Gamification strategies as formative assessment methods.
A systematic review

1Manuel Larrosa, 2Leandro Wives, 3Virginia Rodés

1DINTER - Universidad de la República, Montevideo, Uruguay.
2 PGIE - Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.
3Instituto del Futuro de la Educación - Tecnológico de Monterrey, México.

manuel.larrosa@gmail.com, lwives@gmail.com, virginia.rodes@gmail.com

Abstract. This article aims to deepen the analysis of gamification strategies as an appropriate method to offer feedback in the form of formative assessment for elementary school students (K-12). Although gamification strategies have been crucial in the last decade, studies do not discriminate between game elements, such as leaderboards, levels, rankings, etc. For this reason, this study focuses specifically on one of them: awarding badges to learn more about effects. Beyond its potential, incorporating these practices also supposes a significant challenge: the need for evaluation that formal education systems demand. Based on these topics, a systematic review of the literature covering the period of 2011 to 2020 is proposed, specifically delving into the link between gamification strategies in formal education environments and the implementation of formative assessment techniques through gamification by badges.

Keywords: gamification, badges, formative assessment, motivation.

1. Introduction
The constant interest from educators and researchers in the field of innovation in education has led in recent years to a growing focus on game dynamics as a motivating mechanism appropriate to deliver concepts framed in various disciplines. Several strategies, known as game-based learning, serious games, and gamification, have been developed as practical tools for educational content. This interest is accentuated by the place that mobile devices, digital applications, and video games occupy in audiences of all ages. Digital environments permeate all manifestations of recent human activity and, therefore, also permeate gaming logic. Likewise, education is no stranger to digitization and finds in it as much potential as challenges. At this intersection between educational systems that must adapt on the move and the overwhelming ubiquity of digitization, questions multiply, while some of their answers can be found in gamification. Definitions
agree that it is the application of game logic to other contexts to invoke the motivating and relaxation experience that playing provides. Different definitions link the term gamification with digital environments, tying it to the development and proliferation of video games and gaming. The advantages it introduces seem evident, and there is a remarkable consensus and extensive bibliography. Among them, it can be mentioned the motivation and involvement that generates amid participants; the wealth of practices that it offers for the design of activities by teachers; the horizontal exchanges that it promotes between teachers and students, which also encourage an active role from all those immersed in its dynamics. Despite this, when its application is analyzed within formal educational systems, a question emerges of how to evaluate the development of activities carried out through gamified strategies. Even deeper, when it comes to delivering educational content, it is important to know how students could assess its appropriation. As Sailer et al. (2017) argue, many studies consider gamification as a uniform concept when in practice it takes many different forms, with heterogeneous designs and environments, since gamification can manifest itself in various ways and combine multiple game elements. Therefore, its results are also very diverse and are directly tied to the elements used in each of the strategies designed. The objective of this article is to carry out a systematic review of the literature to identify the game elements that can be part of a gamified environment. In turn, it seeks to deepen the possibilities that gamification offers as a formative evaluation mechanism, as feedback to the activities carried out inside and outside the classroom. Two large areas and their possible link are then considered. On the one hand, gamification strategies themselves, which are made up of various game elements such as badges, missions, points, levels, leaderboards, and within these, specifically the badges earning as a motivating mechanism to encourage student involvement or engagement. On the other hand, what is considered formative evaluation, i.e., the feedback process students receive about their school performance. At the intersection of these two fields, the awarding of badges can be seen as a method of formative assessment for primary school students (K-12), as an element of gamification with multiple purposes: promoting motivation, facilitating game dynamics for teachers, offering attractive and enriching returns for students which channel educational content and focus on curricular content.

2. Systematic study procedure
To assess the potential of this approach, a systematic review is proposed, aimed at finding gaps in the literature. The two large areas mentioned (gamification and formative assessment) have been widely studied but independently. The hypothesis that this study raises is the relatively low presence of research that links the two areas and the potential that they can offer together as a pedagogical practice. Systematic studies are an effective research method to identify the state of the art in a given topic by mapping and classifying the studies available in the specific literature [Klock 2018]. Conforto et al. (2011) argue that the main problem of non-systematic studies is the lack of rigor, which results in a personal interpretation of the texts, but with little critical analysis. Therefore, they agree that the most important contribution of systematic studies is the reliability in which they support conclusions by applying a method with exhaustive rigor, capable of reducing the biases that may influence the investigation. Klock et al. (2020) define a series of steps to follow, which are considered for this study. They are described below:
A. Planning - at this stage, is recommended conducting a first exploratory investigation to obtain a general panorama of the field of study. This overview allows evaluating the real need to carry out the secondary study if extensive previous literature justifies the systematic study. As a central point, in this stage are defined the research questions that guide all subsequent development.

B. Protocol - in this step are determined the methods that will be adopted throughout the procedure to reduce the possibility of biases.

C. Conduction - finally, in this stage is applied the search protocol. In this last step, it is suggested to use a series of selection criteria to systematize the process. Conforto et al. (2011) describe this stage as an iterative process in which search filters are applied to refine the results.

3. Objective and research questions

The central objective guiding this systematic review is to map academic production in the fields of gamification strategies in educational settings and its link with formative assessment mechanisms. The following research questions are posed:

1. Within the studies on gamification: is there evidence that links reward systems through badges with formative assessment?

2. Are these studies directed towards formal education settings for students under 12?

3. What are the main findings regarding formative assessment through gamification strategies by the awarding of badges?

4. Selection criteria

The literature on gamification has been very profuse in recent years, so it is necessary to limit the period in which the analysis will focus. The same happens regarding formative assessment, which has gained significant focus due to the development of skills that cannot be evaluated on a summative basis. Therefore, according to the general panorama obtained in a first exploratory mapping, the selection criteria are the following: studies published between 2011 and 2020; written in English; finished articles, no previews or short articles (6 or more pages), nor complete books, theses or conferences; primary studies, not systematic reviews or mappings; studies that show the link between gamification through reward systems based on badges and formative assessment; studies that focus on formal educational settings, at primary education level (K-12).

5. Search engines and keywords

From a general survey of the topic and other systematic mappings, it appears that the most appropriate search engines for this study are ACM Digital Library, IEEE Xplore, Scopus, and Science Direct. For each of the two major topics on which the search will focus, the keywords and their most common synonyms are: gamification, including its possible variants and terms linked to specific gamification strategies - gamified [gamif*], game-based, badges, open recognition, rewards. Formative assessment [assess*], evaluation [eval*], feedback.
Therefore, the search argument in each case must obey the particularities of the operation of search engines, since each one has its own specificities. The logical operator OR is used within each group, while the logical operator AND is used to link both. In addition, the NOT logical operator was used to eliminate unwanted themes or topics specific to repository's profile, which appear very frequently. Thus, the search argument for each site is configured as follows:

ACM: (gamif* OR "game-based" OR "badges" OR "rewards") AND ("feedback" OR "K-12") NOT ("higher education" OR "high school") Other filters applied in ACM: between 2011 and 2020, only journals, only research articles, PDF, published by ACM.

IEEE: gamif* OR "game-based" OR "badges" OR "rewards" AND assess* OR "feedback" OR "K-12" NOT "higher education" NOT "high school" NOT "software engineering" NOT "computer science" NOT CS NOT engineering.

Scopus: (gamification AND feedback) OR (badges AND assessment) OR (rewards AND open PRE/recognition) AND NOT (high PRE/ school OR higher PRE/ education) [PRE/ - term to indicate that said word precedes the following].

Other filters used: only articles, no conferences, no short papers; by subject area - social sciences, computer science, psychology, arts, and humanities; by language - only in English.

Science Direct: (gamification OR "badges" OR "open recognition" OR "rewards") AND (assessment OR "feedback" OR "K-12" OR education).

In addition, other filters were established by the search engine to refine the search: containing the keywords only in the title and abstract, only articles published in journals, and only in PDF format.

6. Results

From the application of the search arguments in each repository, were obtained the results shown in Table 1. After the results are returned by each repository, an approximation is carried out that analyzes first the title; if it is related to the topic, we move on to the analysis of the abstract. If there are doubts about its relevance to the topic, in-depth reading is carried out to include it or discard it from the study.

<table>
<thead>
<tr>
<th>Search engine</th>
<th>Abstract</th>
<th>Title</th>
<th>Intro</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>479</td>
<td>14</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>IEEE</td>
<td>574</td>
<td>21</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>280</td>
<td>28</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Science Direct</td>
<td>236</td>
<td>10</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>
One aspect that immediately emerges is the evolution that gamification as a subject of study has developed in the studied period. Graph 1 shows the years of publication of the selected articles. Also included in this graph are the articles selected in this review from the cross-searches that derives from the selected papers. Beyond this systematic review, which due to its scope, is not representative of the entire field of gamification, multiple authors highlight the exponential increase in the topic in recent years [Hamari 2017]. The reasons suggested by these authors link this increase to the psychological experience that playing arouses through gamified practices, which through playful stimuli, seek to motivate those involved to achieve specific objectives, modify their behavior or develop new skills [Klock et al. 2015].

Graph 1 - Evolution of publications on gamification during the period 2011 - 2020

7. Discussion

Studies on gamification have been approached in very different ways, in some cases analyzing its potential as means to deliver other knowledge, in other cases emphasizing its own characteristics as a pedagogical tool to promote student involvement, motivation and provide specific achievement criteria. Therefore, it is necessary to clarify the field of study since not all experiences that integrate game dynamics can be considered within the field of gamification. In this sense, Chu et al. (2020) distinguishes between what they consider learning supported by games and gamification. In the first case, it involves environments formally designed to channel specific concepts in a particular area and with a function not directed by entertainment but by the content they transmit, known as serious games. While in the second case, gamification focuses on using game elements in a combined or isolated way, framed in educational experiences incorporated in real-life contexts. Regarding the game elements considered within what is defined as gamification, the literature has had an exponential production in the last decade, but in general, they have been analyzed jointly. Aligned with the core of studies on the subject, Chu et al. (2020) mentions among the main game elements that participate in gamified experiences: points, levels, badges, leaderboards, rewards, progress bars, and narratives. However, as they point out, not all experiences include all these elements but emphasis is placed on some of them to make them more effective. In general, they affirm that, although the experiences may integrate several of these, most use only 2 or 3, enough to improve the students' learning results. Nevertheless, there is a real need to analyze some of them in isolation to delve into their effects and psychological mechanisms. Specially in the phenomenon of rewards based on badges, which have received particular attention and investment in online education in recent years. Using the research questions posed as a guide, they are answered one by one below:
1. Is there research linking badge reward systems with formative assessment?

There is a broad consensus in defining gamification in educational environments as a series of practices that seek to adapt elements of video game design to other areas of daily life [Deterding 2011; Hamari et al. 2014; Sailer et al. 2017], to encourage motivation and participation, recognize and validate skills, and offer credentials indicating the knowledge acquired. Among the most commonly mentioned elements are the delivery of points, badges, rewards, leaderboards, rankings, levels, campaigns, and progress bars, among others. However, as Van Roy et al. (2019) mention, each of these elements separately has a different degree of significance in the effects it generates, determining a specific motivational function. For example, personalization promotes autonomy; badges aim to establish clear goals to achieve; leaderboards encourage competition. Therefore, when making a more detailed study of the game design elements grouped within what are considered gamified strategies, it is necessary to analyze each separately to quantify their effects. In this case, the focus on badges is because it is an element that has gained relevance as a subject of study in recent years, due in part to its widespread presence in many studies, but above all due to initiatives such as Mozilla Open Badges or Khan Academy, which positions them as an alternative credential system to be used in the educational or work environment. Gibson et al. (2015) define badges as the representation of an achievement obtained in a digital, visual format and available online, containing metadata that helps to contextualize its meaning, the process, or the activity by which it was obtained. For the approach pursued by this systematic review, the definition introduced by the author coincides with the hypothesis that links badges in digital format, with the possible return that they transmit. Another symbolically similar example, previous in time but that follows the same purpose, are the badges used by Scouts, which in physical format have the same characteristics described by the author: they offer members of the organization the possibility to certify certain knowledge, which are generally not part of formal education; it makes this recognition visible, which within the internal structure of the organization functions like a CV, as a sample of their skills. One of the common features that both examples share (digital badges and physical badges) is the motivating characteristic, since they act as milestones in a learning path, establishing specific goals to aim for, tracing a path in which there are certain obstacles to go through, while at the same time making those achievements visible once they have been completed. Another critical feature they share is that both examples are developed outside formal education settings, allowing for the recognition of skills that would not be certified otherwise. In short, reward systems based on badges have the advantage of validating abilities not contemplated in traditional school curricula and providing external links that extend the information the badge certifies. As a weakness, in most cases, the validation of the skills developed is only recognized by the organizations that grant them, but there is still no fully shared standard.

2. What is referred to when talking about formative assessment?

Perhaps the best way to explain it is in comparison to summative assessment. The first thing to mention is that any form of evaluation is a way of contrasting student performance through categories that allow progressively segmenting levels. While the summative evaluation does it from a numerical scale or through standardized judgments, which seek to measure the learning developed by a student and place it on that scale, the formative assessment emphasizes the conceptual content, focusing on generating feedback about the activity carried out. Hattie & Timperley (2007) refer to the qualities...
an effective evaluation must have. First, they mention that the evaluation must contain specific information about the task or learning process. Its function is to be a frame of reference that the student can confirm, contrast, and restructure. Therefore, formative assessment is a form of return (feedback) that cannot be sustained in a vacuum, by the contrary, its potential is given by the learning context on which it is built. According to the authors, formative assessment must answer three questions: where is the student going? How is his process going? What are his next steps? In this way, to be truly effective, it must be loaded with the conceptual content intended to be transmitted, trace objectives achieved, and future goals towards which to go. On the contrary, the summative assessment lacks conceptual or superficial content since it only transmits the student's location on the measurement scale through a numerical value or a standardized judgment. A widely cited study linking the psychological qualities that develop motivation with formative assessment is that of Abramovich et al. (2013). The authors mention that one of the strengths of a system that promotes formative assessment through badges is that it allows much more freedom when facilitating student feedback. Especially if one takes into account that the focus is on those skills not contemplated within the traditional curricular framework, this is the case of the so-called 21st-century skills, social skills such as teamwork, communication skills, skills to overcome frustration, entrepreneurship, creativity, among others; that in general are not part of academic or professional certifications and therefore are not easily validated, but that nonetheless are increasingly valued as elements to be developed for the growth of the individual. As the authors suggests, the institutions that award educational badges can validate through these any type of skill, knowledge acquired, or completed achievement as a complementary recognition to the summative evaluations that make up the official certifications. This way, badges can be viewed and shared with others as evidence of the competence developed. For this reason, they can be an element that influences and encourages involvement in educational activities and therefore learning.

3. Are these studies directed toward formal education settings for K-12 students?

This is one of the main findings of this systematic review. Given the nature of badge systems, which make it possible to recognize and validate skills that are not necessarily part of formal curricula, they have been used mainly in non-formal education settings. With the development of Virtual Learning Environments and educational platforms that schools adopt due to increasing digitalization, their use has intensified. However, since they are still considered as a complementary playful element, more as a motivator than an evaluative tool or as feedback, their use is still marginal. There are exceptions, such as Khan Academy, a widely used platform that does not escape the logic of being an accessory tool [Van Roy et al. 2019]. However, given that badges allow the skills developed to be added to a digital portfolio, and that these become part of a sort of CV of a candidate, they find significant potential among advanced students of secondary, technical, or tertiary education in general, besides workers seeking to validate skills or be certified in fields that add professional value. Therefore, most studies that focus on gamified environments that award badges do so by analyzing these qualities, mainly oriented towards adults, advanced students, or professionally active populations; but to a lesser extent, they are directed towards students at the school level. Even though its use is more widespread among young people and adults, studies show that its potential as a solid tool for formative assessment makes it viable to be adopted at any level. This fact is mainly reflected in the playful and visual qualities that badges integrate, especially among
school-age children. But their potential is not only given by their visual characteristics, much further by the content they can deliver. As in the case of badges that contain integrated information about the activity for which they were awarded, skills obtained, number of hours dedicated, date, and institution. In the same way, badges can channel for younger students, information related to the concepts sought to be transmitted, linked to the school curriculum.

4. What are the main findings regarding formative assessment through badges?

It should not be ignored that a reward system based on delivering educational badges is not exempt from possible negative connotations. In this sense, the literature about intrinsic and extrinsic motivation is forceful in its conclusions, and its results also apply to the case of a badge system. As Van Roy et al. (2019) argue, gamification in educational contexts is conceived as an instrument to enhance motivation, understanding that this factor is one of the most determining factors in school performance. Therefore, the effect that a badge system has on motivation is central to discussing its scope. First, it is necessary to differentiate between intrinsic and extrinsic motivation. The first refers to the inherent tendency of human beings to seek novelty and challenge, to extend and exercise their own abilities, to explore and learn. The second refers to the performance of an activity to obtain some result and therefore contrasts with an intrinsic motivation that suggests doing an activity for the inherent satisfaction that the activity itself causes. To analyze the scope of motivation, Deci & Ryan (2010) introduce the Self-Determination Theory (SDT), with which there is a broad consensus on its usefulness in explaining the phenomenon of motivation. The SDT exposes that the activities involving feedback, communications, and rewards, which lead to feelings of competence as part of their development, can amplify the intrinsic motivation in the process. Likewise, appropriate challenges to its participants and positive feedback on performance also enhance intrinsic motivation, while negative feedback has the opposite effect. In the same way, what the authors call functional significance explains that a stimulus by itself is not capable of determining motivation, but rather its motivational impact is mediated by the functional significance that the individual gives to it. Therefore, the way in which a game element is capable of motivating is determined by how the user applies it. In other words, there is no direct relationship between the game elements and the motivational effect they generate. Each of these can trigger a different effect on the user, depending on its meaning and how it is put into practice. This means that a particular game element can have a certain effect (or none), and it depends directly on the individual's intrinsic motivation, which functions according to his psychological needs. In this way, intrinsic motivation is manifested when an individual, for example, practices a sport or plays a musical instrument, simply because he enjoys it and is satisfied. This is how he regulates himself by his own choice and interest. This intrinsically motivated behavior does not require any kind of reinforcement and is a prototypical example of self-determination. The internalization of external motivators gives the opposite extreme of self-regulation. These manifest when the motivation is directed towards a benefit that does not come from the individual himself, but is external, such as receiving a congratulation, a reward, or a gift for reaching a goal. The internalization process occurs when a functional significance is given to an activity that is incapable of developing intrinsic motivation, which is driven by external motivators that give it value, and in the ideal case, can even promote self-regulation and self-determination. In the same way, Auvinen et al. (2015) reinforce the idea that, in educational contexts, different students can react differently to the same
gamification method. Even gamification can produce a result contrary to what is expected when it reduces the internal motivation towards the activity, replacing it with an external motivation. This happens when external rewards are perceived as controlling and not informational or reinforcing something that is intended to be emphasized. In this sense, the so-called achievement badges have a function that occurs parallel to the game's goals or activity. As Hamari & Eranti (2011) mention, they introduce an optional goal independent of the game objectives. However, they have an essential function, which is to visualize certain stages in progress in the activity and offer feedback on the performance of the participant. The social function they introduce is even as important as their feedback function since they allow these achievements to be exposed to other participants or spectators. Certain game elements are more effective than others when it comes to providing feedback, such as badges, as the recognized element with the most weight in this regard. Badges are defined as visual representations of an achievement that can be obtained or collected within the gamified environment. Its function is to make the achievement visible, confirm and clarify it for the user, but more importantly, show the scope of this goal to the other participants. In this sense, badges fulfill a double function; on the one hand, they officiate as a return for the user that indicates that they are in the right direction, but at the same time, they socialize the objective achieved. It is worth asking, what psychological need does each of the elements satisfy? In the case of badges, they offer a visual return, show achievement, and are a way of evaluating the user's performance. This assessment in the form of feedback evokes the feeling of competence, by directly communicating the success in the development of the activities by the player. But we must not lose sight of the fact that all the elements mentioned are external motivators, and if they do not acquire a functional significance that positively internalizes them, they can generate a negative motivating effect. As Van Roy et al. (2019) mention, a classroom environment that promotes autonomy can functionalize the teacher's feedback as informational and enriching, and therefore encourage the development of intrinsic motivation. But, on the contrary, a controlling environment can frustrate autonomy and therefore foster a feeling of extrinsic motivation, which is consistent with negative consequences. The authors state a direct link between highly self-determined environments driven by intrinsic motivation and positive educational outcomes, whereas, in the opposite case, environments extrinsically motivated through control correlate with negative outcomes.

**Figure 1. Theoretical framework from Van Roy et al. 2019.**

Hamari and Eranti (2011) define badges as optional rewards and goals outside the objectives that make up the core of an activity. This definition suggests that the optional
character accompanying the badge does not involve a controlling attitude, and therefore a negative extrinsic motivation, which, although external, can be positively internalized. This aspect is highlighted in the fact that obtaining or not medals does not affect the participant's progress in the activity, but rather, in such a case, they enhance or visualize it. Still, they do not suppose a goal by itself. Instead, the goal is the activity or challenge that awards the badge. One of the most important conclusions derived from the study by Abramovich et al. (2013) is that different types of badges affect student motivation in different ways. In the case of low-performing students, badges work as an incentive to participate in educational activities, but do not significantly affect skills acquisition. Even more importantly, they can have a negative effect on learning, since the study confirms that extrinsic motivators result in a certain rejection of the activities they promote. Therefore, one of the recommendations for instructional designers is to consider the capabilities of the public to which they are directed, since, if badges do not channel appropriate content for a specific public, there is a possibility of generating a motivational effect contrary to what is desired.

8. Conclusions
From the research questions, a clear gap in the literature regarding the link between both topics emerges in the first instance. Although gamification strategies have been addressed from different academic fields, in general, the focus of analysis and interest in its application is given by its motivational qualities or by introducing methodologies that promote active dynamics between teachers and students. However, the focus on formative assessment has been outlined more superficially, and it is here where a future study can emphasize this technic's advantages. A recurring criticism of reward systems is the questioning of their behaviourism nature. In this sense, the studies by Hamari & Eranti (2011), Abramovich et al. (2013), and Van Roy et al. (2019) agree that there is no direct determinism between the function of the badge and its effect. Because many other elements come into play, such as each player's characteristics, context, and motivation toward the activity, it is challenging to establish causality. Mainly due to the SDT, the motivating scope is directly determined by how each one functionalize the stimuli. Van Roy et al. (2019) mentions it when stating they "question the existence of a one-to-one relationship between some game element and its motivational function". It suggests that any game element can trigger a motivational function (or not), depending on the meaning a user gives it and the relationship established between it and their psychological needs.

In conclusion, it can be deduced that if there is no direct relationship between the influence of the badges and the effect they generate, it is difficult to sustain the behaviourist hypothesis, by which it is understood that the delivery of badges is an external reinforcer and therefore its function is to change behavior. Based on the arguments presented, it is understood that the elements that come into play in motivation are multiple, very complex, and that their effects do not always have a direct and unilateral relationship with the function given to them. Furthermore, a strong two-way link between the user and the different game elements draws up diverse relationships. Game elements are linked to each other since they rarely exist in isolation; at the same time, they are nourished by the psychological needs of the users and how each one functionalizes their presence according to their motivation, obtaining a different effect on the game in each case.
References


