A Systematic Mapping of Negative Effects of Gamification in Education/Learning Systems

Cláuvin Almeida
Department of Informatics
PUC-Rio
Rio de Janeiro, Brazil
almeidaclauvin@gmail.com

Marcos Kalinowski
Department of Informatics
PUC-Rio
Rio de Janeiro, Brazil
kalinowski@inf.puc-rio.br

Bruno Feijó
Department of Informatics
PUC-Rio
Rio de Janeiro, Brazil
bfeijo@inf.puc-rio.br

***Abstract*—While most research shows positive effects of gamification, the focus on its adverse effects is considerably smaller. Having this in mind, we conducted a systematic mapping study of the negative effects of game design elements on education/learning systems. The study revealed 77 papers reporting undesired effects of game design elements. We found that badges, competitions, leaderboards, and points are the game design elements most often reported as causing negative effects. The most cited negative effects were lack of effect, lack of understanding, irrelevance, motivational issues, and worsened performance. The ethical issue of cheating was also often reported. As part of our results, we map the relations between game design elements and the negative effects that they may cause. Our mapping study can help gamification designers make more informed decisions when selecting game design elements to be included in education/learning systems, raising awareness on potential negative effects.**

Keywords— gamification; negative effects; education; learning; systematic mapping; snowballing.

I. INTRODUCTION

There are plenty of digital platforms for education with a massive number of users, as Duolingo, a language teaching service used by 300 million people worldwide. It boasts that it gives opportunities for people to learn new languages, no matter their financial standing [1].

Duolingo and other services worldwide use gamification - applying game-playing elements to non-game contexts [2] that are typically tedious, discouraging, or inefficient - as a strategy to make their objectives more achievable. It's a strategy with a strong presence in education, and other domains [3], representing a market predicted to grow over 30% through 2019-2025, with an expected value of more than 32 billion in 2025 [4]. This context means that a significant demand for digital gamified systems exists, which calls for software engineers to create them.

Software development is not a trivial task, and it is more complicated in the case of gamified software solutions. These cases require specialized expertise, going beyond what is expected by an average software engineer [5], for instance:

- Effective gamification requires knowledge of human psychology, similar to how serious games require knowledge regarding the subject they deal with. This necessity arises because gamified software aims to change human behavior (see Volkswagen's Fun Theory videos ([6],[7],[8]);
- Software engineers need a good understanding of the game design mechanics used as tools and how they contribute to functional and non-functional requirements;
- Software engineers face the fact that gamified software has a more limited design space and different objectives to focus on than a game [9].

Hence, selecting the correct gamification elements when designing gamified systems is strongly related to requirements engineering and can affect the overall project success [10]. Moreover, defects in requirement are the most expensive to fix when found in production [11][12]. Indeed, given that gamification deals with changing human behavior, when gamified software is ill-specified, the system may not hit the intended target or even be counterproductive, which can have serious consequences when applied to education.

Education software is the main target of the present work for an important reason: approximately 617 million children and adolescents of primary and lower secondary school age (roughly 55% of the global total) have not reached minimum reading and mathematics proficiency in 2015 [13]. The reasons for this global learning crisis are manifold, such as inequality and poverty, but the poor quality of education is one of the critical causes. In this context, applying gamification to education and learning systems represents a promising means to allow educators to make learning fun, contextualize learning quickly, speak the language of young people, and directly deal with soft skills, improving education quality. However, these systems are prone to generate harmful effects, usually unknown to designers and engineers. These unexpected effects happen because current gamification research lacks a critical lens capable of exploring unintended design consequences [3].

This paper presents a systematic mapping of adverse, unintended gamification effects in digital education and learning systems (such as modified learning management systems, coursewares and digital learning environments). Our search method is driven by broad research questions and uses a hybrid search strategy [14], combining database search with backward and forward snowballing. As far as we know, this paper is the most precise, comprehensive, and up-to-date systematic mapping study that organizes evidence regarding the adverse effects of game design elements in gamified education systems.

Our study included 77 papers reporting several different negative effects. Based on this material, we present valuable information for software engineers and designers of gamified education/learning systems, such as the game design elements that have most often been reported to cause adverse effects; the most common negative effects on students; the adverse consequences affecting teachers; the relation between game design elements and negative effects; the fields in which research on negative effects has been conducted; and the types of empirical studies conducted to assess adverse effects.

The organization of the paper is as follows. Sections II and III present the fundamentals of gamification and gamification effects. Section IV discusses related work. We explain our systematic mapping method in Section V and show the results in Section VI. Finally, Sections VII and VIII present the limitations and conclusions, respectively.

978-1-6654-2705-0/21/$31.00 ©2021 IEEE
DOI 10.1109/SEAA53835.2021.00011

II. GAMIFICATION

The term "gamification" became known after 2010, but we can find precursors in user interface studies of the 80's [2]. If we use the broader perspective of work's gamification, we find similar concepts even earlier, such as the Soviet Union's experiments to motivate workers of the first half of the 20th century [15]. At the beginning of the '2000s, the American movement of "fun at work" in the academic management literature was also a form of work's gamification [15].

Yu-Kai Chou [16] defines gamification as the craft of deriving fun and engaging elements found typically in games and thoughtfully applying them to real-world or productive activities. He calls this process Human-Focused Design because it concentrates on humans' feelings, motivations, and engagement in the experience.

Unfortunately, the term "gamification" remains inconsistently used, and a general theory of gamification is yet to be developed [17]. Aware of so many different definitions, discrepancies, distinctions, and discretionary delimitations, we use the original concept proposed by Deterding *et al.* [2]:

Gamification is the use of game design elements in non-game contexts.

To allow a more precise understanding of this definition, we define game design according to Brathwaite and Schreiber [20], i.e.: *"Game design is the process of creating the content and rules of a game. Good game design is the process of creating goals that a player feels motivated to reach and rules that a player must follow as he makes meaningful decisions in pursuit of those goals."*.

For game design elements, we use the definition by Deterding *et al.* [2] *"(...) elements that are characteristic to games – (...) that are found in most (but not necessarily all) games, readily associated with games, and found to play a significant role in gameplay."*

III. GAMIFICATION EFFECTS

Some scholars have a negative opinion on gamification. For example, Ian Bogost [21] stated that gamification is primarily an opportunistic marketing strategy. Despite such criticism, research over the years found that gamification does bring benefits when properly used.

Using an action research design, Putz *et al.* [22] found that gamification can have a positive effect on students' knowledge retention, independent of age and gender. Positive effects of gamification on enhancing interaction with learning materials and performance on studies were also reported by Klock *et al.* [23].

Systematic reviews have revealed both positive and negative effects. Zainuddin *et al.* [24] found evidence that the use of game design elements such as badges, points, trophies, leader boards, avatars, and virtual gifts not only promotes students' extrinsic motivation but also increases their intrinsic value for learning; however, they also reported studies with contradictory findings. Johnson *et al.* [25] conducted a systematic literature review of empirical studies on gamification for health and well-being. From the papers identified, the impact of gamified interventions was found to be positive by 59% of the papers reviewed, with effects including empowerment, motivation, health monitoring, and more healthy habits taken. However, 41% - a significant portion of the studies - reported mixed or neutral effects.

Hamari *et al.* [26] corroborated the point about mixed effects. Most quantitative studies identified in their review concluded positive effects of gamification elements to exist only partially. Also, they observed (through qualitative analysis) that gamification is more manifold than previous studies often assumed. Koivisto and Hamari [27] reached the same conclusion, having found papers pointing to a mixed effect of gamification and a small amount of purely negative results, which they attributed to a possible confirmation bias.

Indeed, gamification manipulates human psychology through game design elements, and it is natural to expect that such manipulation can have adverse effects. One issue occurred at the Disneyland Resort Hotel, California, in 2018. The hotel decided to use leaderboards updated in real-time to stimulate its workers to clean sheets and towels more efficiently. The initiative backfired hard because the competition degenerated the environment's quality, caused extra stress, and increased the number of injuries on the job [28]. Another case happened with Go365, a gamified app, when imposed on public school teachers in West Virginia, forcing them to provide sensitive medical data and have their positions tracked daily. This gamified app's enforcement was the last in a series of events that lead to a wildcat strike in 2018 [28]. Recent research points out that a gamified activity should never be mandatory; otherwise, it loses its fun value and leads to the collapse of play and work value [29].

IV. RELATED WORK

The academic gamification research does not ignore game design elements causing negative effects. Algashami [30] cataloged various negative effects, which the author called "gamification risks", dividing them into five categories of risk factors: performance, societal & personal, goals, tasks, and gamification design. The author identified 20 gamification risks, amongst them: performance misjudgments, lowering self-esteem, counterproductive comparison, lack of group coherence, lack of engagement, reduce the quality, and kill the joy. In comparison with our work, Algashami's research [30] is not focused on gamification elements and neither on gamification applied to education, but on risk factors' identification and management strategies in large-scale businesses using gamification in their workplaces. Besides, the proposed taxonomy requires further improvement.

Hyrynsalmi *et al.* [31] pointed a lack of secondary studies about the negative effects of gamification. They categorized adverse gamification implications into limiting and harmful issues: the first category discusses gamification limiting the full capabilities of an artifact, and the second concerns the harmful consequences of gamification. We also found examples of both categories in [32]. In the mentioned context of a lack of secondary studies on the harmful effects of gamification, we decided to focus our research on them.

A. Secondary Studies

We searched for related secondary studies but noticed variations concerning our purpose. We found significant differences for at least one of the following: subject, data analysis, date range, or a lack of focus on the negative effects of game design elements in gamification.

Peixoto and Silva's review [33] had a different focus, aiming at building a gamification requirements catalog connecting game design elements to Bartle's Personality Types. Majuri *et al.* [34] present a review of 128 empirical

research papers on gamification of education and learning and point out an excessive focus on quantifiable performance metrics and positive aspects. However, their work is not focused on negative effects and only covers the literature until 2015. Also, Klock *et al.* [23] did also not focus on negative effects, besides having a data range limited from 2013 to 2016. Alhammad and Moreno's secondary study [35] had its scope limited to gamification in software engineering education.

Finally, the secondary study by Toda *et al.* [36] is the work closest to ours, as they also focus on negative effects. However, as we noticed a significant amount of work in recent years, and their study reports on only 17 papers within the date range from 2012 to the first half of 2016, we identified the need for an update [37]. In contrast with Toda *et al.* [36], we decided to run a new and complete mapping study to address our specific purpose more precisely, e.g., focusing directly on game design elements and identifying the type of empirical studies that revealed the negative effects.

V. Systematic Mapping

According to Kitchenham and Charters [38], a systematic mapping is "a broad review of primary studies in a specific topic area that aims to identify what evidence is available on the topic". We follow the procedures and guidelines described in [38] and [39], using a hybrid search strategy [14], combining database search with backward and forward snowballing.

A. The Research Questions

Our goal was to organize evidence regarding the negative effects of Game Design Elements (hereinafter referred to as GDEs) in the context of gamified education/learning systems. Therefore, we derived the following research questions:

RQ1 - *What game design elements cause which negative effects in the field of digital education/learning?*

RQ2 – *In what fields of digital education/learning were the negative effects of game design elements found?*

RQ3 – *Which types of empirical studies were conducted to assess the negative effects?*

To properly answer RQ1, we divided it into three more focused questions, organizing information on the GDEs causing negative effects (RQ1.A) and also on which negative effects affect which kind of user (RQ1.B and RQ1.C). Finally, we answer RQ1 by mapping the GDEs against their reported negative effects.

RQ1.A – *What game design elements caused negative effects in the field of digital education/learning?*

RQ1.B - *What negative effects of game design elements were found affecting those using the digital system or being in the role of a student?*

RQ1.C – *What negative effects of game design elements were found affecting those maintaining the digital system or being in the role of a teacher?*

B. Search Strategy

We decided to use a hybrid search strategy, combining a database search on Scopus with forward and backward snowballing [14]. Hybrid strategies were found to be capable of achieving an appropriate balance of precision and recall when looking for primary studies [14].

To design the search string for the database search on Scopus, we used the PICO (*Population, Intervention, Comparison, Outcome*) criteria [40] as follows:

- *Population*: gamification in the context of education/learning.
- *Intervention*: game design elements.
- *Comparison*: none.
- *Outcomes*: negative effects.

After that, we extracted the basic terms from the PICO criteria (gamification, education/learning, negative effects) and added synonyms and related terms. We decided not to include the intervention's terms, as we conducted the database search based on title, keywords, and abstract, where details on game design elements could have been omitted.

We added the following synonyms and related terms:

- Gamification: *gamify, gamified, gamifying.*
- Education/learning: *information, teaching, curriculum, pedagogy, didactics, training, instruction.*
- Negative: *damaging, prejudicious, prejudicial, detrimental, counterproductive, inappropriate, harmful, perilous, limiting.*

Finally, we applied AND and OR logic operators to connect the terms, resulting in the following search string:

(gamification OR gamify OR gamified OR gamifying) AND (education OR learning OR information OR teaching OR curriculum OR pedagogy OR didactics OR training OR instruction) AND (negative OR damaging OR prejudicious OR prejudicial OR detrimental OR counterproductive OR inappropriate OR harmful OR perilous OR limiting)

As snowballing support tool, we used Publish or Perish [41], a software program that allows retrieving academic citations using information from Scopus and Google Scholar.

C. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria can be found in Table I and II, respectively. The exclusion criteria also provide details on our three-phase filtering procedure. To organize and filter the documents throughout the systematic mapping, we used Rayyan, a free web application to support systematic review authors [42].

D. Applying the Search Strategy

We applied the search string on Scopus on July 28th, 2020, searching within the title, abstract, and keywords. It returned 180 documents, upon which we applied the exclusion criteria through three filtering phases, as described in Table III. After this initial filtering a set of 64 papers remained.

Thereafter, we conducted backward and forward snowballing using these 64 papers as seed set, both on August 18th, 2020. The papers retrieved from backward snowballing and from forward snowballing using Scopus citation information were merged with the seed set, resulting in 2338 unique entries. Additionally, considering that Mourão *et al.* [14] suggest using Google Scholar for forward snowballing, besides doing it using Scopus citation information, we also conducted forward snowballing using citation information from Google Scholar (on September 4th, 2020). The forward snowballing through Google Scholar found 738 additional

unique entries. Hence, we ended up with 3076 unique entries (including the seed set 64). We applied our inclusion and exclusion criteria to the title, abstract, and keywords of the remaining 3012 papers, as shown in Table IV.

After the title, abstract, and keyword filtering, we conducted full-text-based filtering for the remaining 140 papers. The result of this full-text-based filtering is shown in Table V, resulting in a set of 68 included papers. Out of those, 32 were found by the initial Scopus search, 18 by forward snowballing, 15 by backward snowballing, and 3 were retrieved by both forward and backward snowballing. These numbers also help to illustrate how snowballing can be complementary to database searches.

TABLE I. INCLUSION CRITERIA

Inclusion Criteria	Reasoning
IC1 - Papers that include negative effects of GDE applied in the field of education/learning	Research subject
IC2 – Papers that passed through peer review	To ensure a minimum level of quality
IC3 - Papers in English	Quality verifiable by other researchers

TABLE II. EXCLUSION CRITERIA OF THE THREE FILTER PHASES

Exclusion Criteria	Filter Phase	Reasoning
EC0 – Papers not in English	First Filter Phase	Quality not verifiable by other researchers
EC1 - Papers which were not about effects of GDE applied in the field of education/learning	First Filter Phase	Not about the research subject
EC2 – Duplicated papers	First Filter Phase	Duplicated
EC3 – Papers that did not report negative effects	Second Filter Phase	Not about the research subject
EC4 – Papers that have a more up to date version (*e.g.*, journal extension)	Second Filter Phase	Between two peer-reviewed versions reporting the same results, the most recent is to be used
EC5 - The paper is grey literature	Second Filter Phase	Typically not peer reviewed
EC6 - The paper represents a secondary or tertiary study	Third Filter Phase	Our study is a secondary study
EC7 - The paper is mainly about the non-digital use of GDE	Third Filter Phase	Focus of this paper is on digital artifacts
EC8 - The paper is a short paper (less than 4 pages)	Third Filter Phase	Typically does not represent complete research results
EC9 - The paper was inaccessible to the authors	Third Filter Phase	No means to access the paper
EC10 - Books and chapters are off	Third Filter Phase	Problems with verifying the quality

TABLE III. FILTERING AFTER SCOPUS DATABASE SEARCH RESULT OF 180 PAPERS, BASED ON TITLE, ABSTRACT, AND KEYWORDS

Removed because of	Amount
EC0	3
EC1	88
EC3	6
EC5	5
EC6	9
EC7	1
EC8	1
EC10	3
Remaining papers	64

TABLE IV. FILTERING OF 3012 UNIQUE ENTRIES RETRIEVED FROM SNOWBALLING

Removed because of	Amount
EC0	113
EC1	2192
EC2	83
EC3	28
EC4	1
EC5	27
EC6	177
EC7	1
EC8	6
EC9	20
EC10	288
Remaining articles (including the seed set)	140

TABLE V. FULL-TEXT-BASED FILTERING OF 140 PAPERS

Removed because of	Amount
No access (not even requesting authors)	6
EC1	45
EC3	13
EC6	5
EC7	1
EC8	1
EC10	2
Included papers	68

It is noteworthy that we conducted the full-text-based assessment only after snowballing on purpose, as we thought that applying snowballing on some additional closely related papers would not be detrimental. Nevertheless, this decision indeed increased our snowballing effort.

Finally, to complement our search strategy, we compared our set of included papers against the 17 papers included by Toda *et al.* [36]. While our set of 68 papers to be included comprised 29 papers ranging from 2012 to 2016, only seven of them were also included by Toda *et al.* [36]. I.e., their search strategy did not retrieve 22 papers reporting negative effects of gamification in education/learning systems that were retrieved by our search strategy. On the other hand, our search strategy missed nine papers included in their mapping (the remaining one was retrieved but eliminated from our mapping for not being related to "digital" GDEs – EC7). As a result of this comparison, to present a mapping including all papers that we were aware of, we manually included the papers found by Toda *et al.* that were missed by our search strategy, ending up with a final set of 77 included papers.

The differences could be explained by using a different and independently elaborated search strategy. It is also noteworthy that we applied only one iteration of forward and backward snowballing. We are confident that most of the missed papers would also have been retrieved by subsequent snowballing iterations. For instance, we verified that five out of the nine missed ones would have been retrieved as part of the second iteration (i.e., they cited or were cited by papers retrieved through our first snowballing iteration).

In summary, the scope of our search strategy (3,192 articles) and the added value (we extended the previously mapped evidence from 17 to 77 articles) provide an unbiased and significant overview on the adverse effects of gamification in educational systems.

E. Data Extraction

We extracted data from the 77 included papers focusing on answering our research questions. We used Google Sheets

to organize the extracted data. The spreadsheet with all the extracted data is available in an online Zenodo open science repository (www.doi.org/10.5281/zenodo.4540909).

We answered RQ1 by extracting data for RQ1.A, RQ1.B, and RQ1.C and connecting the GDEs with their respective negative effects. For RQ1.A, we extracted the GDEs that were related to negative effects. For RQ1.B and RQ1.C, we respectively extracted the negative effects caused to main users (in this case, students) and those keeping the system working or in the position of teachers. We followed the open coding guidelines proposed in [43] to assign the text of the papers to design elements and negative effects. During this process, different terms perceived as related to the same element or effect were associated with a single code. In case of doubt concerning coding, discussions were held among the three authors of the present work.

To answer RQ2, we extracted the fields of education/learning where the gamified systems were used (e.g., computer science, medicine). Finally, to answer RQ3 we extracted the types of empirical studies conducted within each paper reporting the negative effects.

VI. Systematic Mapping Results

Figure 1 shows the distribution of the 77 included papers throughout the publication years (search strategy applied during the second semester of 2020). Results for each of our research questions based on the extracted data follow.

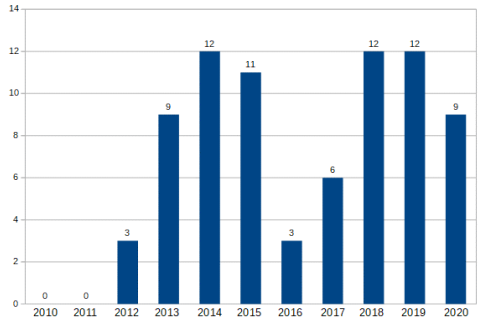

Fig. 1. Publication years of the included papers.

A. RQ1 - What game design elements cause which negative effects in the field of education/learning?

Overall, the papers reported 88 different GDEs, 64 different negative effects caused to the user, and ten different negative effects caused to the person maintaining the system or in the role of a teacher. We decided to conservatively ground the answer to this question on associations reported by more than one paper instead of correlation and causation, strengthening, in the process, our confidence in the results. We adopted this approach, considering the vast amount of GDEs and effects, plus the fact that many of the articles found had GDEs grouped and used together instead of individually. The complete extracted data, allowing different analyses, can be found in our open science repository.

RQ1.A – *What game design elements caused negative effects in the field of education/learning?* Table VI lists the GDEs mentioned by at least two papers and the number of papers that referred to each of them as causing negative effects. The list of papers referring to each element can be identified in the online repository. It is possible to observe that most of the reported negative effects were associated with the use of badges, competitions, leaderboards, and points. This makes sense given that these are GDEs commonly used in gamification, which may be related to creating competitive environments. It is also noteworthy that there were several (59) other GDEs, which had only one paper each indicating negative effects. Further analysis is required to answer whether this can be explained by the lack of negative effects caused by these elements or the lack of investigations involving them.

TABLE VI. Game Design Elements and Number of Papers Indicating Negative Effects

GDEs	#Papers	GDEs	#Papers
Badges	27	Teams	3
Competitions	22	Virtual Currency	3
Leaderboards	21	Collaboration	2
Points	17	Enjoyment	2
Challenges	11	Luck	2
Achievements	10	Progression	2
Quizzes	10	Quests	2
XP	10	Rankings	2
Levels	9	Rewards	2
Feedback	6	Score	2
Scoreboards	6	Simulations	2
Rewards	5	Skill trees	2
Goals	4	Tutorials	2
Avatars	3	Virtual goods	2
Immediate feedback	3	Others	59

RQ1.B - *What negative effects of game design elements were found affecting those using the system or being in the role of a student?* Table VII shows the negative effects caused to the user mentioned by at least two papers and the number of times that papers referenced those negative effects. It is possible to observe that the most cited negative effects concern the lack of effect, lack of understanding, irrelevance, motivational issues, and worsened performance. The ethical issue of cheating was another recurrent effect, usually motivated by creating competitive reward environments and/or systems with failures that enable users to easily score by cheating. Again, several (42) other negative effects caused to the user were mentioned only once.

TABLE VII. Negative Effects Caused to the User

Negative effects caused to the user	#Papers	Negative effects caused to the user	#Papers
Lack of effect	16	Dislike of gamification	3
Lack of understanding	9	Lack of improvement	3
Irrelevance	8	Time constrains	3
Lack of motivation	8	Dislike of competition	2
Demotivation	6	Discouragement	2
Loss of performance	6	Lack of flow	2
Cheating	5	Lack of granularity on grading	2
Gaming the system	5	Novelty effect	2
Reduction of intrinsic motivation	5	Perception of high workload	2
Alienation or confusion for non-gamers	3	Sabotaged cooperation	2
Anxiety	3	Unintentional sabotage of weaker students	2

Hence, the most common negative effect was that using the gamified system resulted in no difference when compared to not using the gamified system. Someone may argue that the negative effects characterized as being the "lack of" something are not negative, given that nothing bad effectively happened. However, for each of such neutral results to happen, gamification elements were designed and

implemented, requiring human effort, time, and money. It can also be argued that just because a literary study didn't found an effect, it doesn't mean that does not exist. We accept this argument, but prefer to err from excess than from fault removing that effect. Another one that needs explanation is the "Novelty effect", which is a negative effect, in the sense that potential positive effects may be temporary. I.e., as soon as the user's interest goes away, the positive effects will not apply anymore, and if they weren't present for enough time, they may not be enough in terms of cost/benefit.

RQ1.C – *Which negative effects of game design elements were found affecting those maintaining the system or being in the role of a teacher?* Table VIII shows the negative effects caused to those keeping the system or being in the position of a teacher, which were mentioned more than once and the number of times those negative effects were mentioned within the analyzed papers. It is possible to observe that the most common negative effects concern technical challenges and extra required effort or resources.

TABLE VIII. NEGATIVE EFFECTS CAUSED TO THE TEACHER/PERSON MAINTAINING THE SYSTEM

Negative effects caused to the teacher/person keeping the system	#Papers
general technical challenges (bugs, difficulties with the software/hardware)	7
extra resources needed (e.g., money, time, people, effort)	5
engineering problems with the LMS (Learning Management System) used	2

Technical challenges are a part of software engineering in general and may also appear in gamified systems. Extra human effort and resources needed typically appear as a negative effect when the gamified systems imply having to create additional content and taking care of additional tasks on top of the everyday tasks related to education. Finally, engineering problems typically appeared when learning management systems did not cover what the designers wanted them to do, leading to implementation workarounds and potentially lower quality. Seven other negative effects were cited only once.

To complete the answer to RQ1, we mapped the GDEs against the related negative effects. The bubble plot in Figure 2 shows the GDE and negative effect combinations that appeared more than once in our systematic mapping. This mapping can help raise gamification designers' awareness of potential undesired negative effects of GDEs on education/learning systems.

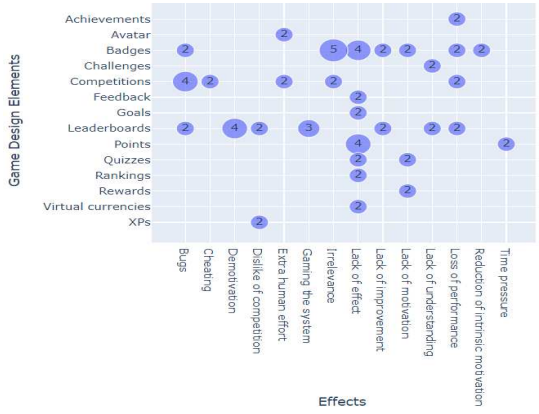

Fig. 2. GDEs and their reported negative effects.

It is possible to observe, for instance, that the use of *Badges* may be irrelevant [44][45][46][47][48], have no effect [49][50][51][52], not result in expected improvements [53][54], not be motivating [55][56], and even lead to losses in performance [57][58] or reduction of intrinsic motivation [59][60]. Furthermore, it may be related to technical problems [47][61].

Similar interpretations can be made for the remaining GDEs. We provide a table containing all the specific references related to the bubble plot of Figure 2 in our online repository due to space constraints. It is noteworthy that the primary studies included in our mapping vary in context and empirical strategy. Digging deeper into the strength of the empirical evidence and the specific contexts in which the negative effects of the GDEs were observed would require analyzing the primary studies beyond the typical scope of mapping studies. It may be part of a more detailed review extending this paper.

B. *RQ2 – In what fields of education/learning were the negative effects of game design elements found?*

The fields where negative effects of GDEs were reported more than once are shown in Table IX. Besides the listed ones, there were 29 other fields reported that were cited only once. It can be observed that the negative effects were reported in several different areas. Given the closeness between games, gamification, and digital technology, computer science being the most covered subject was expected.

TABLE IX. FIELDS OF EDUCATION/LEARNING WHERE NEGATIVE EFFECTS OF GAME DESIGN ELEMENTS WERE FOUND

Fields	#Papers
Computer Science	27
Medicine	6
Business	4
Education (e.g., Pedagogy)	3
Mathematics	3
University courses	3
Communication	2
Game development	3
Multimedia content and production	2
Seventh grade (middle school)	2

C. *RQ3 – What types of empirical studies were conducted to assess the negative effects?*

To answer this question, we used Wohlin's classification [64], which divides empirical strategies into surveys, case studies, and controlled experiments. As shown in Table X, most of the research was reported as concerning case studies or controlled experiments, complemented by surveys. The positive aspect is that all papers reported applying at least one empirical strategy.

TABLE X. EMPIRICAL STUDIES CONDUCTED TO ASSESS THE NEGATIVE EFFECTS (COUNTING MULTI-TYPES AS DIFFERENT ENTRIES)

Types of empirical studies	Amount
Case Study & Survey	31
Controlled Experiment & Survey	28
Controlled Experiment	9
Survey	5
Case Study	4

VII. LIMITATIONS

We applied the search string on Scopus on July 28th, 2020, followed by single iteration snowballing searches conducted

on August 18th, 2020 (backward and forward using Scopus citation information) and September 4th, 2020 (complementing forward snowballing using Google Scholar citation information). Because of that, any papers published in 2020 after those dates were not retrieved.

Another limitation concerns potentially missing papers. After analyzing 3192 papers (see list in our online repository), based on our inclusion and exclusion criteria, we initially included 68 papers. While our search strategy allowed identifying significantly more papers than the database search strategy employed by Toda *et al.* [36] (e.g., 29 published between 2012 and 2016 against 17), the sets had differences. Therefore, we manually included the papers found by Toda *et al.,* which were missed by our search strategy, ending up with a final set of 77 papers.

It is noteworthy that we verified that most of the missed papers would also have been retrieved by subsequent snowballing iterations. Nevertheless, an extension applying subsequent snowballing iterations and investigating different hybrid strategies [14] would require significant additional effort, which is planned as part of a journal extension. We are confident that our final set of included papers as part of this publication already allowed providing an unbiased and meaningful overview of the adverse effects of GDEs in gamified education systems.

Another risk of false negatives concerns the filtering process. We screened all papers considering only titles, abstracts, and keywords, which may not contain sufficient information to decide upon inclusion. We avoided applying EC1 and EC3 during the initial screening to lower this risk, only excluding papers that we had high confidence of not investigating GDE effects (EC1) and reporting negative effects (EC3). In case of any doubt, the paper was left for full-text-based assessment. Moreover, the application of the inclusion and exclusion criteria made by the first author was reviewed by the second and third authors in meetings. In case of doubt, during the initial screening or full-text-based assessment, discussions were held to reach a consensus.

Furthermore, we chose not to consider grey literature as part of our inclusion criteria. There is the possibility of relevant grey literature that does not have equivalent in non-grey academic papers. On the other hand, even though we did not explicitly evaluate the strength of evidence, the results herein reported are based on peer-reviewed research and backed by empirical studies.

Finally, while research that reaches negative results is essential because it shows us what does not work [62], there still seems to be a publication bias towards positive outcomes. Research reporting negative results tends to have less scientific interest, fewer citations, and to be less often published [63]. Hence, there may be additional negative results that were not published and which, for that reason, could not be included in our mapping study.

VIII. Concluding Remarks

Gamification, when properly applied, can have positive effects on education/learning systems. However, gamified systems are also prone to generate harmful effects.

This paper reported a systematic mapping of negative effects of GDEs on digital education/learning systems. Based on data extracted from 77 identified papers, we provided a comprehensive overview with valuable information for software engineers and designers of such systems. For instance, we identified the GDEs that have most often been reported to cause adverse effects, the most common negative effects, and the relation between GDEs and negative effects. That can help gamification designers to consider potential negative effects when selecting GDEs. Also, researchers can benefit from the overview of available evidence to identify topics on which more primary studies should be conducted

IX. Funding

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel (CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brazil, and by the National Council for Scientific and Technological Development (CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil - Finance Code 001.

References

[1] Duolingo, "About Us: Mission", Who We Are. https://www.duolingo.com/info (accessed Sep 2020).

[2] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From Game Design Elements to Gamefulness: Defining Gamification", In Proceedings of the 15th International Academic MindTrek Conference, Tampere, Pirkanmaa, Finland, pp. 9-15, 2011.

[3] A. Rapp, F. Hopfgartner, J. Hamari, C. Linehan, F. Cena, "Strengthening gamification studies: Current trends and future opportunities of gamification research", Int. J. Human Computer Studies, 127, 1-6, 2019.

[4] Orbis Research, Gamification Market by Deployment (On premises, Cloud based, and Hybrid), Enterprise Size (Small and Medium Enterprises, and Large Enterprises), End User, and Geography (North America, Europe, APAC, and Row)-Global Forecast up to 2025, 2019.

[5] L. Piras, E. Paja, P. Giorgini *et al*., "Gamification Solutions for Software Acceptance: A Comparative Study of Requirements Engineering and Organizational Behavior Techniques", In Proceedings of the 11th RCIS, 2017.

[6] Volkswagen. "The Fun Theory 1 – Piano Staircase Initiative". Youtube https://www.youtube.com/watch?v=SByymar3bds. (accessed Aug 2020)

[7] Volkswagen. "The Fun Theory 2 – an initiative of Volkswagen: The World's Deepest Bin". Youtube https://www.youtube.com/watch?v=qRgWttqFKu8. (accessed Aug 2020)

[8] Volkswagen. "The Fun Theory 3 – an initiative of Volkswagen: Bottle Bank Arcade Machine". Youtube https://www.youtube.com/watch?v=zCt_MzsnIUk. (accessed Aug 2020)

[9] B. Morschheuser, K. Werder, J. Hamari and J. Abe, "How to Gamify? A Method For Designing Gamification", In Proceedings of the HICSS 2017, 2017.

[10] D.M. Fernández, S.Wagner, M. Kalinowski *et al.,* "Naming the pain in requirements engineering", Empirical Software Engineering, vol. 22 (5), 2298-2338, 2017.

[11] B. Boehm, V. Basili, "Software Defect Reduction Top 10 List" in Foundations of Empirical Software Engineering, Berlin: Springer Berlin Heidelberg, 426-431, 2005.

[12] T. Menzies, W. Nichols, F. Shull, L. Layman, "Are delayed issues harder to resolve? Revisiting cost-to-fix of defects throughout the lifecycle", Empirical Software Engineering, 22, 4, 1903-1935, 2016.

[13] United Nations. "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all", United Nations, https://unstats.un.org/sdgs/report/2019/goal-04 (accessed Mar 2021).

[14] E. Mourão, J.F. Pimentel, L. Murta, M. Kalinowski, E. Mendes and C. Wohlin, "On the performance of hybrid search strategies for systematic literature reviews in software engineering", Information and Software Technology, 123, 106294, 2020.

[15] M. Nelson, "Soviet and American precursors to the gamification of work", AcademicMindTrek'12, Tampere, Pirkanmaa, Finland, 2012.

[16] Y. Chou, Actionable Gamification: Beyond Points, Badges and Leaderboards, 2019th edition, Milpitas, CA, USA: Octalysis Media, 2019.

[17] K. Seaborn and Fels, D. “Gamification in theory and action: A survey”, Int. J. Human Computer Studies, 74, 14-31, 2015.

[18] Molleindustria. “Game definitions”, The Definition of Game, http://gamedefinitions.com/ (accessed Sep 2020).

[19] J. McGonigal, Reality Is Broken: Why Games Make Us Better and How They Can Change the World, 1st edition, New York, NY, USA: Penguin Group (USA), 2011.

[20] B. Brathwaite, I. Schreiber, Challenges for Game Designers, 1st edition, Boston, MA, USA: Course Technology, 2009.

[21] I. Bogost, “Why Gamification is Bullshit”, In: S. P. Walz and S. Deterding (eds.), The Gameful World: Approaches, Issues, Applications, The MIT Press, pp. 65-78, 2015.

[22] L. Putz, F. Hofbauer and H. Treiblmaier, “Can gamification help to improve education? Findings from a longitudinal study”, Computers in Human Behavior, 110, 106392, 2020.

[23] A. Klock, I. Gasparini, A. Ogawa and M. Pimenta, “Does gamification matter? A systematic mapping about the evaluation of gamification in educational environments”, In Proceedings of the SAC 2018, 2018.

[24] Z. Zaimudin, S. Chu, M. Shujahat and C. Perera, “The impact of gamification on learning and instruction: A systematic review of empirical evidence”, Educational Research Review, 30, 2020.

[25] D. Johnson, S. Deterding, K. Kuhn *et al.,* “Gamification for health and well-being: A systematic review of the literature”, Internet Interventions, 6, 89-106, 2016.

[26] J. Hamari, J. Koivisto and H. Sarsa, “Does Gamification Work? — A Literature Review of Empirical Studies on Gamification”, In Proceedings of the 47th HICSS, Waikoloa, Hawaii, USA, 2014.

[27] J. Koivisto and J. Hamari, “The rise of the motivational information systems: A review of gamification research”, International Journal of Information Management, 45, 191, 2019.

[28] V. Gabrielle, “The Dark Side of Gamifying Work”, Fast Company, (https://www.fastcompany.com/90260703/the-dark-side-of-gamifying-work) (accessed in Dec. 2020)

[29] W. Hammedi, T. Leclercq, I. Poncin, and L. Alkire, Uncovering the dark side of gamification at work: Impacts on engagement and well-being, Journal of Business Research, 122, pp. 256-269, 2021.

[30] A. Algashami, “Gamification Risks in Collaborative Information Systems: Identification and Management Method”, Ph.D. dissertation, Dept. of Computing, BU, Bournemouth, Dorset, 2019.

[31] S. Hyrynsalmi, J. Smed, and K. Kimppa. “The Dark Side of Gamification: How We Should Stop Worrying and Study also the Negative Impacts of Bringing Game Design Elements to Everywhere” In Proceedings of the GamiFIN 2017, 2017.

[32] S. Diefenbach, A. Müssig, “Counterproductive effects of gamification: An analysis on the example of the gamified task manager Habitica”, Int. J. Human Computer Studies, 127, 190-210, 2019.

[33] M. Peixoto, C. Silva, “A Gamification Requirements Catalog for Educational Software: Results from a Systematic Literature Review and a Survey with Experts”, In Proceedings of the SAC’17, 2017.

[34] J. Majuri, J. Koivisto, J. Hamari, “Gamification of education and learning: A review of empirical literature”, In Proceedings of the GamiFIN 2018, Pori, Satakunta, Finland, 2018.

[35] M Alhammad, A. Moreno, “Gamification in software engineering education: A systematic mapping”, Journal of Systems and Software, 141, 131-150, 2018.

[36] A. Toda, P. Valle, S. Isotani, “The dark side of gamification: An overview of negative effects of gamification in education”, In Proceedings of the HEFA 2017 , 2017.

[37] E. Mendes, C. Wohlin, K. Felizardo, M. Kalinowski, “When to Update Systematic Literature Reviews in Software Engineering”, Journal of Systems and Software, 167, 1-24, Sep. 2020.

[38] B. Kitchenham and S. Charters, “Guidelines For Performing Systematic Literature Reviews In Software Engineering - Technical Report (EBSE 2007-001)”, Keele University & University of Durham, 2007.

[39] K. Petersen, S. Vakkalanka and L. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update”, Information and Software Technology, 64, pp.1-18, 2015.

[40] L. Roever, “PICO: Model for Clinical Questions”, Evidence Based Medicine and Practice, 4, 2018.

[41] A. Harzing, “Publish or Perish”, Harzing.com – Research in International Management, https://harzing.com/resources/publish-or-perish. (accessed in Jul 2020)

[42] Qatar Computing Research Institute (Data Analytics), “Rayyan”, Rayyan QCRI, the Systematic Reviews web app, https://rayyan.qcri.org/welcome (accessed in Aug 2020)

[43] J. Saldaña, The coding manual for qualitative researchers, 4th Edition, SAGE Publications Limited, 2021.

[44] C. Pilkington, “A Playful Approach to Fostering Motivation in a Distance Education Computer Programming Course: Behaviour Change and Student Perceptions”, The International Review of Research in Open and Distributed Learning, 19, 3, 2018.

[45] R. Roy, S. Deterding and B. Zaman, “Collecting Pokémon or receiving rewards? How people functionalise badges in gamified online learning environments in the wild”, Int. J. Human Computer Studies, 127, 62-80, 2019.

[46] L. Hakulinen, T. Auvinen and A. Korhonen, “Empirical study on the effect of achievement badges in TRAKLA2 online learning environment”, In Proceedings of the LATICE '14, 2013.

[47] L. Haaranen, P. Ihantola, L. Hakulinen and A. Korhonen, “How (not) to introduce badges to online exercises”, In Proceedings of the SIGCSE’14, Atlanta, Georgia, USA, 2014.

[48] M. Ohn, K.-M. Ohn, “An evaluation study on gamified online learning experiences and its acceptance among medical students.”, Tzu Chi Medical Journal, 32, 2, 211-215, 2019.

[49] B. J. Morris, C. Dragovich, R. Todaro, S. Balci and E. Dalton, “Comparing badges and learning goals in low-and high-stakes learning contexts”, Journal of Computing in Higher Educ., 31, 573-603, 2019.

[50] G. Barata, S. Gama, J.Jorge, D.Gonçalves, “So fun it hurts -Gamifying an engineering course”, In International Conference on Augmented Cognition AC’2013, Las Vegas, Nevada, USA, 2013.

[51] L. Z. Pedro, A. M. Z. Lopes, B. G. Prates, J. Vassileva and S. Isotani, “Does gamification work for boys and girls? An exploratory study with a virtual learning environment”, In Proceedings of the SAC’15, Salamanca, Spain, 2015.

[52] M. Ortiz-Rojas, K. Chiluiza and M. Valcke, “Gamification in Computer Programming: Effects on learning, engagement, self-efficacy and intrinsic motivation”, In Proceedings of the ECGBL 2017, Graz, Austria, 2017.

[53] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, “Improving participation and learning with gamification”, In Proceedings of the Gamification '13, 2013.

[54] W. Ghaban and R. Hendley, “Understanding the Effect of Gamification on Learners with Different Personalities”, In Proceedings of the CSEDU 2019, Heraklion, Crete, Greece, 2019.

[55] K. Davis, H. Sridharan, L. Koepke, S. Singh and R. Boiko, “Learning and engagement in a gamified course: Investigating the effects of student characteristics”, Jour. Comp. Assist. Learn, 34, 5, 1-12, 2018.

[56] K. Davis and S. Singh, “Digital badges in afterschool learning: Documenting the perspectives and experiences of students and educators”, Computer & Education, 88, 72-83, 2015.

[57] L. de-Marcos, A. Domínguez, J. Saenz-de-Navarrete and C. Pagés, “An empirical study comparing gamification and social networking on e-learning” , Computer & Education, 75, 82-91, 2014.

[58] V. Naik and V. Kamat, “Adaptive and Gamified Learning Environment (AGLE)”, In Proceedings of the 7th T4E, , 2015.

[59] M. Hanus and J. Fox, “Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance”, Computer & Education, 80, 152-161, 2015.

[60] L. Facey-Shaw, M. Specht, P. Rosmalen and J. Bartley-Bryan, “Do Badges Affect Intrinsic Motivation in Introductory Programming Students?”, Simulation & Gaming, 51, 1, 2020.

[61] A. Campos, E. Batista, A. Signoretti, R. Gardiman and C. Madeira, “Gamifying activities in a higher education course”, In Proceedings of the 9th ECGBL, 2015.

[62] J. A. Teixeira da Silva, “Negative results: negative perceptions limit their potential for increasing reproducibility”, Journal of Negative Results in Biomedicine, 14, 1, 12, July, 2015.

[63] N. Matosin, E. Frank, M. Engel, J. Lum and K. Newell, “Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture”, Disease Models and Mechanisms, 7, 171-173, 2014.

[64] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, and A. Wesslén, Experimentation in software engineering. Springer Science & Business Media, 2012.