108	.146	.625
	Mc	Test Statistic	12.028	6.368	4.000	0.000
		Asymp. Sig.	.525	.703	.549	1.000

Source: Author's elaboration

Table 21

Summary of within groups' hypotheses

Interval A		**Interval B**		**Interval C**		**Interval D**	
H	**Result**	**H**	**Result**	**H**	**Result**	**H**	**Result**
H_1	Accepted	H_{16}	Accepted	H_{31}	Accepted	H_{46}	Accepted
H_2	Accepted	H_{17}	Rejected	H_{32}	Rejected	H_{47}	Rejected
H_3	Accepted	H_{18}	Accepted	H_{33}	Accepted	H_{48}	Accepted
H_4	Rejected	H_{19}	Rejected	H_{34}	Rejected	H_{49}	Rejected
H_5	Rejected	H_{20}	Rejected	H_{35}	Rejected	H_{50}	Rejected
H_6	Accepted	H_{21}	Accepted	H_{36}	Accepted	H_{51}	Accepted
H_7	Accepted	H_{22}	Rejected	H_{37}	Rejected	H_{52}	Rejected
H_8	Rejected	H_{23}	Rejected	H_{38}	Rejected	H_{53}	Rejected
H_9	Accepted	H_{24}	Accepted	H_{39}	Accepted	H_{54}	Accepted
H_{10}	Rejected	H_{25}	Rejected	H_{40}	Rejected	H_{55}	Rejected
H_{11}	Rejected	H_{26}	Rejected	H_{41}	Rejected	H_{56}	Rejected

Interval A		Interval B		Interval C		Interval D	
H	**Result**	**H**	**Result**	**H**	**Result**	**H**	**Result**
H_{12}	Accepted	H_{27}	Accepted	H_{42}	Accepted	H_{57}	Accepted
H_{13}	Accepted	H_{28}	Rejected	H_{43}	Rejected	H_{58}	Rejected
H_{14}	Rejected	H_{29}	Rejected	H_{44}	Rejected	H_{59}	Rejected
H_{15}	Accepted	H_{30}	Accepted	H_{45}	Accepted	H_{60}	Accepted

Source: Author's elaboration

6.2.2 Between groups' findings

The main purpose of this phase is to determine if the median of the rates created (Ta, Rp, Bd, Ge and Mc) of at least one group is different from the median of another group (Experimental Group 1, Experimental Group 2 and Control Group) in the fourteen months measured in this study. The Kruskal-Wallis H test was run to determine if there were differences between the three groups. Distributions of the rates were not similar, as assessed by visual inspection of a boxplot. Results made will be presented rate by rate, showing the most pronounced differences found, which will be worked out in the discussion section of this research.

No statistical result was found for Ta when comparing the groups in all the months analyzed. Three aspects deserve attention: (1) the Control Group had a higher rate at the PA cycle ends before the intervention; (2) both Experimental Group 1 and Experimental Group 2 surpassed the Control Group in the first month after the intervention; and (3) in the second month after the intervention, all the groups already had similar scores in this rate. Figure 18 highlights the Mean Ranks obtained in these periods.

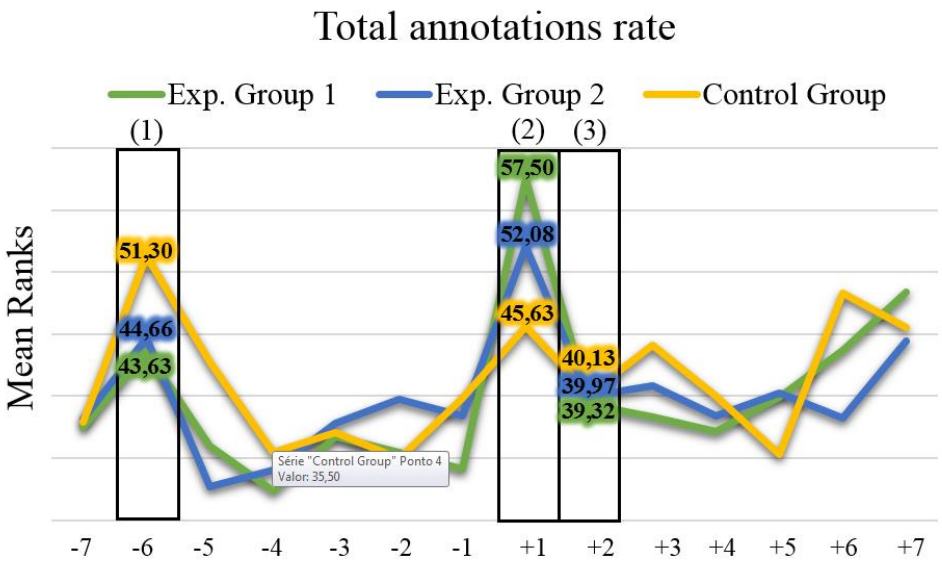

Figure 18. Total annotations rate – Between groups' comparison
Source: Author's elaboration

Likewise, there were no statistically significant differences in the Rp rate in all periods' comparison. Knowing that this rate measures participants' unwanted behavior, two aspects are of interest in this study to be discussed in the next section: (1) in the first month after the intervention, all three groups have approximate mean rank values; and (2) at the end of the PA cycle, the Control Group (only one not trained) is the one with the highest mean ranks for the Rp rate. See Figure 19 to observe the differences pointed.

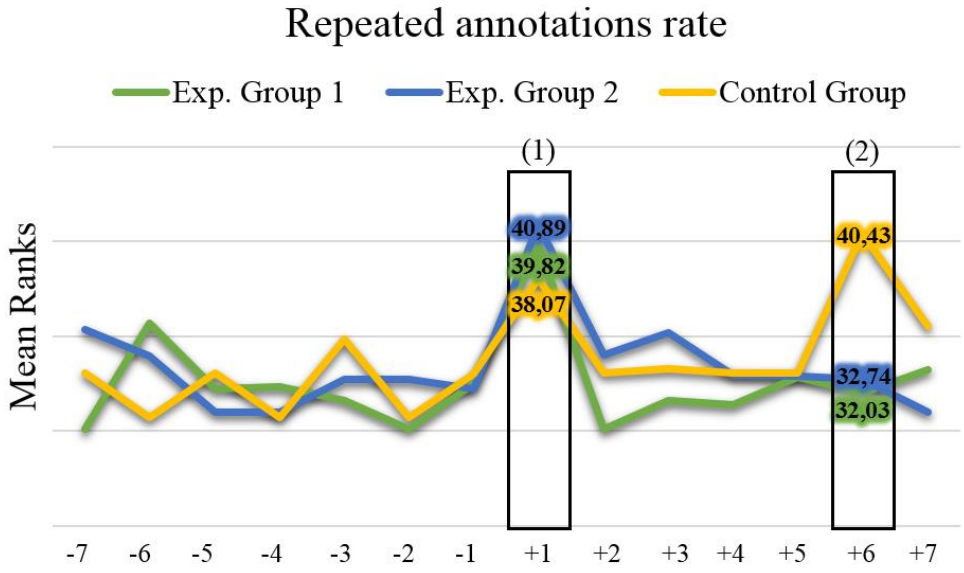

Figure 19. Repeated annotations rate – Between groups' comparison
Source: Author's elaboration

The Kruskal-Wallis H test found statistically significant differences for Bd in the first month after the quasi-experiment, $\chi 2(2) = 8.655$, $p = .013$. It was necessary to run a Post Hoc test to determine which groups were the difference. Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Statistical significance was accepted at the $p < .016$ level. This post hoc analysis revealed statistically significant differences in the behavior description rates (Bd) between the gamified group (mean rank = 33.66) and the Control Group (mean rank = 19.73) (p= .004), but not between the Experimental Group 2 (mean rank = 26.08) or any other group combination. Figure 20 graphically represents this finding.

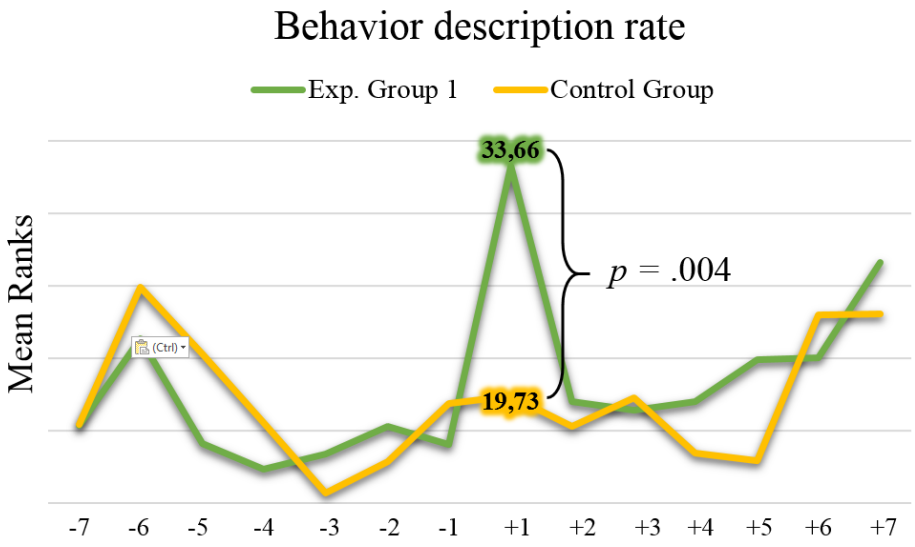

Figure 20. Bd – Significant differences (Exp. Group 1 and Control Group)
Source: Author's elaboration

No statistical result was found for Ge when comparing the groups in all the months analyzed. One aspect should be considered the Experimental Group 1 was the one that got the lowest mean ranks at this rate after the intervention. This rate is associated with one type of annotation: improvement. Recognition annotations do not require the participant to complete this field. Figure 21 shows the Ge Mean Ranks for the three groups in the fourteen-month period.

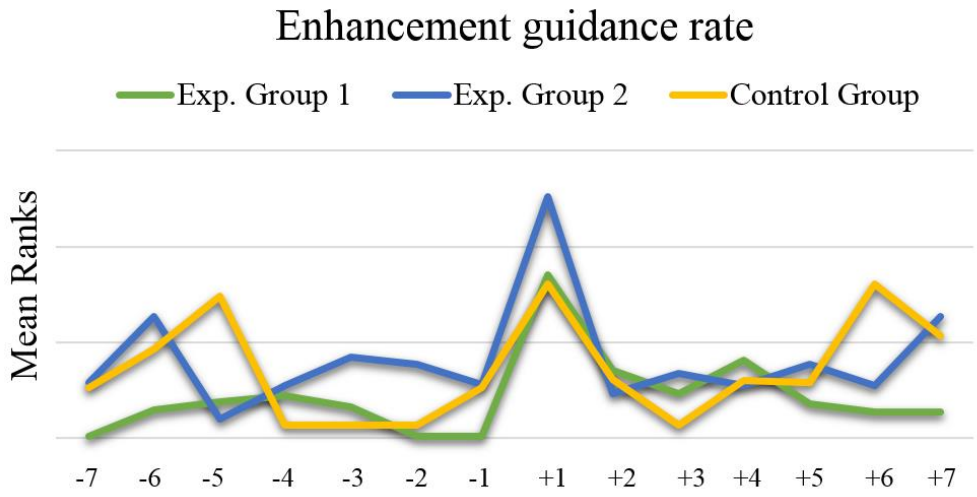

Figure 21. Ge – Between groups' comparison
Source: Author's elaboration

Lastly, Kruskal-Wallis H test found statistically significant differences for Mc in the two periods investigated: first and fourth months after the quasi-experiment, $\chi 2(2) = 9.696$, $p = .008$ and $\chi 2(2) = 7.582$, $p = .023$, respectively. Again, it was necessary to run a Post Hoc test to determine between which groups were the difference. Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Statistical significance was accepted once more at the $p < .016$ level.

The post hoc analysis for the first month after the intervention revealed statistically significant differences in the management commitment rate (Mc) between the gamified group (mean rank = 32.95) and the Control Group (mean rank = 20.50) ($p= .002$), but not between the Experimental Group 2 (mean rank = 26.18) or any other group combination. The post hoc analysis for the fourth month after the intervention revealed statistically significant differences in Mc between the gamified group (mean rank = 30.58) and the Experimental Group 2 (mean rank = 25.00) ($p= .015$), though not between the Control Group (mean rank = 25.00) or any other group combination. Figures 22 and 23 graphically represents these findings.

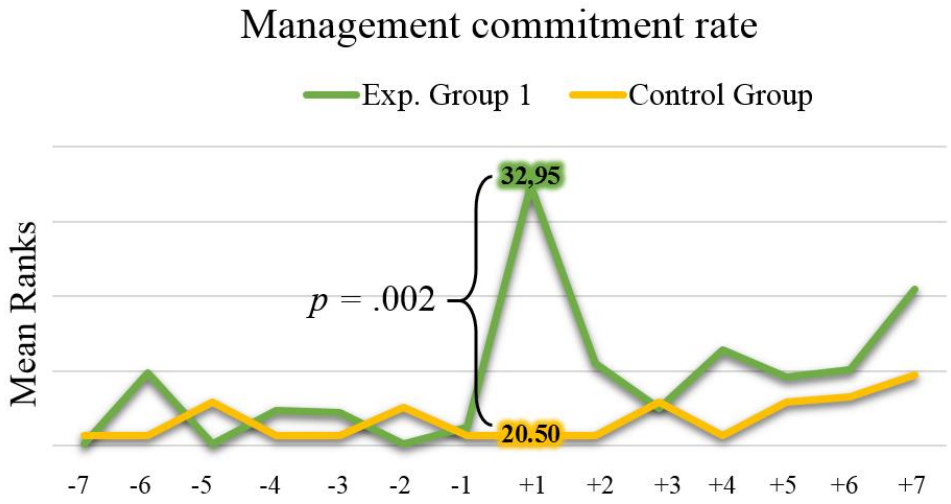

Figure 22. Mc – Significant differences (Exp. Group 1 and Control Group)
Source: Author's elaboration

Figure 23. Mc – Significant differences (Exp. Group 1 and Exp. Group 2)
Source: Author's elaboration

A findings synthesis can be seen in Table 22. The hypotheses acception or rejection can be summarized in: (1) There were no differences between groups in all comparison in the periods before the intervention – hypotheses from H_{61} to H_{65} are accepted; (2) only three significant divergences were found in all comparisons after the intervention - H_{97}, H_{115} and H_{129} are accepted; all the other hypotheses for between groups comparison after intervention are rejected.

Table 22
Kruskal-Wallis H tests synthesis

Variable		Periods													
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Ta	Kruskal-Wallis H	0.012	0.718	4.689	0.650	0.013	1.297	1.443	3.831	0.048	0.902	0.179	1.267	2.717	1.180
	Asymm. Sig.	.994	.698	.096	.723	.993	.523	.486	.147	.976	.637	.915	.531	.257	.554
Rp	Kruskal-Wallis H	1.971	2.433	1.171	1.789	0.900	1.789	0.051	0.191	1.921	0.400	0.051	0.363	5.118	2.460
	Asymm. Sig.	.373	.296	.557	.409	.638	.409	.975	.909	.383	.819	.975	.834	.077	.292
Bd	Kruskal-Wallis H	0.185	0.318	3.229	0.984	1.673	0.883	0.705	8.655	0.646	0.049	1.743	2.676	3.145	2.787
	Asymm. Sig.	.912	.853	.199	.612	.433	.643	.703	.013	.724	.976	.418	.262	.208	.248
Ge	Kruskal-Wallis H	1.171	1.164	4.694	0.806	1.713	3.648	1.171	0.537	0.381	0.806	1.312	0.430	2.938	1.298
	Asymm. Sig.	.557	.559	.096	.668	.425	.161	.557	.764	.827	.668	.519	.807	.230	.522
Mc	Kruskal-Wallis H	1.789	1.713	2.533	0.806	0.806	2.028	0.806	9.696	3.264	0.023	7.582	0.661	0.703	2.437
	Asymm. Sig.	.409	.425	.282	.668	.668	.363	.668	.008	.196	.989	.023	.719	.703	.296

Source: Author's elaboration

7 Discussion

Initially, it is important to mention that this research met all the objectives proposed and tested all the hypotheses. This section will begin discussing the main qualitative results and after discuss the findings of the quasi-experiments.

The first point to be discussed in this section is study 1, which was the intervention's creation. The gamification literature is not sufficiently clear when defines gamification as only the use of game elements in non-game contexts, to solve real life problems. This definition is insufficient to allow the design of a compelling experience. In this research, gamification was allied to instructional theories fundaments. It is understood that this conjunction was what allowed the achievement of the results obtained. Here, game elements were focused primarily on competition rather than cooperation between participants. It was a risk to make that choice because it is known that competition can hinder learning (Hanus & Fox, 2015). In this research, the risk was taken because it was in a financial institution context, where managers are used to be measured by their daily performances. The method applied for evaluation and reward in competition may have mitigated the possible adverse effects. Assessments in the event were carried out by the participants themselves with the use of an interactive digital feedback system. They assessed their performances in practical activities by alleviating the burden of an assessment by an instructor or a judge (McGonigal, 2009). Instead, traditional instructional design, even with the use of practical activities, may have been inadequate to the daily work pace of these participants (complex and challenging tasks), becoming repetitive and tiring, although achieving the same instructional objectives (Prince, 2004).

Despite the theory on gamification does not provide a basis for how to use each game element aligned with each purpose and behavior one wishes to work on, the theoretical basis provided by the studies on training and development can cover this lack of theoretical basis in gamification studies (Abbad et al. al., 2006b). On the one hand, T&D area has over 100 years of research expertise and sufficient empirical evidence to demonstrate the effectiveness of methods in certain situations and inefficiency in others (Bell et al., 2017). On the other hand, gamification has a weak and recent theoretical basis – in the search made with the inclusion and exclusion criteria, the oldest article dated the year of 2013 (Dominguez et al., 2013).

Regarding situational performance tests developments and the emergence of the assessment criteria, this task was only possible due to the analysis of the pilot group, an indication that organizations may not be able to predict quality criteria during the instructional design stage. Develop detailed criteria is an exhaustive and laborious task. It is essential to

discuss the importance of the pilot class not only to test instructional design but to evaluate with higher levels of learning with open tests, identifying criteria to redesign entirely then the instruction based in the underlying criteria that were not previously visible.

Study 4 measured generalization, and of the five hypotheses formulated for this study, only one was rejected. Although gamification studies demonstrate some contradiction in learning performance results, most point positively to the use of gamification, despite no study measured generalization with situational performance tests. Even though, the result obtained in this research was not surprising. Gamification applied in this training was intentionally competitive, requiring long instruction periods before the activity beginning, and also a long time to compute the scores obtained by each participant and by each group on the gamified activities. This extra time that Experimental Group 2 had in class was used to expand group discussions and theoretical and dialogued expositions on the content and practical activities evaluation. The similar results achieved for both groups indicate this extra time spent discussing concepts did not result in better generalization for the Experimental Group 2. One reason may lie in the cognitive load construct. Gamification may work as a relieve tool when participants are overloaded with repeated information for more extended periods (Su & Chen, 2015).

Another relevant result obtained in study 4 refers to the comparative analyses between pretest and posttest scores for the three groups. All hypotheses were accepted, showing that the trained groups obtained improvement in learning, while the group that did not undergo training did not improve – confirmed by the statistical analysis. This outcome is essential to confirm the quality of the training events applied, but also showed the equivalence between the instruments applied to measure generalization. This research was unconventional when applying open-format situational performance tests to measure high cognitive levels such as synthesis and evaluation. Another significant result that corroborates the internal validity of the quasi-experiment was the measure of the initial repertoire levels of the participants of the three groups. In the pretest scores, no significant differences were found. This fact is relevant since the individuals were not randomly assigned in each experimental group.

At this point it is important to mention how the threats to the internal validity of this research were treated: (1) ambiguous temporal precedence – eliminated with pretest and posttest implementation; (2) selection – pretest scores did not differ for the three groups; (3) history – presence of a Control Group (valid only for the quasi-experiment II with 14 month periods measured); (4) maturation – small time lag for study 4 and presence of the Control Group in study 5 (had a 14-month time series measures); (5) regression – scores were not extreme; (6) attrition – no loss of subjects; (7) testing – different but equivalent situational tests; (8)

instrumentation – there was no change in the nature of the measure. Additionally, the same instructor could not apply 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.

It is essential to discuss the relevance of the use of instruments that measure learning in organizations – especially those situated in more complex levels of the cognitive domain, as was the case of this research. Developing, applying and assessing learning with open answers items is an extra effort for everyone involved in the T&D solutions creation, though the benefits outweigh this effort. Evaluating transfer of training with criteria similar to those used in the generalization learning assessment was only possible after an intense analytical process done with the pilot group participants pretest and posttests answers. Further, the pilot group analysis should not only be intended to test the instructional design but primarily to allow situational performance tests to fundament decision making, creating additional criteria expected for the participants' performances, not possible to be identified at the instructional design stage. Organizations may be failing to perform a needs analysis and instructional design based on instructional goals constructed without situational performance tests and generalization learning assessments.

Transfer of training performance was assessed through rates created to measure gamification effects as an instructional strategy, comparing it to an instructional design with a traditional approach and an untrained Control Group. The analyses were executed with measures of comparison between groups and within groups in a total interval of 14 months (7 before and 7 after the intervention).

The first moment of statistical analysis on the transfer of training was the one that observed a comparison within groups. Sectional intervals were made in the total period of 14 months, to verify significant differences within each group in the five rates. One rate produced statistically significant results, once more for the Experimental Group 1. These results were repeated in all the intervals, with significant differences between months before the intervention and the first month after it. The rate that showed this difference was the only one that did not deal with quality criteria but measured the total number of annotations per team member. Notwithstanding the statistically significant results found, there was a steady decay of this behavior soon in the first month after the intervention – the same happened to the other rates measuring the annotation’s quality.

Regarding between groups analysis and observing statistically significant results, two rates demonstrated that gamification produced a higher transfer of training performance in the individuals who participated in the Experimental Group 1, with gamification as an instructional strategy. The rate that measured the management commitment produced a significant difference comparing Experimental Group 1 with both Experimental Group 2 and Control Group. Considering that this rate reflected criteria worked explicitly in the training events, it was not expected that the Control Group would demonstrate this behavior – which was confirmed by the data analysis. Instead, the Experimental Group 2 had the same experience Experimental Group 1 had in classroom. Thus, Experimental Group 1 presented an effect on individuals so that they, even having contact with the same criteria in training, applied it more in the work environment, with statistically significant results, when compared to Experimental Group 2 participants.

There is a second rate that indicated a significant positive difference for the Experimental Group 1 - the one that dealt with behavior description quality. This rate originated from quality criteria in managers' annotations regarding the behavior descriptions in their workplace. It is important to note that the two rates that obtained significant difference favorable to the Experimental Group 1 in comparison with the other groups exhibit the difference in the first month after the training intervention. Only one of the comparisons showed a difference in another period, positioned in the fourth month after the training. The results on this two quality rates were timid (represented only three hypotheses accepted), in fact not representative compared to all formulated hypotheses that were rejected.

Yet regarding the comparisons between groups, repeated annotations rate revealed an interesting result, despite not being statistically significant. It was possible to observe that even though in the first month after the intervention the three groups had similar rate scores, a few months after the training (more precisely at the end of the evaluation cycle), the groups that experienced the training had a decrease in the use of repeated annotations (behavior considered inadequate). The untrained Control Group maintained the same levels in the use of this type of annotation, showing that the difference may be an effect of the training emphasis on the inadequacy of this type of annotation – which tends to massify and treat individuals collectively, confronting the foundations of the PA model used by the FI.

To conclude this discussion section, it is relevant to mention the occurrence of peaks reflecting behavioral change in the evaluated managers. They manifested almost exclusively in the end-of-cycle periods (the intervention was performed near the end of one of the evaluated cycles). Therefore, if transfer of training should be manifested as lasting behavioral change, in

fact this did not occur in this research. There were episodic changes in the behavior of the managers, which were demonstrated in the evaluation cycles final periods. Possible causes of these results can be related to factors that extrapolate the training and its effects. The volume of daily tasks of these managers can induce them to perform the evaluation of their team members only at the closing moments of the cycle. Another factor may be the belief that these managers have in this methodology of performance appraisal adopted by FI to develop KSAs in individuals. Regarding the training, its effectiveness and its possible consequences, it can be noticed that an absence of stimuli for the participants to do annotations during the whole cycle may have been the key to the fact that transfer did not occur as expected. To sum up, a training of only 16 hours duration should be supported by complementary actions, stimulating the demonstration of KSAs developed in the workplace (Ford, Baldwin, & Prasad, 2018).

Next session presents the contributions of this research to the scientific field as well as to organizations.

8 Final Words

This section presents the final thoughts about this research. Initially, it is pointed this work's contributions of this each theme and study addressed: gamification, learning assessment in organizations. In a second moment, the limitations of this research are brought. Finally, there is a proposal of future research gamification as a strategy for organizations, seeking to promote behavior change through learning.

8.1 Contributions

A first contribution is theoretical, regarding the gamification and game elements regarding systematization. The revised literature does not define nor adequately organize the game elements, do not indicate where, when and how they should be adopted to promote behavioral change. Some of the manifested elements were indeed closer to principles, design fundaments. This work used the MDE model and sought to fit the game elements in the given concepts, especially in the description of each gamification mechanic.

The second contribution of this work is related to the instructional design area. There was an attempt to align the entire instructional design with the transfer of training. Many instructional design methods still insist on working with intermediate instructional objectives, without concerning with their connection to tasks performed later in the workplace. The

instructional design built for this quasi-experiment used PA theoretical foundations though was mainly focused on the practical activities that sought to simulate tasks executed at work. Organizations should seek out models and methods to design their training solutions, with instructional objectives strictly aligned with the transfer of training.

A third contribution relates to the instructional design applying gamification as an instructional strategy, using fundaments of instructional design theories and approaches. Most of the researches on gamification had not enough theoretical basis to justify the elements' choices and the design's construction. What was desired here was to use gamification in a way that is in line with the instructional objectives defined for the training, articulating gamification complementing the other created or selected strategies to facilitate learning.

Another significant contribution of this work refers to the construction of situational performance tests, also previously and intentionally aligned with the transfer of training. In the same way that the activities built in the instructional design had objectives aligned to the expected performances in work, the situational tests were also created pursuing to simulate and have proximity to the situations that the participants encounter when in their functions in the financial institution context of this study.

A next contribution was the construction of the criteria for generalization-level learning assessment, measured in the situational performance tests. These criteria construction was made based on the tests analysis of the tests conducted with the pilot group data. These criteria were fundamental for the learning measures at the generalization level and later for the transfer measures. Perhaps the most significant contribution of this work is to empirically demonstrate the importance of constructing criteria representing expected quality standards prior to the provision of any training solution for an organization's employees. Most of the criteria identified were not predictable for instructional designers before testing. This work contributes not only to the field of learning assessment but also to the area of training needs assessment, which could also benefit from the criteria of situational tests. Workshops to identify performance gaps and develop training, even with content experts and training target members, are probably unable to identify and categorize all performance standards that derive from the analytical work of responses given in situational performance tests. This recommendation can be valuable to organizations contributing to the effectiveness of their training solutions.

Another contribution of this work refers to the assessment of transfer of training with the same criteria created to assess generalization, something only possible because there was an alignment that began in the instructional objectives construction. The alignment continued through the situational performance tests development, to finally arrive at the criteria

formulation, always aligned with the transfer of training, pursuing to reflect trainees real performances in the workplace. Only in this way was it possible to arrive at this contribution of being able to assess transfer of training with participants' real data reflecting performances in their functions, based on the same criteria that were trained and assessed through situational performance tests.

Finally, this research is innovative using gamification: within a financial institution; investigating behavior of managers; performing a training intervention with experienced professionals in management, to measure learning outcomes with situational performance tests with open answers; to assess transfer of training – application of KSAs in the workplace.

8.2 Limitations

Among the research limitations of this study is the choice of the game elements adopted to implement gamification. As it was applied, it was not possible to measure the effectiveness of one element compared to another. Gamification encompassed the entire learning environment from start to finish of the event. Elements of inherently competitive games were used. Some principles were also 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 this research is the reconstruction of the instructional design after the pilot group. Due to the short time interval between the completion of the intervention classes and the pilot group, it was not possible to perform the situational performance tests between these moments, using them to improve the practical activities seeking greater effectiveness of the training. This research could not randomly assign subjects to each group, but this gap was circumvented by statistical tests that did not result in differences between groups in pretest scores. Also this research did not test the instruments by means of factorial analyzes.

A research gap was the non-identification of some profile variables of participants – e.g., learning preferences. The last limitation identified regards the non-identification of interfering variables (mediators or moderators) that could influence learning performance – e.g., levels of motivation and engagement of the participants and their previous experiences with gamification.

8.3 Future Research

Many challenges could not be contemplated in this research and continue as possibilities for future studies in the area of gamification and learning in organizations.

Concerning theoretical aspects, gamification requires more robust theoretical models, in the form of taxonomies, organizing the game elements and their use conforming to the level or type of behavior expected of the participant after the experience. Models found in the reviewed literature do not contemplate how each element can be favorable when used in a particular situation, participant profile, environment, or type of KSAs that designers wish to develop on the participants. Prescriptive models have been applied effectively in the instructional design for over 50 years, facilitating designers' work in creating and developing instructional solutions within organizations.

Another recommendation for future research would be to approach other KSAs with characteristics different from those used here. An example would be KSAs that are reflected in harsh business indicators such as numbers that show a direct impact on the business conducted by sellers in the most diverse types of organizations. These indicators, although similar to those adopted in this research (regarding the demonstration of workplace behaviors) differ because of the complexity or the particularity of being assessed after training. An example of KSA at another extreme would be the use of communication area themes. Usually, these areas have no way of measuring behavior and its effectiveness. Therefore, it would be challenging to work on these topics where the results of learning and their application in work are not as tangible as those worked on this research.

Concerning designing gamification, it was noticed that the literature agrees to affirm that some designs are neglected in empirical researches. It is worth here to reinforce this aspect indicating that there is an effort on the part of the researchers to investigate the effects of these elements in learning environments, inside and outside the organizations. One can cite among the elements or principles of gamification with an empirical gap the use of environments that promote autonomy and freedom of choice. Also, cooperative environments, where the participants promote their self-management, cooperating and building knowledge using gamification as support for the collaborative process. There is also a gap in the investigation of environments that personalize the learning experience. Gamification, in this sense, can collect behavioral data, suggesting personalized recommendations and paths for the participants, as already exist in many current video games, where there is no single end, though several endings depending on the player's choices.

Specifically, regarding training transfer, the literature indicates that in order to avoid decaying the positive effects of training, empirically proven in this research, organizations need to use post-training interventions so that participants continue to use KSAs seized in the environment of training and maintain the quality standard presented in the first moments after training. These interventions may be supported by technology. Surveys on gamification may also be combined with post-training interventions applying hybrid learning environments, for example, a classroom course with post-training interventions performed remotely with the use of gamification on online platforms.

Finally, as the last recommendation, it is understood that gamification should assess learning in organizations using complementary instruments to document data collections or self-report instruments. An example of an assessment that already occurs in some organizations and can be used for scientific purposes refers to the evaluation of external clients on the performance of participants in training events. After all, these clients experience and can help to measure (having appropriate instruments) if the performances demonstrated by the employees are within the proper and expected by the organization and by the clients themselves.

Concluding, gamification applied in learning research within organizations is a subject still little explored, but that should continue to be investigated. After all, many organizations continue to look for innovative ways to improve individual and team's performances. Gamification has proved to be effective, but to have its use expanded within organizations, it needs support from scientific methods and results. So, it is crucial the participation of researchers and science to provide a theoretical and empirical base, helping companies that sell and buy gamification to design, apply and assess solutions that promote behavior change, bringing the expected positive outcomes for organizations.

References

Abbad, G. S. (1999) *Um modelo integrado de avaliação do impacto do treinamento no trabalho – IMPACT*. Tese de doutorado não publicada. Universidade de Brasília.

Abbad, G. S., & Borges-Andrade, J. E. (2006). Aprendizagem humana em organizações de trabalho. In J. C. Zanelli, J. E. Borges-Andrade, & A. V. B. Bastos (Org.), *Psicologia, organizações e trabalho no Brasil* (pp. 237-275). Porto Alegre: Artmed.

Abbad, G. S., Borges-Ferreira, M. F. & Nogueira, R. (2006a) Medidas de aprendizagem em avaliação de TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 469-488). Porto Alegre: Artmed.

Abbad, G. S., Nogueira, R. & Walter A. M (2006b). Abordagens instrucionais em planejamento de TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 255-281). Porto Alegre: Artmed.

Abbad, G. S., Zerbini, T., Carvalho, R. S. & Meneses, P. P. M. (2006c) Planejamento Instrucional em TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 289-321). Porto Alegre: Artmed.

Adukaite, A., van Zyl, I., Er, Ş., & Cantoni, L. (2017). Teacher perceptions on the use of digital gamified learning in tourism education: The case of South African secondary schools. *Computers & Education*, 111, 172–190.

Alcivar, I., & Abad, A. G. (2016). Design and evaluation of a gamified system for ERP training. *Computers in Human Behavior*, 58, 109–118.

Alves, F. (2015) *Gamification: como criar experiências de aprendizagem engajadoras. Um guia completo: do conceito à prática*. DVS Editora.

Ambrosio, D. M., & Garofalo, P. F. (2016). Expect the Unexpected: Simulation Games as a Teaching Strategy. *Clinical Simulation in Nursing*, 12(4), 132–136.

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.

Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011). Towards a science of user engagement. *WSDM Workshop on User Modelling for Web Applications*, Hong Kong, China.

Auvinen, T., Hakulinen, L., & Malmi, L. (2015). Increasing Students' Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges. *IEEE Transactions on Learning Technologies*, 8(3), 261–273.

Baldwin , T., Ford , J. K., & Blume , B. (2009). Transfer of training 1988–2008: An updated review and new agenda for future research . In G. P. Hodgkinson and J. K. Ford (Eds.), *International review of industrial and organizational psychology*, 24, 41-70. Chichester, England : Wiley.

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

Baldwin, T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41, 63-105.

Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs: Prentice-Hall.

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.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612-637.

Beed, P. L., Hawkins, E. M., & Roller, C. M. (1991). Moving learners toward independence: the power of scaffolded instruction. *The Reading Teacher*, 44, 648-655.

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, 305–23.

Bell, B., & Kozlowski, S. (2008). Active Learning: Effects of Core Training Design Elements on Self-Regulatory Processes, Learning, and Adaptability. *Journal of Applied Psychology*, 93 (2), 296–316.

Bharathi, G. Ajay, S., Abhinav, T., Conrad, N. (2016). Knowledge discovery of game design features by mining user-generated feedback. *Computers in Human Behavior.* 60. 361-371.

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

Blume B., Ford J. K., Baldwin T., & Huang J. (2010). Transfer of training: a meta-analytic review. *Journal of Management.* 36, 1065-1105.

Blume B., Ford J. K., Surface E., & Olenick J. (2017). A dynamic model of training transfer. Apresentado em: *Annual Conference Social Industrial Organizational Psychology*, 32nd, Abril, Orlando, FL.

Borges-Andrade, J. (2012) Avaliação Integrada e Somativa em TD&E In G. S. Abbad, L. Mourão, P. P. M. Meneses; T. Zerbini, J. E. Borges-Andrade, & R. Vilas-Boas (Org.), *Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: Ferramentas para gestão de pessoas.* (pp. 289-321). Porto Alegre: Artmed.

Borges-Andrade, J. E., Abbad, G. S., & Mourão, L. (2012) Modelos de avaliação e aplicação em TD&E. In G. S. Abbad, L. Mourão, P. P. M. Meneses; T. Zerbini, J. E. Borges-Andrade, & R. Vilas-Boas (Org.), *Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: Ferramentas para gestão de pessoas.* (pp. 20-35). Porto Alegre: Artmed.

Brown P., Roediger H., & McDaniel M. (2014). *Make It Stick.* Cambridge, MA: Harvard University Press.

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.

Burke L. (2001). Training transfer: ensuring training gets used on the job. In *High Impact Training Solutions*, ed. LA Burke, pp. 89–116. Westport, CT: Quorum Books.

Cain, J., & Piascik, P. (2015). Are Serious Games a Good Strategy for Pharmacy Education? *American Journal Of Pharmaceutical Education*, 79(4), 47.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction.* Princeton, NJ: Princeton University Press.

Cheng, E., & Hampson I. (2008). Transfer of training: A review and new insights. *International Journal of Management Reviews.* 10 (4), 327–341.

Cheong, C., Flilippou, J., & Cheong, F. (2014). Towards the Gamification of Learning : Investigating Student Perceptions of Game Elements. *Journal of Information Systems Education*, 25(3), 233–245.

Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, 78, 66–77.

Clark, M. C., & Rossiter, M. (2008). Narrative learning in adulthood. *New Directions for Adult and Continuing Education*, 119, 61-70.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678–707.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Creswel, J. W. (2010) *Projeto de Pesquisa: métodos qualitativo, quantitativo e misto.* Porto Alegre: Artmed.

Davis, K., & Singh, S. (2015). Digital badges in afterschool learning: Documenting the perspectives and experiences of students and educators. *Computers & Education*, 88, 72–83.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research.* 71(1), 1-27.

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. (2012). Gamification: Designing for Motivation. Interactions, 19(4), 14.

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).

Dichev, C., & Dicheva, D. (2017). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education*, 14.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society*, 18(3), 75–88.

Ding, D., Guan, C., & Yu, Y. (2017). Game-Based Learning in Tertiary Education: A New Learning Experience for the Generation Z. *International Journal of Information and Education Technology*, 7(2), 148–152.

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.

Dunlosky J, Rawson K., Marsh E., Nathan M., & Willingham D. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science*. Public Interest, 1, 4-58.

Dunn, O. J. (1964). *Multiple comparisons using rank sums.* Technometrics, 6, 241-252.

El Tantawi, M., Sadaf, S., Alhumaid, J., Lamb, L. C., DiFiori, M. M., Jayaraman, V., & Feeney, J. M. (2016). Using gamification to develop academic writing skills in dental undergraduate students. *European Journal of Dental Education*, 1(15), 1–8.

Ellis, H. (1965). *The transfer of learning*. New York: Macmillan.

Elverdam, C., & Aarseth, E. (2007). Game classification and game design construction through critical analysis. *Games and Culture*, 2(1), 3—22.

Entertainment Software Association (2015). *Essential Facts About the Computer and Video Game Industry*. Recuperado em 3 abril, 2018, site: http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf

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.

Fontanive, N. S. (1982) Técnicas e meios de ensino para educação a distância. *Tecnologia Educacional*, 11 (45), 37-42.

Ford J. K., & Meyer T. (2013). Advances in training technology: meeting the workplace challenges of talent development, deep specialization, and collaborative learning. Em *The Psychology of Workplace Technology*, ed.M Coovert, L Thompson, 43–76. New York: Routledge.

Ford J. K., Baldwin T., & Prasad J. (2018). Transfer of Training: The Known and the Unknown. *The Annual Review of Organizational Psychology and Organizational Behavior*. 5, 1–5.

Ford J. K., Quinones M., Sego D., & Sorra J. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Pers. Psychology*. 45, 511–27.

Ford, J. K., & Kraiger, K. (1995). The application of cognitive constructs and principles to the instructional systems model of training: Implications for needs assessment, design, and transfer. In C. I. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (10, pp. 1-48). New York: John Wiley.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59-109.

Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H., et al. (2014) Active learning increases student performance in science, engineering, and mathematics. *PNAS*, 111(23), 8410–8415.

Gagne, R. (1992). *Principles of Instructional Design*. Fort Worth, TX: Harcourt Brace

Gaudine, A., & Saks, A. (2004). A longitudinal quasi-experiment on the effects of posttraining transfer interventions. *Human Resource Development Quarterly*, 15, 57–76.

Giannetto, D., Chao, J., & Fontana, A. (2013). Gamification in a Social Learning Environment. *Issues in Informing Science and Information Technology*, 10, 195–207.

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). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Goldstein L., & Ford, J. K. (2002). *Training in Organizations: Needs Assessment, Development, and Evaluation.* Belmont, CA: 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., Bastiaens T., & Kirschner P. (2004). A Five-Dimensional Framework for Authentic Assessment. *ETR&D*, 52 (3), 67–86.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3025–3034.

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 Educational Research & 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.

Hunicke, R., Leblanc, M. & Zubek, R. (2004). MDA: a formal approach to game design and game research. *In Proceedings of the AAAIWorkshop on Challenges in Game AI* (pp. 04–04).

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: Instructors guide*. Amherst, MA: HRD Press.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco: Pfeiffer.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*. 93, 59-69.

Kirkpatrick D. (1967). Evaluation of training. In R. L. Craig, & L. R. Bittel (Eds.), *Training and development handbook* (pp. 87-112). New York: McGraw-Hill.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, 35, 179–188.

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

Kraiger, K., Salas, E., & Cannon-Bowers, J. (1995). Measuring knowledge organization as a method for assessing learning during training. *Human Factors*, 37, 804–816.

Laerd Statistics (2017). *Statistical tutorials and software guides.* Retrieved from https://statistics.laerd.com/

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.

Landers, R. (2014). Developing a Theory of Gamified Learning: Linking Serious Games and Gamification of Learning. *Simulation & Gaming*, 45(6), 752–768.

Landers, R., & Callan, R. (2011). Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. In M. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 399– 424). Surrey, UK: Springer.

Lazzaro, N. (2005). Why we play games: Four keys to more emotion without story, *Design*, 18, 1-8.

Lee, J. J., & Hammer, J. (2011). Gamification in education: what, how, Why Bother? Definitions and uses. *Exchange Organizational Behavior Teaching Journal*, 15(2), 1–5.

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. *American Journal of Surgery*, 209(1), 79–85.

Martin, B., Kolomitro, K., & Lan, T. (2014). Training Methods: A Review and Analysis. *Human Resource Development Review*, 13 (1), 11–35.

McGehee W, Thayer P. (1961). *Training in Business and Industry*. New York: Wiley.

McGonigal, J. (2011) *Reality is broken: why games make us better and how they can change the world.* Nova Iorque: Penguin press.

Meneses, P. P. M., Abbad, G. S., Zerbini, T., & Lacerda, E. R. M. (2006) Medidas de características da clientela em avaliação de TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 422-442). Porto Alegre: Artmed.

Mordor Intelligence (2018). *Global Gamification Market*. Retrieved in February 8th , 2019. https://www.mordorintelligence.com/industry-reports/gamification-market

Muntean, C. I. (2011). Raising engagement in e-learning through gamification. In *Proceedings of the 6th international conference on virtual learning* (pp. 323-329).

Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2016). The higher the score, the higher the learning outcome? Heterogeneous impacts of leaderboards and choice within educational videogames. *Computers in Human Behavior*, 65, 391–401.

Noe R., Clarke A., & Klein H. (2014). Learning in the twenty-first-century workplace. *Annual Review Organanizational Psychology and Organizational Behavior*. (1), 245–75.

Noe R., Clarke A., & Klein, H. (2014). Learning in the Twenty-First-Century Workplace. *The Annual Review of Organizational Psychology and Organizational Behavior*, (1), 245-275.

Paiva, R., Bittencourt, I. I., Tenorio, T., Jaques, P., & Isotani, S. (2016). What do students do on-line? Modeling students' interactions to improve their learning experience. *Computers in Human Behavior*, 64, 769–781.

Pasquali, L., & Alves, A. R. (1999) Testes referentes a conteúdo: medidas educacionais. In: L. Pasquali (Org.). *Instrumentos psicológicos: manual prático de elaboração.* (pp.114-155) Brasília: IBAPP.

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.

Pilati, R. & Abbad, G. (2005). Análise fatorial confirmatória da escala de impacto do treinamento no trabalho. *Psicologia: Teoria e Pesquisa*, 20(1), 31-8.

Pilati, R., & Borges-Andrade, J. E. (2006). Construção de medidas e delineamentos em avaliação de TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 359-384). Porto Alegre: Artmed.

Pilati, R., Riether, M., & Porto, J. B. (2006) Estratégias de análise de dados e retroalimentação do sistema de TD&E. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 514-529). Porto Alegre: Artmed.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*. 50(4), 258-283.

Prestopnik, N., Crowston, K., & Wang, J. (2017). Gamers, citizen scientists, and data: Exploring participant contributions in two games with a purpose. *Computers in Human Behavior*, 68, 254–268.

Prince, M. (2004) Does active learning work? A review of the research. *Journal of Engineering Education*. 93(3):223–31

Queiroga, F., Andrade, J. M., Borges-Ferreira, M. F., Nogueira, R., & Abbad, G. S. (2012) Medidas de aprendizagem em TD&E: Fundamentos teóricos e metodológicos. In G. S. Abbad, L. Mourão, P. P. M. Meneses; T. Zerbini, J. E. Borges-Andrade, & R. Vilas-Boas (Org.), *Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: Ferramentas para gestão de pessoas.* (pp. 108-126). Porto Alegre: Artmed.

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

Roediger H., & Butler A. (2011). The critical role of retrieval practice in long-term retention. *Trends Cognitive Science*. (1), 20–27.

Saks, A., & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management*, 45, 629–648.

Santhanam, R., Liu, D., Shen, W. M., & Santhanam, R. (2016). Gamification of Technology-Mediated Training: Not All Competitions Are the Same. *Information Systems Research*, 27(2), 453–465.

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*. Boston, MA: Houghton Mifflin.

Sheldon, L. (2012) *The Multiplayer Classroom: Designing Coursework as a Game.* Boston, MA: Cengage Learning.

Siegle, D. (2015). Technology: Learning can be fun and games. *Gifted Child Today*, 38(3), 192-97.

Siemens, G. (2013). "Massive open online courses: Innovation in education", *Open educational resources: Innovation, research and practice*, 5.

Skinner, B. F. (1938) *The behaviour of organisms: An experimental analysis*. New York: Appleton-Century Company Incorporated.

Su, C. (2015). 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.

Su, C., & Chen, K. (2015). An Empirical Study on the Implementation and Evaluation of Gamifying Learning Motivated Achievement Model. *Applied Mechanics and Materials*, 764–765, 822–826.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3), 1-24.

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.

Taylor, J., Russ-Eft, D., & Chan, D. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, 90, 692-709.

Tenorio, T., Bittencourt, I. I., Isotani, S., Pedro, A., & Ospina, P. P. (2016). A gamified peer assessment model for on-line learning environments in a competitive context. *Computers in Human Behavior*, 64, 247–263.

Thorndike E., & Woodworth R. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*. 8, 247-261.

Training Magazine (2018). *Training Magazine's 2018 Training Industry Report*. Retrieved in February 6th, 2019. https://trainingmag.com/trgmag-article/2018-training-industry-report/

Vargas, M. R. M., & Abbad, G. S. (2006) Bases Conceituais em Treinamento, Desenvolvimento e Educação. In J. E. Borges-Andrade, G. S. Abbad, & L. Mourão (Org.), *Treinamento, Desenvolvimento e Educação em Organizações e Trabalho: Ferramentas para gestão de pessoas* (pp. 289-321). Porto Alegre: Artmed.

Wang, J.-H., Chen, S. Y., & Chan, T.-W. (2016). An Investigation of a Joyful Peer Response System: High Ability vs. Low Ability. *International Journal Of Human-Computer Interaction*, 32(6), 431–444.

Wexley K., & Latham G. (1981). *Developing and Training Human Resources in Organizations.* Glenview, IL: Scott Foresman.

Yamabe, T., & Nakajima, T. (2013). Playful training with augmented reality games: case studies towards reality-oriented system design. *Multimedia Tools and Applications*, 62(1), 259–286.

Yelon , S., Ford , J. K. , & Anderson , W. (2014). Twelve tips for increasing transfer of training from faculty development programs. *Medical Teacher*, 36, 945-950.

Yelon S, Ford J. K., & Bhatia S. (2015). How trainees transfer what they have learned: toward a taxonomy of use. *Performance Improvment.* 27, 27–52.

Yildirim, I. (2017). The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education*, 33(2016), 86–92.

Annex

Annex 1 – Situational performance tests

(1) Pretest

You have just taken over as General Manager of Independence branch, located in a medium-sized city in the interior of Brazil. The unit has nine employees, in the following functions: 1 general manager, one service manager, one assistant, and six clerks. Among them, only 30% have postgraduate degrees and the course indexes at T&D are lower than expected, as well as the internal knowledge certifications obtained by the employees. Communication team is formed and is active. The branch has a history of many health permits. The latest climate survey had unsatisfactory results, slightly lower than expected.

Concerning business and processes, the goal agreement was not fulfilled in the last semester. The physical customers and companies customers' credit percentages have not been delivered for more than one semester. Compliance rates are satisfactory. The branch has a significant flow of face-to-face services, with many retired clients. In the city, there are other competing banks besides very strong and active cooperatives.

BASED ON THIS SCENARIO, EVALUATE AND ALTER, EDITING THE PERFORMANCE AGREEMENTS BELOW, INCLUDING AND EXCLUDING INFORMATION:

Skills performance agreement

Email sent by the general manager for the science of all employees for record of the PA Performance Agreement, where it was determined what is expected, based on the previous semester's goals program, climate research, managing expectations and the purpose of the FI.

- People: it was advised that everyone would do everyone's tasks, especially in absences; the general manager will name the new Communication team and choose the leader; each employee will do two face-to-face courses in the semester. Have an attitude and be the leader of oneself.

- Processes: all should maintain excellence in process compliance. All employees must offer digital solutions to customers.

Goal Performance Agreement

- Business: achieve the goal program with a maximum score in all indicators and deliver 120% of all challenges proposed by Regional Superintendence.

The monitoring will be done by the reports of the programs of goals, with annotation of improvement in the PA system.

As the days go by in your new role, you get to know your team's employees better. What you were able to observe two of the employees are described below:

Based on the employee profile reading, perform the following tasks:

Carefully read the employee's profile

(1) Select ONLY ONE skill to record a note to the employee.

(2) Select the annotation type.

(3) Write an annotation.

(4) At the end of the previous tasks, issue a concept to the employee ONLY IN THE OBSERVED COMPETENCE. For this, consider what description was maintained throughout the PA's assessment cycle.

Peter is a clerk and is 35 years old, single and has been in the FI for ten years. Graduated in Geography, he is outstanding in the team for his technical knowledge, being a specialist in internal regulations. His sales performance is timid, and although he relates very well to customers, he rarely takes advantage of the opportunities that arise in the interactions. He has

indicated enrollment courses. Peter is repeatedly delayed or forgetting to check out his electronic check-in, generating many occurrences. He has manifested family problems concerning his mother's health.

(1) Select ONLY ONE skill to record a note to the employee.

☐ *Fundamental / Internal Processes - Adopts security practices and risk prevention measures, acting following internal norms, legislation and external regulations.*

☐ *Fundamental / Internal Processes - Communicates in writing and orally, in a clear, objective, accessible to the interlocutor and appropriate to the context.*

☐ *Fundamental / Clients - Negotiate with clients (internal or external) transparently, establishing agreements that satisfy the parties and generate lasting relationships.*

(2) Select the annotation type.

☐ Recognition ☐ Enhancement

(3) Write an annotation to the employee

Behavior or situation

_____

_____

Guidance enhancement

_____

_____

(4) At the end of the previous tasks, issue a concept to the employee ONLY IN THE OBSERVED COMPETENCE. For this, consider what description was maintained throughout the PA's assessment cycle.

Martha is currently a Service Manager and is 38 years old, 15 years in the FI and has been in managerial position for four years. She holds a degree in Accounting but does not contain a postgraduate degree. She took started in the FI in this branch. Very responsible, it focuses on the compliance of the processes. Her previous administrators have always rated it saying she could sell more and induce its team to take advantage of all business calls, but never gave her concepts below 5 or made improvement notes on its PA. The PA process she realizes is superficial. Martha indicated only self-instructional courses of short duration. It is observed that Martha spends the whole day attending clients in a very similar way to the clerks of her team. No meetings are held, nor are there individual conversations with your staff for direction and evaluation. No planning was done to reach the budgeted numbers.

(1) Select ONLY ONE skill to record a note to the employee.

☐ *Fundamental / Clients - Serves the internal and external clients, with attention and agility, offering adequate solutions and striving to satisfy their needs.*

☐ *Management / Learning and Growth - Develops the team through the orientation of training actions and the use of frequent feedback and recognition, seeking to reconcile career expectations of the function with the needs of the FI.*

☐ *Fundamental / Learning and Growth - Relates to empathy and courtesy, contributing to team productivity and the excellent atmosphere of the work environment.*

(2) Select the annotation type.

☐ Recognition ☐ Enhancement

(3) Write an annotation to the employee

Behavior or situation

_____

_____

Guidance enhancement

_____

_____

(4) At the end of the previous tasks, issue a concept to the employee ONLY IN THE OBSERVED COMPETENCE. For this, consider what description was maintained throughout the GDP's assessment cycle.

(2) Posttest

You have just taken over as General Manager of Energy Branch, one of the largest branches in the region where you are located. The unit has 25 employees. The branch has four vacancies of clerks not completed for eight months. The team has 50% of the employees with postgraduate and a good percentage of training and many internal certifications of knowledge. The communications team is formed and is active, but even so, the branch climate is far from ideal. The last climate research had a bad result, due to the formation of "cliques" that do not interact and compete a lot among themselves. There are occasional cases of depressed staff members, which may have increased the number of medical leave. The average age of the clients is quite young who use a lot of the digital channels offered by the FI. The branch has fulfilled the latest goal programs, always with emphasis, even losing many points due to the excessive amount of complaints and denunciations by unrecognized sales, the inadequacy of investment products and delays in the queues of face-to-face calls. The branch is always framing compliance at the end of the semester with great difficulty.

BASED ON THIS SCENARIO, **EVALUATE AND ALTER, EDITING** THE PERFORMANCE AGREEMENTS BELOW, **INCLUDING AND EXCLUDING** INFORMATION:

Skills performance agreement

Conduct training at T&D aiming at improving the branch's climate (example non-violent communication course). Take care of health to avoid repeating the unacceptable amount of health left of the previous semester. Focus on the development of competencies related to customer experience focused on the client's vision, carrying out attendance courses and financial investments. Managers should report monthly to a committee meeting on the progress of teams concerning the aspects established herein, especially about team integration.

Goal Performance Agreement

Agreement of goals discussed between the Agency Committee and registered for all employees for the current half-year of PA, where it was agreed what is expected, based on the result of the previous half of the FI and the branch, in the climate research, perceptions and the purpose of the FI.

- Program of Goals: to reach 1 thousand points or the average of Regional Superintendence, whichever is greater.

- Individual mobilizations: deliver 100% of the expected mobilizations to the branch.

- People: hold committee meetings with the participation of all trust function (CF) employees.

The monitoring will be done by the perception of the general manager, with annotations in the PA.

As the days go by in your new role, you get to know your team's employees better. What you were able to observe two of the employees are described below:

Based on the employee profile reading, perform the following tasks:
Carefully read the employee's profile

(1) Select **ONLY ONE** skill to record a note to the employee.

(2) Select the annotation type.

(3) Write an annotation.

(4) At the end of the previous tasks, **issue a concept** to the employee **ONLY IN THE OBSERVED COMPETENCE**. For this, consider what description was maintained throughout the GDP's assessment cycle.

Helen is an assistant, 32 years old, four years in the FI, and was appointed six months ago in the current role. He holds a degree in Information Technology and a postgraduate degree in Business Management. You are excited at the new function. It is well focused on operational activities but has little knowledge of products and services. In the previous agency was a clerk and participated in the Communications team. In her PA he has grades 5 or 6 in all skills, but you do not have annotations. At the courses plan for this semester, Helen demanded three courses on sales and product arguments. She participates in a community on a social network and has captured several sales arguments about pension and capitalization, sharing the findings with colleagues. Helen has celebrated almost every day increasing her internal sales revenue.

(1) Select ONLY ONE skill to record a note to the employee.

☐ *Fundamental / Financial - Disseminates and uses digital solutions to streamline processes, improve customer experience and generate results for the Bank.*

☐ *Fundamental / Clients - Negotiates with clients (internal and external) transparently, establishing agreements that satisfy the parties and generate lasting relationships.*

☐ *Fundamental / Socioambiental - Age considering social, environmental and economic impacts, demonstrating a commitment to sustainable development.*

(2) Select the annotation type.

☐ Recognition ☐ Enhancement

(3) Write an annotation to the employee

Behavior or situation

_____

_____

Guidance enhancement

_____

_____

(4) At the end of the previous tasks, issue a concept to the employee ONLY IN THE OBSERVED COMPETENCE. For this, consider what description was maintained throughout the GDP's assessment cycle.

1 2 3 4 5 6 7

Bryan is a physical relationship manager and has a high business performance. He is always featured in the goals program and is one of the greatest champions in the challenges of the Regional and State. He has a good relationship with his teammates, but he does not talk to colleagues in other sectors. In general, the employees who do not work close to Bryan, do not have a good impression of him, due to the history of fights, discussions, and disagreements. His relationship with colleagues has prevented Bryan from replacing the first manager, even though he deserves the negotiating aspect. Bryan's relationship with customers is, and he has no annotations of non-compliance or complaints.

(1) Select ONLY ONE skill to record a note to the employee.

☐ *Fundamental / Financial - Identifies and seizes opportunities for the Bank, promptly, with a focus on business sustainability.*

☐ *Fundamental / Learning and Growth - Relates to empathy and courtesy, contributing to team productivity and the excellent atmosphere of the work environment.*

☐ *Fundamental / Financial - Disseminates and uses digital solutions to streamline processes, improve customer experience and generate results for the Bank.*

(2) Select the annotation type.

☐ Recognition ☐ Enhancement

(3) Write an annotation to the employee

Behavior or situation

_____

_____

Guidance enhancement

_____

_____

(4) At the end of the previous tasks, issue a concept to the employee ONLY IN THE OBSERVED COMPETENCE. For this, consider what description was maintained throughout the GDP's assessment cycle.

1 2 3 4 5 6 7

Annex 2 – Feedback tab

(1) Pretests

Orientation - Question 1: During the correction, it should be noted that, in addition to the inclusion of aspects related to meeting and follow-up, there was adjustment or exclusion of texts presented purposely inappropriate for the participants.

Question 1

EXPLANATIONS OF THE POINTS IN THE TEXTS

The participant receives a score that:

(1) Training: To adjust or exclude the writing that proposes to carry out two face-to-face courses and offer an additional solution for the low levels of qualification.

(2) Echo: Indicate that Communication team must be maintained or elected with the participation of all employees of the unit and not by imposition of the manager (or exclude the appointment of Communication team only by the general manager) and propose action for Communication team to assist in solving some of the contextual problems of the unit.

(3) Health Licenses: Describe strategy seeking a solution to the problem, such as training actions and improving the quality of life at work.

(4) Climate: Indicate activity that aims to solve the climate problem of the unit.

(5) Credit: Indicate action to improve the results in credit, which are not being fulfilled by the group.

(6) Attendance: Add solution to the dissemination of digital strategies, aiming to solve excess presence that may be reducing the negotiation capacity of the unit.

(7) Absences: Indicate a plan for solving problems arising from deficiencies of employees, taking into account the specific competencies needed to replace absent employees.

(8) Targets: The expression "maximum mark in all" should be removed and the text written indicating what is necessary to comply with the agreement.

(a) Meeting: Inform accomplishment of a meeting of the team for construction, because the agreements should not be a managerial imposition, but something discussed in the team to generate the commitment necessary to achieve the agreed performances. (Only punctuate item a - meeting).

(b) Monitoring: Include the word "recognition" or another form of follow-up agreed upon at this point in the text.

Note: Items marked with a and b (meeting and follow-up) of question 1 and do not punctuate item c (contextual aspects).

Guidance - questions 2 and 3: You can not mention aspects related to the work day (electronic check-in), health problems or disciplinary issues in annotations in PA. The participant who suggests these themes will not receive punctuation in any item of the question. Incorrect marking on part 1 results in no punctuation in section 3 but does not override punctuation in part 2.

Question 2 (Peter)

Expected responses:

PART 1 (Association)

(X1) Adopts practical ...

(X1) Recognition

(X2) Negotiates with customers ...

(X2) Enhancement

PART 2 (Annotation)

Expected response (X2)

Behavior or situation

The main activity in compliance with internal regulations

Guidance enhancement

Peter, according to the feedback provided today, I would like to congratulate you for your performance in complying with FI regulations. Stay focused on process compliance by leveraging your knowledge to leverage business. I am on hand to support you on this journey. Let's combine a conversation for the next month. Congratulations!

Expected response (X2)

Behavior or situation

Failure to take advantage of business opportunities

Guidance enhancement

Peter, according to the conversation held today, urged you to improve your position regarding business opportunities in interaction with clients. I suggest that, in addition to seeking knowledge at T&D, talk to more experienced colleagues to grow their business performance. I'll make myself available if there are any difficulties. We talked again in two weeks to check on your development.

PART 3 (Concept)

(X1) 5 or greater

(X2) 4 or less

Question 3 (Martha)

Expected responses:

PART 1 (Association)

(X) Develop the team ...

(X) Enhancement

PART 2 (Annotation)

Behavior or situation

Martha, according to the feedback issued today, third in this semester, it is necessary that there is an urgent change of performance regarding the development of the team, in keeping with the managerial function that you occupy.

Guidance enhancement

Martha, I recommend an immediate change of attitude regarding the management of your team. Take the leadership trail courses this week, and I'll give you support for that. Have a meeting with your team's staff and set up a plan of action for the next 30 days. Act as a point of support for the team and establish strategies for reaching the budgeted figures. I already left a scheduled date for our next conversation in two weeks. I am available.

PART 3 (Concept)

3 or less

(2) Posttest

Question 1

EXPLANATIONS OF THE POINTS IN THE TEXTS

The participant receives a score that:

(1) Climate: Indicate action that seeks to solve the climate problem of the unit other than the indicated course.

(2) Health Licenses: Adjust or delete text that mentions an unacceptable amount of licenses and proposes action that seeks to solve the problem presented.

(3) Complaints in regulator: Indicate in the agreement of goals the need to follow up this number.

(4) Unrecognized sales: Indicate an action proposal to solve this severe problem.
(5) Investments: Add another strategy to the text that proposes the accomplishment of training to solve the problem that is reflected in the inadequacy of products.
(6) Attendance: Add another strategy to the text that proposes the realization of practice to solve the problem.
(7) Conformity: Indicate plan to seek to address the problematic framework of compliance indicators.
(8) Goals: Change or delete the punctuation or the word "average" by a percentage, more accessible to follow.
(a) Meeting: Inform in the text of the competency agreement the holding of the meeting with the team.
(b) Follow-up: Add to the text follow-up action that extrapolates the perception of the general manager.

Note: Items marked with a and b (meeting and follow-up) of question 1 and do not punctuate item c (contextual aspects).

Question 2 (Helen)
Expected responses:
PART 1 (Association)
(X1) Spread and use ...
(X2) Negotiates with customers ...
(X) Recognition

PART 2 (Annotation)
Expected response (X1):
Behavior or situation
Recognition for the dissemination and sharing of a sales pitch with teammates.
Guidance enhancement
Helen, as we talked yesterday, I come here to congratulate you for the initiative in seeking knowledge and sharing them with the team. I suggest you stay focused and dig deeper into other FI products, such as credit and investments. You can look for me to talk whenever you feel the need. I will follow your footsteps closely, for I see in you a future FI manager.

Expected response (X2):

Behavior or situation

Improved business performance reflected in increased sales revenue

Guidance enhancement

Helen, as shared with the entire team at our general meeting, I congratulate you on improving your business performance. I suggest that you deepen your knowledge through T&D, internal regulations and contact with more experienced colleagues. I am available to assist your professional development process. We'll talk again early next month. Carry on, congratulations.

PART 3 (Concept)

5 or greater

Question 3 (Bruno)

PART 1 (Association)

(X1) Identify and enjoy ...

(X1) Recognition

(X2) Relates to empathy ...

(X2) Orientation ...

PART 2 (Annotation)

Expected response (X1):

Behavior or situation

Great business performance

Guidance Enhancement

Bryan, as a conversation after recognition of Super, I congratulate for the highlight achieved in another regional challenge. Seek to harness its full potential to help other colleagues in the unit to produce results, after all; we are a team. I will continue to monitor your progress, understanding that you have the potential to become the first manager. We meet again before the end of the month.

Expected response (X2):

Behavior or situation

Relationship with colleagues has been detrimental to the performance of the unit.

Guidance Enhancement

Bryan. After repeated conversations on the subject, the last being held today, I determine that you change your behavior immediately. Conduct leadership training and nonviolent communication this week. I make myself available for a weekly meeting to report on your progress concerning the colleagues in the other teams in the unit. If you do not reposition yourself, we will have to deal with the matter in a disciplinary manner.

PART 3 (Concept)

(X1) 5 or greater

(X2) 3 or less

Annex 3 – Assessment sheet

Pre test ( ) **Post test ( )**	**Final Grade>>>>>>>>>>**	0		
ASSESSMENT SHEET		YES	NO	BLANK
Question 1				
a) Did the participant report if the performance agreement was built with the team's participation?				
b) Did the participant report how the performance agreement will be monitored?				
c) Did the participant write action plans to improve the performance of the fictional unit, including, excluding or altering the given performance agreement? (Inform the amount - limited to 5)				
Question 2				
PART 1 (Association)				
Did the participant properly select a competence and the type to write the annotation to the fictitious character, observing the performance description?				
PART 2 (Annotation)				
a) Did the participant describe behavior or situation appropriate to the scenario given, annotation's type and the competence selected?				
b) Did the participant describe enhancement guidance appropriate to the scenario given and the competence selected indicating path with action to improve performance?				

Pre test ( ) **Post test ( )** **Final Grade>>>>>>>>>>**	0		
ASSESSMENT SHEET	YES	NO	BLANK
c) Did the participant report if there was dialogue (verbal feedback) with the simulated employee before the annotation?			
d) Did the participant report how the simulated employee will be monitored?			
e) Did the participant report being available to assist the performance improvement of the simulated employee?			
f) Did the participant use direct speech in the annotation, directing it to the simulated employee?			
PART 3 (Score)			
Did the participant give an adequate score for the selected competence, considering that the behavior description was maintained throughout the assessment cycle?			
Question 3			
PART 1 (Association)			
Did the participant properly select a competence and the type to write the annotation to the fictitious character, observing the performance description?			
PART 2 (Annotation)			
a) Did the participant describe behavior or situation appropriate to the scenario given, annotation's type and the competence selected?			

Pre test ( ) **Post test ( )** **Final Grade**>>>>>>>>>>	0		
ASSESSMENT SHEET	YES	NO	BLANK
b) Did the participant describe enhancement guidance appropriate to the scenario given and the competence selected indicating path with action to improve performance?			
c) Did the participant report if there was dialogue (verbal feedback) with the simulated employee before the annotation?			
d) Did the participant report how the simulated employee will be monitored?			
e) Did the participant report being available to assist the performance improvement of the simulated employee?			
f) Did the participant use direct speech in the annotation, directing it to the simulated employee?			
PART 3 (Score)			
Did the participant give an adequate score for the selected competence, considering that the behavior description was maintained throughout the assessment cycle?			