information

Article

GATUGU: Six Perspectives of Evaluation of Gamified Systems

Jakub Swacha [1,*], Ricardo Queirós [2,3] and José Carlos Paiva [2,4]

[1] Department of Information Technology in Management, University of Szczecin, 71-004 Szczecin, Poland
[2] CRACS—INESC Porto LA, 4169-007 Porto, Portugal
[3] uniMAD—ESMAD, Polytechnic of Porto, 4480-876 Vila do Conde, Portugal
[4] Department of Computer Science, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal
* Correspondence: jakub.swacha@usz.edu.pl

Abstract: As gamification spreads to new areas, new applications are being developed and the interest in evaluating gamified systems continues to grow. To date, however, no one has comprehensively approached this topic: multiple evaluation dimensions and measures have been proposed and applied without any effort to organize them into a full gamut of tools for the multi-dimensional evaluation of gamified systems. This paper addresses this gap by proposing GATUGU, a set of six perspectives of evaluation of gamified systems: General effects of gamification, Area-specific effects of gamification, Technical quality of gamified systems, Use of gamified systems, Gamefulness of gamified systems, and User experience of gamified systems. For each perspective, GATUGU indicates the relevant dimensions of evaluation, and, for each dimension, one measure is suggested. GATUGU does not introduce any new measurement tools but merely recommends one of the available tools for each dimension, considering their popularity and ease of use. GATUGU can guide researchers in selecting gamification system evaluation perspectives and dimensions and in finding adequate measurement tools. Thanks to conforming to GATUGU, the published gamification system evaluation results will become easier to compare and to perform various kinds of meta-analyses on them.

Keywords: gamification assessment; gamification evaluation; evaluation methods

Citation: Swacha, J.; Queirós, R.; Paiva, J.C. GATUGU: Six Perspectives of Evaluation of Gamified Systems. *Information* **2023**, *14*, 136. https://doi.org/10.3390/info14020136

Academic Editor: Gennady Agre

Received: 1 December 2022
Revised: 10 January 2023
Accepted: 15 February 2023
Published: 19 February 2023

1. Introduction

In recent years, there has been a meteoric growth in the gaming market. In fact, there are several factors that have contributed to this growth, such as: the increased proliferation of mobile devices, the easy availability of games on the Internet, and innovations in software and, mainly, in hardware, with the advent of consoles and of advanced graphics engines [1].

Given their popularity, games and gameful approaches began to be applied to solve problems and create better experiences in disparate domains such as sports, social networks, sustainability, e-learning, finance and shopping, productivity, etc. In this context, gamification has been used in contexts other than games, in order to change user behaviors into alignment with organizational goals.

Typically, gamification is applied in two contexts: organizational and educational. In the first, gamification is used in several areas, including job search, recruitment, onboarding, consolidation of work teams, corporate training, productivity, marketing, and sales. In the second, gamification is used in educational institutions to help students in the teaching and learning process and promote their involvement to learn in a natural and fun way.

However, there is always a reverse side of the medal. When misapplied, gamification can lead to distraction, addiction, frustration, and boredom [2]. The problem is often due to the inappropriate use of its elements that neglect the types of users and that lead to the exaggerated use of PBL (Points, Badges, and Leaderboards) that only enhances the extrinsic motivation (easily achieved but volatile) instead of the intrinsic motivation (more complex to achieve but longer lasting) [3].

Taking into account the formalization of the design process of a gamified strategy, there is no consensus on the use of formal processes. In fact, several researchers [4] claim that the design of games or gamified environments are too complex activities to be reduced to a formal procedure. Other authors [5] consider that the design process should be treated as an agile process that does not always follow a specific formality. Despite the existence of dozens of frameworks around the world, several researchers [6,7] claim that gamification as an academic topic is still quite young and only a few well-established models and frameworks can be useful. In this realm, the most well-known frameworks are MDA (Mechanics, Dynamics, and Aesthetics) [8], Octalysis [9], GAME (Gather, Act, Measure, and Enrich) [10], and 6D [11]. The 6D Framework [11] is one of the most consensual frameworks used mainly for organizational environments. The framework organizes its methodology in six steps: goals definition, target group behaviour definition, user types identification, activity cycles planning, fun highlighting, and appropriate tools selection. From these six steps and, apart from the user types identification, the last one plays a central role in the gamification strategy where the gamification designer should decide which tools should be used, depending, much of the time, on the domain and the type of objectives to be achieved. From the panoply of types of tools, common solutions are learning platforms [12,13], Customer Relationship Management (CRM) [14], Enterprise Resource Planning (ERP) [15–17], Knowledge Management Systems (KMS) [18,19], Mobile Apps [20–22], and Web Pages [23,24].

Evaluation is considered as one of the main components or phases of the design and development process of gamified systems [25–29]. In practice, however, gamified systems are often not evaluated at all. For instance, only 63% of papers focused at gamification in the context of software engineering analyzed in [30] included evaluation; this ratio, while much better than the appalling 11% reported in [31] where educational gamification papers were considered, is still far from impressive.

One of the reasons of this situation may be that, after over ten years having passed since the confluence of web technologies, digital business models, and online and location-based gaming paving the way for the emergence of gamified systems [32], there is still no generally agreed framework for their evaluation. While various methods have been proposed in prior studies (see, e.g., [28,33,34]), the authors are not aware of one that would approach the problem of gamified systems evaluation in a comprehensive way, i.e., one covering all key perspectives of evaluation.

The aim of this paper is to fulfill this gap by proposing GATUGU, a set of six perspectives of evaluation of gamified systems. Its name is an acronym of the names of the covered perspectives, i.e., General effects of gamification, Area-specific effects of gamification, Technical quality of gamified systems, Use of gamified systems, Gamefulness of gamified systems, and User experience of gamified systems. For each perspective, we define the relevant dimensions of evaluation, and, for each dimension, we propose one measure capable of its measurement.

The paper is structured as follows. In the following section, we briefly review the methods of evaluation of gamified systems proposed in prior studies. In Section 4, we explain our approach as well as provide arguments for the proposed selection of perspectives and dimensions within them. In the subsequent section, we describe the components of GATUGU, suggesting measures and measurement tools relevant to the evaluation of respective dimensions of each perspective. In Section 5, we try to map the relevance of respective evaluation perspectives to popular types of gamified systems. Next, we present the results of a survey among gamification practitioners who were asked to evaluate the appropriateness of the measures selected for the respective evaluation dimensions. The final section discusses the advantages and limitations of GATUGU, and concludes the paper.

2. Methods of Evaluation of Gamified Systems

There are various ways of evaluating gamified systems. Monteiro et al. identified 21 different evaluation dimensions used in the analyzed 100 studies on gamification in

software engineering [30]. Ranked from the most frequent, these were: "engagement" (30 studies), "performance" (29 studies), "satisfaction" (28 studies), "motivation" (19 studies), "usability" (9 studies), "learning" (6 studies), "utility" (5 studies), "efficiency" (5 studies), "perception" (4 studies), "fun" (3 studies), and, finally, "communication", "user experience", "acceptance", "involvement", "sociability", "kindness", "confidence", "flow", "gameplay", "awareness", and "consciousness" (all used in up to 2 studies). In their literature review on gamification in education [35], Inocencio followed the distinction introduced by Liu et al. [36] of instrumental (i.e., relevant to the goals specific for the context for which the system was developed) and experiential (i.e., relevant to how the system is experienced by its users) outcomes of gamification. In the first group, he identified 45 measured constructs that he assigned to four dimensions: Participation based on user activity (including, e.g., Behavioral Engagement, User Activity, and Participation), Performance based on self-perception (including, e.g., Meaningful Learning, Perceived Achievement, and Declarative Knowledge), Performance based on student grades (including, e.g., Learning Achievement, Learning Performance, and Retention of Knowledge), and Retention based on user activity (including, e.g., Attrition, Retention, and Persistence). In the second group, he identified 31 measured constructs, the most frequently used of which were: Perceptions (experience in general), Motivation Attitude (toward gamification, lessons, and badges, in general), Satisfaction Engagement (emotional and cognitive), Self-Efficacy, and Flow/Cognitive Absorption.

There were few attempts to guide gamification designers among this multiplicity of options. Gasca-Hurtado et al. describe a multi-layer assessment framework for gamified environments whose last element is the "environment assessment layer" aimed at obtaining user data to measure changes in the behaviour of gamified environment users [33]. They suggest the use of neuroscience equipment to collect data on user behaviour and then analyze these data to measure the effectiveness of gamification in improving motivation, commitment, and behaviour changes of users. Due to the required neuroscience equipment, this method is hardly suitable for a wide use.

Dal Sasso et al. [28] propose a set of five methods for the evaluation of gamification:

1. *Success metrics* consist of defining a set of goals at design time and verifying them after the system has been in production for a while. Dal Sasso et al. only provide two exemplary metrics (number of new users in the last month, average activity increase per user).
2. *Analytics* consist of measuring the number of active users and calculating the ratio of Daily Active Users (DAU, the number of unique users that interact with the software during a day) to Monthly Active Users (MAU, the average number of unique users that interacted with the software in the previous 30 days) to observe the trend of usage of the software in any given moment.
3. *Conflicts* consist of listing and prioritizing the conflicts between gamification elements and other elements of the information system to identify gamification elements that should be pulled out of the environment.
4. *Jen ratio* consists of computing the ratio of total positive interactions among users (e.g., virtual gifts and acknowledgments) over the total negative interactions (e.g., misbehavior and rude comments) in a given period of time and context to assess how positive the attitude of the users is.
5. *Survey* consists of surveying selected users of all expertise levels to quantify how the use of gamification impacted the activity of the users and if it brought actual benefits.

While the multi-aspectuality of Dal Sasso et al.'s proposal is appreciated, it is far from comprehensive; moreover, three of the component methods (Analytics, Conflicts, and Jen ratio) precisely address the defined evaluation dimensions, whereas the other two (Survey and Success metrics) are general, and can thus be used to measure almost anything.

Fitz–Walter [29] introduces a 9-step gamification design process whose steps from 7 to 9 are dedicated to evaluation, including, respectively:

- (Step 7) Identification of any gamification-related problems using heuristic evaluation;

- (Step 8) Evaluation of the software design as both a tool (considering aspects such as utility, usability, and user experience) and game (measuring the fun and enjoyment of players by playtesting, as well as checking if the gamification experience is having the desired effect (with regard to, e.g., concentration, user enjoyment, perceived control, user challenge, perceived usefulness, and perceived ease of use) using field experiments that compare a gamified application to a non-gamified application;
- (Step 9) Balancing the tool and the game, using working prototypes to identify any issues.

Fitz–Walter's proposal is also not comprehensive; moreover, steps 7 and 9 are clearly aimed at obtaining suggestions for changes in the system rather than its evaluation as such.

The lack of a comprehensive method for the evaluation of gamified systems led to the development of GATUGU, which will be presented in the following section.

3. A Comprehensive Approach to Evaluation of Gamified Systems: GATUGU

3.1. Approach

In order to avoid mixing evaluation concepts of different levels (e.g., methods, measures, and measurement tools), which is quite a common shortcoming of the methods described in the previous section, GATUGU adopts a hierarchical approach.

On the most general level, we identify evaluation perspectives, denoting different aspects of interest relevant to gamified systems. After careful examination of both the literature and the authors' own experiences as gamified systems' designers, we have identified six such perspectives. We consider the obtained selection of evaluation perspectives comprehensive as it covers most of the gamified systems aspects evaluated in the literature. Nonetheless, the proposed selection is generic and could not be exhaustive. We are aware of the evaluation dimensions that do not closely fit any of the selected perspectives (e.g., user attachment). However, we consider them as niche topics requiring dedicated tools, whereas GATUGU aims to be a solution addressing most but not all kinds of evaluation needs. Limiting the scope of the set of covered perspectives (and, as will be explained below, the dimensions and measurement tools) is the primary of the efforts to address the current problem of the variety of measures, which causes the lack of comparability of reported evaluation results.

For each evaluation perspective, we propose between three and five relevant dimensions. Looking at the number of possible evaluation dimensions reported in the literature [30,35], it is obvious that, at this level, we no longer attempted to include the most of what could be measured, but tried to limit the number of covered dimensions, to increase the comparability of evaluation results obtained following GATUGU. The criteria we considered when selecting the dimensions were:

- Having available measurement tools, preferably validated;
- Being able to provide valuable information to stakeholders, which is easy to interpret by them (e.g., Valence lacks this trait);
- Ensuring clearly distinct character of dimensions included in a single perspective (for instance, the combination of Learning, Learning Gain, and Learning Achievement lacks this trait);
- For the sake of wider interpretability and applicability, we preferred more general dimensions over more specialized ones (e.g., we have included Engagement yet not Flow, which is a state of high engagement).

It is important to note here that GATUGU provides a set of perspectives, dimensions, and measures. The users of GATUGU are free to use any subset of them, based on their relevance to the application context of the evaluated gamified system (see Section 5), the goals of the evaluation, the availability of the required data (or the ability to obtain them), and their own preference; while the choice of considered perspectives is usually dictated by the first two reasons, the choice of evaluation dimensions is where the other two matter more.

On the lower level, for each evaluation dimension, we suggest a single measure and one measurement tool capable of measuring it. If, for a given dimension, both objective and subjective measures are available, the objective measures are preferred but only if their measurement does not require dedicated hardware or special measurement conditions. By limiting the number of suggested dimension measures to just one, we ensure the comparability of evaluation results in that the dimension for all who will follow our suggestion.

3.2. Selection of Evaluation Perspectives and Dimensions

The application of gamification is often motivated by the benefits it promises, such as building motivation or rising engagement. These benefits are general, i.e., they can be observed regardless of the context in which the gamified system is used. We therefore defined the first perspective as the General effects of gamification. We looked at existing gamification definitions (which there are many of) to identify such effects. In particular, two definitions earned our attention [37]:

- Brian Burke: "the use of game mechanics and experience design to digitally engage and motivate people to achieve their goals";
- Mark Schreiber: "use of game elements to increase engagement and make life and work more fun".

Consequently, three dimensions have been assigned to the first perspective: Engagement, Motivation, and Fun.

Gamification is nonetheless usually applied to attain goals specific to the area in which the gamified system operates. Hence, the second considered evaluation perspective is the Area-specific effects of gamification. As each area of such applications may have its own evaluation dimensions, which are non-applicable to other areas, in this perspective, we chose but one dimension for each area that we know as a frequent target of gamification: Learning [38–40], Work Performance [41–43], Sales Revenue [44,45], Social Relations [46,47], and Stress Reduction [48,49].

The gamified systems are also information systems, and, as such, they can be evaluated in terms of their technical quality. Therefore, the third perspective is the Technical quality of gamified systems. While ISO/IEC 9126 [50] defines a considerable number of potential dimensions, we decided to follow what the industry considers as important [51], and have chosen the following four: Performance, Reliability, Availability, and Scalability.

While the use of some gamified systems is compulsory (e.g., it is the primary tool which an employee uses in their everyday work), many are not. In such cases, the use of the system is a necessary requirement for (sometimes, even a sufficient proof of) its success. Hence, the fourth perspective is the Use of gamified systems. In this perspective, we include five dimensions: Adoption Rate, Frequency of Use, Session Length, Continued Use Intention, and User Retention. These capture the adoption, use, and retention phases of a gamified system.

While the gamified systems are tools, they can be viewed as games as well [29]. For this reason, their game-related aspects can (and should) also be evaluated. The fifth perspective is, therefore, the Gamefulness of gamified systems. As there are multiple dimensions in which games are evaluated [52], in this case, we tried to select those that are most relevant to the main design goals. For this reason, we omitted several of the dimensions discussed in [52] (e.g., Accomplishment or Guided) that can hardly be seen as the potential main design goals. Consequently, four dimensions were chosen, assessing whether a gamified system is visually attractive for its users, challenging to play, well-balanced, and immersive.

Back to the quality of gamified systems, not all of its aspects are of technical nature [50]. No less important is the experience of users. For this reason, the sixth and last perspective is the User-perceived quality of gamified systems. In it, we follow the proposal of [53] and consider the two dimensions defined in the Technology Acceptance Model (TAM) [54]: Perceived Usefulness and Perceived Ease of Use.

3.3. Collecting Aggregate Measurements for Respective Dimensions

In the next section, measures will be suggested for the respective GATUGU dimensions. These measures may either be objective (obtained from given formulas, adopting objectively measurable variables as their input parameters) or subjective (obtained from a survey among users, who self-report their feelings and/or opinions).

All the suggested subjective measures use some form of Likert-scale questions. Following [53], for all measures based on the Likert scale, if not specified otherwise, we recommend using the numbers (e.g., 1...7) as point labels (rather than statements such as *Agree...Disagree*).

The results in the Likert scale can be effectively visualized using a histogram as well as presented in a tabular form containing the shares of the respondents who answered positively (5–7) and/or the difference between the share of respondents who definitely agreed (only answering 6–7) and disagreed (1–2) ("Net Top Box").

In order to obtain an aggregate of such a measure, taking into consideration that the Likert scale is an ordinal scale, instead of calculating the arithmetic average of respective responses, we propose to use the arithmetic average of the Net Top Box percentages. A Net Top Box percentage $ntb(i)$ for a single item i can be calculated with the following formula:

$$ntb(i) = \frac{100\%}{|X|} \sum_{x \in X(i)} \begin{cases} 1 & \text{if } rc(x) > top - 2, \\ -1 & \text{if } rc(x) < 3, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where:

- i is the measured component item;
- $X(i)$ is the set of responses measuring component item i;
- top is the maximum response value (i.e., 5 for a 5-point Likert scale, and 7 for a 7-point Likert scale);
- x is a single response (provided as an integer between 1 and top);
- $rc(x)$ is the value of response x adjusted for reverse coding (see below).

As some measures include items that are reverse coded (i.e., the higher their value, the smaller the aggregate measure should be), all responses should be adjusted with the following formula:

$$rc(x) = \begin{cases} 1 + top - x & \text{if } x \text{ is reverse-coded,} \\ x & \text{otherwise,} \end{cases} \tag{2}$$

where x is the original measurement.

The aggregate measure $M(d)$ of dimension d can therefore be calculated with the following formula:

$$M(d) = \frac{1}{|I(d)|} \sum_{i \in I(d)} ntb(i), \tag{3}$$

where:

- d is the measured dimension;
- $I(d)$ is the set of items of which the measure of d is composed.

The interpretation of the measurements is straightforward:

- if the aggregate measure is a positive number, it indicates the prevalence of positive evaluation of a given dimension;
- if the aggregate measure is a negative number, it indicates the prevalence of negative evaluation of a given dimension;
- if the aggregate measure is zero, it indicates a balance between a positive and negative evaluation of a given dimension.

4. GATUGU Components

4.1. Evaluating General Effects of Gamification

4.1.1. Motivation

Motivation is defined as "a phenomenon that explains the processes giving rise to individual energy and endurance, and it is also a pathway towards a specific behavior" [55]. The motivation of gamified system users can be measured using a survey. There are various scales used to measure motivation, based on various theories of motivation (see the Appendix A to [35] for a list of six of them).

The measurement tool we suggest to use is SIMS (Situational Motivation Scale) proposed by Guay, Vallerand, and Blanchard [56], in particular, its two subscales, measuring, respectively, intrinsic motivation and identified regulation (extrinsic motivation)—as, in the context of evaluating gamification, we are not interested in measuring external regulation (when someone does what he/she is obliged to do) or amotivation (when someone does not see a reason for what he/she is doing). It is based on the Self-Determination Theory (SDT) [57] (p. 68), which is the one by far most often referred to in gamification research [58].

An important practical advantage is that it is a simple tool, as both subscales consist of just four items each, measured using a 7-point Likert scale, ranging from 1 (*does not correspond at all*) to 7 (*corresponds exactly*).

Table 1 lists the component items of the motivation measure. Note that *this activity* pertains to the activity being gamified (e.g., learning if the gamified system is an interactive learning environment).

Table 1. Measuring situational motivation of gamified system users.

Item	Why Are You Currently Engaged in This Activity?	Subscale
1	Because I think that this activity is interesting	Intrinsic
2	Because I think that this activity is pleasant	Intrinsic
3	Because this activity is fun	Intrinsic
4	Because I feel good when doing this activity	Intrinsic
5	Because I am doing it for my own good	Extrinsic
6	Because I think that this activity is good for me	Extrinsic
7	By personal decision	Extrinsic
8	Because I believe that this activity is important for me	Extrinsic

Source: [56].

4.1.2. Engagement

Engagement is defined as the investment of one's physical, cognitive, and emotional energies into the various activities to be performed [59]. Similar to motivation, engagement can be measured using a self-report questionnaire. We recommend using the instrument developed by Hamari et al. [60] (p. 178), which consists of eight items measured using a 5-point Likert scale and considers engagement from the perspective of the Flow theory [61], which is the second most often referred to in the gamification research after SDT [58].

Table 2 lists the component items of the engagement measure. Note that two of the eight items are reverse-coded—these items are marked with (R). Note that *what you were doing* pertains to the activity being gamified (e.g., learning if the gamified system is an interactive learning environment).

Table 2. Measuring engagement of gamified system users.

Item	Question
1	How hard were you concentrating?
2	It provided content that focused my attention
3	How much did you enjoy what you were doing?
4	Interacting with it was entertaining
5	Interacting with it was fun
6	How interesting was what you were doing?
7	Did you feel bored with what you were doing? (R)
8	Did you wish you were doing something else? (R)

Source: adapted from [60].

4.1.3. Fun

Researchers have several different conceptualizations of fun but none commits to a specific definition. Nonetheless, some studies attempted to define fun as the result of combining enjoyment, control of the content, care for the content, and the suspension of disbelief [62]. Within the entertainment area, fun is "pleasure without purpose" [63]. Others have split fun into five types, namely sociability, contentment, achievement, sensual, and ecstatic [64]. Various instruments have been proposed for measuring fun: while it can even be measured with the simple 1-item Smileyometer [65,66], here, we recommend using the FunQ questionnaire, comprising six subscales (Autonomy, Challenge, Delight, Immersion, Loss of Social Barriers, and Stress) and 18 items in total (measured using a 5-point Likert scale) [67].

Table 3 lists the component items of the fun measure. The reverse-coded items are marked with (R). When implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 3. Measuring fun perception of gamified system users.

Item	Question	Subscale
1	While using [*the gamified system*], I knew what to do.	Autonomy
2	I used [*the gamified system*] because I had to. (R)	
3	I used [*the gamified system*] because I wanted to.	
4	While using [*the gamified system*], I felt I was good at it.	Challenge
5	While using [*the gamified system*], I did something new.	
6	While using [*the gamified system*], I was curious.	
7	While using [*the gamified system*], I had fun.	Delight
8	I want to use [*the gamified system*] again.	
9	While using [*the gamified system*], I was happy.	
10	While using [*the gamified system*], I felt that time flew.	Immersion
11	While using [*the gamified system*], I forgot about school/work.	
12	While using [*the gamified system*], I felt good.	
13	While using [*the gamified system*], I made new friends.	Loss of Social Barriers
14	While using [*the gamified system*], I talked to others easier than usual.	
15	While using [*the gamified system*], I felt closer to others more than usual.	
16	While using [*the gamified system*], I felt bad. (R)	Stress
17	While using [*the gamified system*], I felt angry. (R)	
18	While using [*the gamified system*], I felt sad. (R)	

Source: adapted from [67].

4.2. Evaluating Area-Specific Effects of Gamification

4.2.1. Learning

Even though learning has virtually been a central topic in psychological research since psychology became an independent science and was even the most intensively studied

topic within psychology during a large part of the previous century [68,69], researchers are rarely explicit about what they mean by "learning". The most commonly found definitions in psychology textbooks are identical and refer to learning as "a relatively permanent change in behavior as a result of practice or experience" [70,71]. To measure learning, we suggest the use of the classic pretest–posttest approach, in which the student's knowledge improvement due to learning is measured as the difference between the arithmetic average of their percentage of correct answers on the pretest and the arithmetic average of their percentage of correct answers on the posttest, with both the pretest and posttest containing the questions relevant to the subject of learning [72]:

$$KI = 100\% \times \sum_{u \in U} \frac{Posttest(u) - Pretest(u)}{|U| \times |Test|} \tag{4}$$

where:

- $Test$ is the set of questions relevant to the subject of learning;
- U is the set of users whose knowledge improvement is measured;
- $Pretest(u)$ is the number of correct answers to the $Test$ questions provided by user u at the beginning of the period under evaluation;
- $Posttest(u)$ is the number of correct answers to the $Test$ questions provided by user u at the end of the period under evaluation.

Any positive value denotes a positive learning effect. The value of 100% denotes a perfect learning effect (the respondents failed to answer any test question before, and answered all of them after). No value larger than 100% is possible.

4.2.2. Work Performance

Work Performance relates to how well an individual performs a job, role, task, or responsibility [73]. We recommend using the ratio:

$$WP = \frac{Output}{t} \tag{5}$$

as the measure of Work Performance, where:

- $Output$ denotes the given job results of a user or a group of users, measured by the number of produced items or provided services, or the combined value of produced items/provided services, if the items are of notably different value,
- t denotes the time spent on the given job by this user or this group of users.

Work Performance measurements can only be interpreted relatively.

4.2.3. Sales Revenue

Sales Revenue encompasses the income received from the sales of goods or the provision of services [74]. We recommend the direct measurement of Sales Revenue in the respective currency where the sales occurred, or in a common currency if the products were sold in markets using various currencies.

Sales Revenue can only be interpreted relatively.

4.2.4. Social Relations

Positive social connections such as supportive interactions, the sense of belonging, and effective teamwork are important for health and well-being in the context of any activity, particularly those performed regularly such as work and studying [75,76]. We recommend the measurement of Social Relations with the following formula:

$$SocR = \frac{1}{|U| \times |t|} \times \sum_{u \in U} si(u), \tag{6}$$

where:

- U is the set of users whose social activity is measured;
- $si(u)$ is the number of social interactions of user u in the considered time period;
- t is the length of the considered time period.

The measurements of Social Relations can only be interpreted relatively.

4.2.5. Stress Reduction

According to the World Health Organization (WHO), stress "can be defined as any type of change that causes physical, emotional or psychological strain. Stress is the body's response to anything that requires attention or action" [77]. Reducing stress is, thus, beneficial to their performance in the job, role, task, or responsibility they conduct. Although symptoms of stress can be measured using various sensors [78], such sensors may not be available at the place of measurement, so we recommend the use of the Perceived Stress Questionnaire (PSQ), containing 30 questions using a 4-level scale (1–4) [79].

Table 4 lists the component items of the perceived stress. The reverse-coded items are marked with (R). In order to measure stress reduction, one has to calculate the difference:

$$StRed = \sum_{u \in U} (PS_{pre}(u) - PS_{post}(u)), \tag{7}$$

where:

- $PS_{pre}(u)$ is the perceived stress of user u measured at the beginning of the period under evaluation;
- $PS_{post}(u)$ is the perceived stress of user u measured at the end of the period under evaluation.

Any positive value denotes a positive effect of stress reduction; the value of zero denotes no effect of stress reduction, whereas any negative value denotes an increase in perceived stress.

4.3. Evaluating Technical Quality of Gamified Systems

4.3.1. Performance

As defined by ISO/IEC 25010:2011 for Systems and software Quality Requirements and Evaluation (SQuaRE) [80], performance, in terms of time, is the degree to which the response and processing times and throughput rates of a system, while performing its functions, meet the requirements. For measuring the performance, as processing gamification rules is the operation characteristic for most gamified systems, we recommend to use the number of processed gamification rules per second:

$$RPS = \frac{Rules}{t}, \tag{8}$$

where:

- $Rules$ is the number of processed gamification rules;
- t is the total processing time of the gamification rules in seconds.

The measurement can be performed in two ways: by an experiment in a controlled environment or by an analysis of data gathered from a system operating in the real world.

In the first case (see [81] for an exemplary analysis of this kind), the evaluator specifies the rules on which the system performance is evaluated. As rules may have various levels of complexity (compare, e.g., a rule triggered by clicking a single user interface component and a rule triggered by finishing a set of jobs each in a respective specified time), the measurement should comprise numerous test cases involving the triggering of different rules. Similarly, as for various reasons, the execution time of the same rule may vary, so the measurements should be repeated to obtain the average, minimum, and maximum processing times. The chosen number of measurement repetitions should depend on how often outlying measurements occur: typically, 10 is enough, but if, e.g., about every 20th

rule is processed much longer than average, the number of repetitions should be set to at least 40. Moreover, as many gamified systems have multiple concurrent users, and the number of rules triggered simultaneously greatly affects their processing time, we suggest repeating the tests for various numbers of simultaneously triggered rules. This number should mirror the expected number of concurrent users at different daily activity levels (e.g., 10 in early morning, 80 mid-day, and 40 in late afternoon). For the sake of comparability, it is necessary to report the configuration of the test platform, both hardware (e.g., the processor, memory, etc.) and software (e.g., operating system, database engine, used interpreters or compilers, etc.).

Table 4. Measuring perceived stress of gamified system users.

Item	Statement	Subscale
1	You feel that too many demands are being made on you	Harrassment
2	You find yourself in situations of conflict	
3	You are under pressure from other people	
4	You feel criticized or judged	
5	You have too many things to do	Overload
6	You have too many decisions to make	
7	You feel loaded down with responsibility	
8	You have enough time for yourself (R)	
9	You are irritable or grouchy	Irritability
10	You feel calm (R)	
11	You feel lonely or isolated	Lack of joy
12	You feel you are doing things you really like (R)	
13	You feel you are in a hurry	
14	You feel safe and protected (R)	
15	You enjoy yourself (R)	
16	You feel you are doing things because you have to	
17	You are lighthearted (R)	
18	You feel rested (R)	Fatigue
19	You feel tired	
20	You are full of energy (R)	
21	Your problems seem to be piling up	
22	You fear you may not manage to attain your goals	Worries
23	You have many worries	
24	You feel discouraged	
25	You are afraid for the future	
26	You feel under pressure from deadlines	
27	You feel frustrated	Tension
28	You feel tense	
29	You feel mentally exhausted	
30	You have trouble relaxing	

Source: [79].

In the second case, the system performance is evaluated considering the rules actually triggered by real-world users. System logs gathering rule start and end processing times can be maintained to obtain the data needed for the measurement. Various periods of time (hour, day, and week) could be considered to obtain performance data illustrating the varying level of activity in the system.

The main advantages of the first option are its feasibility already at an early stage of system prototype development and the possibility to check the processing times of complex rules that are rarely triggered in real use but do happen. The advantages of the second option are measuring the actual processing time as experienced by users, not what the users are expected to experience, and that it can be automatically obtained without any additional effort from the evaluator (such as preparing data and setting up environment for the experiments); only the provided necessary logs are maintained.

4.3.2. Reliability

Reliability is the degree to which a system, product, or component performs specified functions under specified conditions for a specified period of time, as in ISO/IEC 25010:2011 [80]. Although Reliability can be measured using various measures [82], we recommend using Mean Time Between Failures (MTBF), which is the average time between successive failures of the system [83]:

$$MTBF = \frac{Uptime}{Failures}, \tag{9}$$

where:

- $Uptime$ is the part of the analyzed time span in which the system was available for users;
- $Failures$ is the number of failures of the evaluated system that happened during the analyzed time span.

If the number of registered $Failures$ is zero then it can only be concluded that

$$MTBF > Uptime. \tag{10}$$

Obviously, the analyzed time span should be long enough for the potential failures to materialize.

4.3.3. Availability

According to ISO/IEC 25010:2011 [80], the Availability is the degree to which a system, product, or component is operational and accessible when required for use. The Availability can be measured with the ratio:

$$Avt = 100\% \times \frac{Uptime}{(Uptime + Downtime)}, \tag{11}$$

where

- $Uptime$ is the part of the analyzed time span in which the system was available for users;
- $Downtime$ is the part of the analyzed time span in which the system was not available [84] (p. 175).

The value of 100% denotes a perfectly available system (a value over 100% is not possible).

4.3.4. Scalability

The Scalability is "the ability of a system to accommodate an increasing number of elements or objects, to process growing volumes of work gracefully, and/or to be susceptible to enlargement" [85]. We recommend measuring the Scalability with the ratio:

$$Scl = 100\% \times \frac{LowLoadProcessingTime/LowLoadSize}{HighLoadProcessingTime/HighLoadSize}, \tag{12}$$

where:

- $LowLoadProcessingTime$ is the time spent processing a low number of concurrently triggered gamification rules (provided in $LowLoadSize$),
- $HighLoadProcessingTime$ is the time spent processing a high number of concurrently triggered gamification rules (provided in $HighLoadSize$),
- $HighLoadSize >> LowLoadSize$.

The value of 100% denotes a perfectly scalable system (a value over 100% is still possible, denoting a system that handles large loads relatively faster than small loads).

4.4. Evaluating Use of Gamified Systems

4.4.1. Adoption Rate

The Adoption Rate consists of the pace at which a new technology is acquired and used by the target public [86]. The Adoption Rate can be measured using the ratio:

$$AR = \frac{NewUsers}{t}, \tag{13}$$

where *NewUsers* is the number of users who began to use the system during the analyzed time span of length t.

4.4.2. Frequency of Use

The Frequency of Use (or usage frequency) consists of the number of times a system or application is used in a certain period of time. The Frequency of Use can be measured using the ratio:

$$FoU = \frac{UserLogins}{t}, \tag{14}$$

where *UserLogins* is the number of times users logged in to the system during the analyzed time span of length t.

4.4.3. Session Length

The Session Length is the amount of time from when a user launches the system or application until they stop using or leave it. The Average Session Length can be measured using the ratio:

$$ASL = \frac{TotalTimeOnline}{UserLogins}, \tag{15}$$

where:

- *TotalTimeOnline* is the total time spent by the users in the system during the analyzed time span;
- *UserLogins* is the number of times users logged in to the system during the analyzed time span.

4.4.4. Continued Use Intention

The Continued Use Intention measures one's intention to continually use or reuse a system or application [87]. To measure it, we recommend using the measurement scale developed by Bhattacherjee [88], which, according to [89] is the most widely adopted scale for measuring the Continued Use Intention. It is based on Kieran's behavioral intention scale [90] and consists of three items, one of which is reverse-coded, measured with a 5-point Likert scale.

Table 5 presents the three items that compose the scale for measuring the continued use intention of a gamified system. The reverse-coded item is marked with (R). When implementing the questionnaire, [the gamified system] should be replaced with the name of the gamified system under evaluation.

Table 5. Measuring the continued use intention of a gamified system.

Item	Statement
1	I intend to continue using [*the gamified system*] rather than discontinue its use.
2	My intentions are to continue using [*the gamified system*] rather than use any alternative means.
3	If I could, I would like to discontinue my use of [*the gamified system*]. (R)

Source: adapted from [91].

4.4.5. User Retention

The User Retention calculates the amount of first-time users of the system or application that return in subsequent time periods. The User Retention can be measured using the ratio:

$$UR = 100\% \times \frac{ReturningUsers}{NewUsers}, \tag{16}$$

where:

- *NewUsers* is the number of new users who registered during the analyzed period;
- *ReturningUsers* is the number of new users who used the system at least once after the initial logging in during the same analyzed period.

The value of 100% denotes a perfect user retention, with all newly registered users coming back to the system at least once (a value over 100% is not possible).

4.5. Evaluating Gamefulness of Gamified Systems

4.5.1. Visual Attractiveness

The look-and-feel of a system or an application is the first and arguably the major indicator of its quality for the user. To measure Visual Attractiveness, we recommend using the tool proposed by King et al. [91] that consists of four items referring, respectively, to the layout, choice of colors, overall look, and attracting user's attention, and uses a 5-point Likert scale.

Table 6 presents the four items that compose the scale for measuring the visual attractiveness of a gamified system. When implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 6. Measuring the visual attractiveness of a gamified system.

Item	Statement
1	The layout of the user interface of [*the gamified system*] is attractive.
2	The colors used on the user interface of [*the gamified system*] are attractive.
3	Overall, I think the user interface of [*the gamified system*] looks really good.
4	The user interface design of [*the gamified system*] caught my attention to explore it.

Source: adapted from [91].

4.5.2. Challenge

Being challenged is essential to achieve flow while playing a game [92]. The challenge is, therefore, part of the game experience, and many times even described as a dimension of the game experience [93,94]. When creating a gamified system, the ultimate goal is also to engage users while performing an activity. Hence, the users of a gamified system or application must also experience challenge.

This dimension can be measured using the Gameful Experience Questionnaire (GEQ) [52] that consists of eight items. Table 7 presents the eight-item scale adapted to gamified systems. When implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation, while any *n*-point Likert scale can be used for GEQ, we suggest using a seven-value Likert scale to obtain a finer grade of measurement detail.

Table 7. Measuring the challenge of a gamified system.

Item	Statement
1	Using [*the gamified system*] prompts me to push my limits.
2	Using [*the gamified system*] drives me in a good way to the brink of wanting to give up.
3	Using [*the gamified system*] pressures me in a positive way by its high demands.
4	[*The gamified system*] challenges me.
5	Using [*the gamified system*] calls for a lot of effort in order for me to be successful.
6	Using [*the gamified system*] motivates me to do things that feel highly demanding.
7	Using [*the gamified system*] causes me to feel like I continuously need to improve in order to do well.
8	Using [*the gamified system*] causes me to work at a level close to what I am capable of.

Source: adapted from [52].

4.5.3. Balance

A balanced game ensures the game's difficulty is adequate when compared to player's abilities, i.e., the player should be challenged but not blocked by the complexity of the game [95]. In the case of multiplayer games, balance also weighs the fairness of competition, particularly when multiple choices are available; all must be viable, consistent, and fair, not opening flaws that can be exploited to obtain advantages [96]. According to Newheiser [97], an unbalanced game design might undermine the entire ruleset of the game. To measure game Balance, we recommend the three-item scale presented in [98] (p. 44), using a seven-point Likert scale.

Table 8 presents the adapted three-item scale for the purpose of measuring the balance of a gamified system. When implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 8. Measuring the balance of a gamified system.

Item	Statement
1	In [*the gamified system*], lots of meaningful choices are presented to the users.
2	In [*the gamified system*], sufficient contexts are provided to allow each choice (that a user makes in the system) to have balanced advantages and limitations.
3	In [*the gamified system*], users with equal skills have a roughly same chance to succeed although they might start with different choices (e.g., options, characters, resources, etc).

Source: adapted from [98] (p. 44).

4.5.4. Immersion

Immersion, in games, is the act of causing the player to enter into a cognitive state where their reality mixes with that from the virtual world [99]. This new reality fully absorbs the player's attention [100], apparently isolating them from the real world [101]. Immersion, contrary to flow, might not describe an optimal experience as it may include negative emotions and anxiety [102].

Although immersion in a gamified application or system can be objectively measured using, e.g., eye-tracking techniques [103], these require dedicated hardware, for which reason we recommend measuring it with a questionnaire administered to the user, i.e., measuring the user-perceived immersion. Even though various measurement scales have been developed to measure immersion, our suggestion is the nine-item scale included in the Gameful Experience Questionnaire (GEQ) [52] using a seven-point Likert scale.

Table 9 presents the nine items that compose the scale for measuring the immersion of a gamified system. When implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 9. Measuring the immersion of a gamified system.

Item	Statement
1	Using the [*the gamified system*] provides me with the feeling that time passes quickly.
2	The [*the gamified system*] grabs all of my attention.
3	Using the [*the gamified system*] provides me with a sense of being separated from the real world.
4	Using the [*the gamified system*] causes me to lose myself in what I am doing.
5	Using the [*the gamified system*] causes my actions to seem to come automatically.
6	Using the [*the gamified system*] causes me to stop noticing when I become tired.
7	Using the [*the gamified system*] causes me to forget about my everyday concerns.
8	Using the [*the gamified system*] causes me to ignore everything around me.
9	Using the [*the gamified system*] causes me to become fully emotionally involved.

Source: adapted from [52].

4.6. Evaluating User Experience of Gamified Systems

The Technology Acceptance Model (TAM) was developed by Davis in 1985 with an original purpose of quantifying the likelihood of technology acceptance [104]. In 2019, Lewis showed that the TAM questionnaire can be used to measure user experience, provided its item format is changed from assessments of likelihood to agreement [53]. In

GATUGU, we adopt this approach, using the two measurement scales proposed by Davis for Ease of Use and Usefulness.

4.6.1. Ease of Use

The Ease of Use of a system or an application describes how easily users can use it. The actual Ease of Use can be measured using metrics such as task success (completion ratio), time on task (completion time), number of errors made by users performing a task, or number of help requests by the users performing a task (for all but the first metric, the lesser their value, the better the Ease of Use) [105]. Measuring the actual Ease of Use is, however, difficult to set up, as it requires a good knowledge of what tasks are typical for a given system, and in what order they are typically performed, and to compare across systems that differ in the scope of tasks performed within them. For this reason, in GATUGU, we suggest to measure the perceived Ease of Use, which, although indirect and subjective, is free from the shortcomings mentioned above.

The perceived Ease of Use can be measured using a six-item scale originally proposed by Davis in 1989 [54], in the form developed by Lewis [53], using a 7-point Likert scale, preferably with numbers 1–7 used instead of textual labels ("Disagree"/"Agree").

Table 10 lists the items of the perceived Ease of Use measure. As in previous cases, when implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 10. Measuring the perceived ease of use of a gamified system.

Item	Statement
1	Learning to operate [*the gamified system*] would be easy for me.
2	I would find it easy to cause [*the gamified system*] to do what I want to do.
3	My interaction with [*the gamified system*] would be clear and understandable.
4	I would find [*the gamified system*] clear and understandable.
5	It would be easy for me to become skillful at using [*the gamified system*].
6	I would find [*the gamified system*] easy to use.

Source: adapted from [53].

4.6.2. Usefulness

Usefulness indicates the extent to which a system or an application allows the user to become closer to or meet their goals in a particular context of use [106]. While it is possible to measure the actual Usefulness of a system by assessing to what extent it helped the users to advance them toward their goals, such measurements are even more context-specific and thus difficult to prepare and compare their results than in the case of measuring the actual Ease of Use. For this reason, in GATUGU, we suggest measuring the perceived Usefulness, which can be measured using a six-item scale originally proposed by Davis in 1989 [54], in the form developed by Lewis [53], using a 7-point Likert scale, preferably with numbers 1–7 used instead of textual labels ("Disagree"/"Agree").

Table 11 lists the items of the perceived Usefulness measure. As in previous cases, when implementing the questionnaire, [*the gamified system*] should be replaced with the name of the gamified system under evaluation.

Table 11. Measuring perceived usefulness of a gamified system.

Item	Statement
1	Using [*the gamified system*] would enable me to more quickly accomplish tasks.
2	Using [*the gamified system*] would improve my performance.
3	Using [*the gamified system*] would increase my productivity.
4	Using [*the gamified system*] would enhance my effectiveness.
5	Using [*the gamified system*] would cause it to be easier to meet my goals.
6	I would find [*the gamified system*] useful.

Source: adapted from [53].

5. Relevance of Respective Evaluation Perspectives to Popular Types of Gamified Systems

Gamification is applied in many different scenarios and, thus, on several distinct systems. GATUGU identifies six (6) evaluation perspectives, containing twenty-three (23) dimensions in total. Such an extensive evaluation instrument aims to cover all aspects of the distinct gamified systems. Consequently, there are dimensions relevant to some kind of gamified systems but irrelevant to another. Table 12 highlights the most relevant GATUGU dimensions to popular types of gamified systems, namely: learning platforms (LP) [12,13]; customer relations (CR) [107]; employee relations (ER) [108,109]; enterprise information systems (EIS) [15,18,19,110]; mobile apps (MA) [20–22]; and web pages/applications (WP) [23,24].

Note that an implemented system may fall into multiple types (e.g., learning platform and web application). For this reason, we highlight, for instance, Performance, Reliability, Availability, and Scalability only in mobile and web applications.

Table 12. Most relevant GATUGU dimensions to popular types of gamified systems.

	LP	CR	ER	EIS	MA	WP
Motivation	✓					
Engagement	✓	✓	✓	✓	✓	✓
Fun	✓	✓	✓	✓	✓	✓
Learning	✓					
Work Performance			✓	✓		
Sales Revenue		✓				
Social Relations			✓			
Stress Reduction			✓	✓		
Performance					✓	✓
Reliability					✓	✓
Availability					✓	✓
Scalability					✓	✓
Adoption Rate	✓				✓	✓
Frequency of Use	✓				✓	✓
Session Length	✓				✓	✓
Cont. Use Intention	✓				✓	✓
User Retention	✓				✓	✓
Visual Attractiveness	✓				✓	✓
Challenge	✓					
Balance	✓					
Immersion	✓					
Ease of Use					✓	✓
Usefulness					✓	✓

6. Measures Selected for Evaluation in the Opinion of Gamification Practitioners

The set of measures described in Section 4 was presented to a number of faculty and industry representatives who were involved in some gamification projects, and were aware of the needs of gamification evaluation. The respondents were reached through a personal and professional network of the authors in December 2022. For each of the selected measures, the respondents were provided with their detailed definition and asked two questions: (1) whether the given measure is appropriate to measure the given dimension (the expected response was a number from 1 to 7); (2) if the first answer was negative, specify why and/or suggest alternative measure/-s (a text box was provided for the response). By 28 December 2022, 12 responses were received via a web questionnaire. The answers to the first questions are shown in Figure 1. As can be observed, for all 23 selected measures, there was a clear prevalence of agreeing answers over disagreeing ones.

Looking for the reasons of disagreeing responses, the answers to the second questions were investigated. The most relevant of them are listed in Table 13 (a dash denotes no comments received for a given dimension). After their careful analysis, it was found that no measure should be considered as wrongly chosen. Some of the critical comments stemmed from the fact that the respondents did not know the explanations presented in this paper, and thus did not know that our goal was to propose just one relevant measure for a given dimension, not all relevant measures, and that we understood the potential need of using a same-level scale in a single questionnaire (hence the provided formulas for converting the results between 5- and 7-point scales).

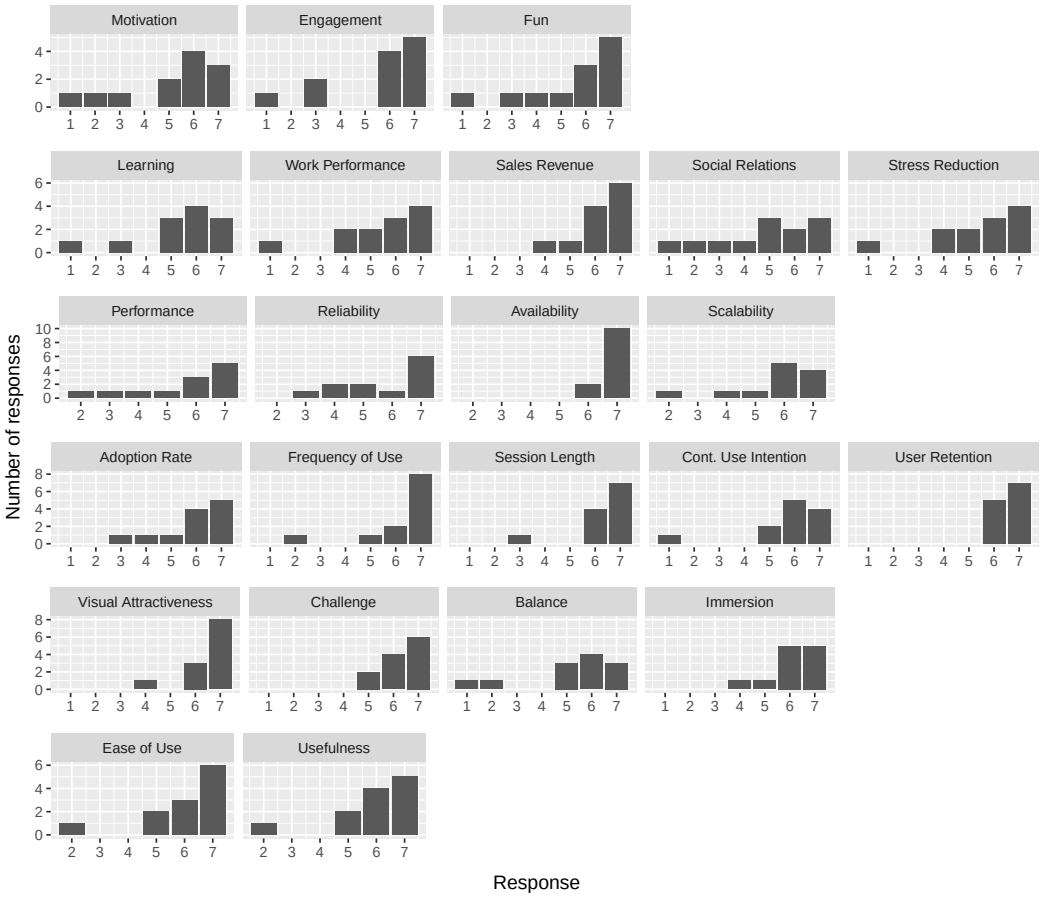

Figure 1. Results of the GATUGU's evaluation questionnaire to gamification experts.

Other comments criticize the form or, in particular, redundancy of the respective scale items. Although we were aware of these weaknesses, all the measurement scales proposed for inclusion in GATUGU were validated and published in peer-reviewed papers, and the respondents did not suggest better alternative measurement scales. In two cases, when alternative metric or measurement scales were proposed, the respondents' proposals were not measuring the exact dimension (MTTR is rather a metric of Availability than Reliability, and SUS is for measuring the Usability rather than the Usefulness).

Table 13. Selected respondents' comments regarding the choice of measures for respective dimensions.

Dimension	Critical Comments
Motivation	"Not sure if this fully covers all aspects of motivation."
	"[Some of the scale items, e.g.,] "for my own good", "is good for me" and "is important for me" are hard to differentiate."
Engagement	"Mix of questions and sentences is confusing"
Fun	"Difficult due to many items. Item 7 is confusing, as fun is the overall construct to be measured"'
Learning	"Other measures may also be relevant in different contexts. For example, self-reported learning outcomes, or observations of changes in behavior or performance, may also be useful for assessing learning."
Work Performance	"Not for all jobs are measurable units."
Sales Revenue	–
Social Relations	"This is purely formal and says nothing about the quality of the social relation."
Stress Reduction	–
Performance	"The response time for user actions and the overall stability of the system may also be important considerations when evaluating the performance of a gamification system."
Reliability	"MTBF does not include the time taken to repair a system after a failure... MTTR is used to measure the average time it takes to repair the system after it has failed, which measures how long the system is offline due to unplanned maintenance. By keeping MTBF high relative to MTTR, the availability of a system is maximised."
Availability	–
Scalability	–
Adoption Rate	–
Frequency of Use	"The length of time that users spend interacting with the system, the number of actions or transactions performed may also be crucial."
Session Length	–
Cont. Use Intention	–
User Retention	–
Visual Attractiveness	–
Challenge	"In general: scales should be the same in one questionnaire if you intend to combine all these questions."
Balance	–
Immersion	–
Ease of Use	–
Usefulness	"SUS Questionnaire."

7. Discussion and Conclusions

The evaluation of gamified systems can be performed in vastly versatile ways. This brings a number of issues with regard to performing evaluation and reporting its results. The lack of awareness of available evaluation perspectives causes gamification system evaluators to limit the scope of their examination to just one or a few dimensions, not necessarily the most relevant ones considering the character and/or purpose of the evaluated gamified application. The lack of typology of evaluation dimensions causes it to be difficult for researchers striving to perform a multi-perspective evaluation to establish a set of adequate measures. The high number of available measurement tools for a single dimension causes it to be very difficult to perform meta-analyses and other types of research based on secondary data, as different reports use different and hard-to-compare measures to measure the same factors. This paper strives to bring order to the field of gamified systems evaluation by:

- Suggesting evaluation perspectives relevant to gamified systems;
- Identifying several dimensions that could be measured in each perspective;
- Recommending one measurement tool for each dimension, so that the reported results conforming to GATUGU could be easily compared across sources.

In our opinion, GATUGU will help researchers in selecting the evaluation perspectives most adequate to the character and/or purpose of the evaluated gamified application, in choosing the dimensions addressing the selected evaluation perspectives, and in finding the measurement tools capable of measuring the chosen dimensions.

There are two types of limitations concerning GATUGU. The first is internal and stems from the fact that GATUGU is based on the authors' subjective (though taking into consideration both their popularity and practical aspects of use) selection of perspectives, dimensions, and measurement tools. As it cannot be said that the omitted perspectives, dimensions, or measurement tools are irrelevant, or even objectively worse for analytical purposes than those that have been selected, not to mention new evaluation concepts that may likely appear in the future, we encourage discussion on extending or improving GATUGU.

The second is external and stems from the fact that the goals of GATUGU will only be achieved if the research community follows its recommendations. When designing GATUGU, we were careful to not to increase the existing chaos; this involved not proposing any new evaluation dimension or measurement tool that has not been used before, therefore GATUGU is safe from such concerns. To truly show its advantages, it will have to be consistently used by a large share of gamification researchers. The authors can only hope that this will happen and support it by their own example, using GATUGU in their future research on gamification evaluation.

Author Contributions: Conceptualization, J.S.; methodology, J.S. and J.C.P.; investigation, J.S., R.Q. and J.C.P.; resources, J.S., R.Q. and J.C.P.; writing—original draft preparation, J.S., R.Q. and J.C.P.; writing—review and editing, J.S., R.Q. and J.C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Karolina Muszyńska for proofreading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ouellette, M. *Examining the Evolution of Gaming and Its Impact on Social, Cultural, and Political Perspectives*; A Volume in the Advances in Human and Social Aspects of Technology (AHSAT) Book Series; IGI Global: Hershey, PA, USA, 2016; pp. 95–122.
2. Andrade, F.R.; Mizoguchi, R.; Isotani, S. The Bright and Dark Sides of Gamification. In Proceedings of the 13th International Conference on Intelligent Tutoring Systems, Zagreb, Croatia, 7–10 June 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9684, pp. 176–186. [CrossRef]
3. Queirós, R.; Pinto, M. The Dangers of Gamification. In Proceedings of the Advanced Research in Technologies, Information, Innovation and Sustainability; Guarda, T., Portela, F., Augusto, M.F., Eds.; Springer Nature: Cham, Switzerland, 2022; pp. 151–162.
4. Crawford, C. *The Art of Computer Game Design*; McGraw-Hill, Inc.: New York, NY, USA, 1984.
5. Julius, K.; Salo, J. Designing Gamification. Master's Thesis, University of Oulu, Oulu, Finland, 2013.
6. Seaborn, K.; Fels, D.I. Gamification in theory and action: A survey. *Int. J. Hum. Comput. Stud.* **2015**, *74*, 14–31. [CrossRef]
7. Hamari, J.; Koivisto, J.; Sarsa, H. Does Gamification Work?—A Literature Review of Empirical studies on Gamification. In Proceedings of the 47th Hawaii International Conference on System Sciences, Washington, DC, USA, 6–9 January 2014; pp. 3025–3034.
8. Hunicke, R.; LeBlanc, M.; Zubek, R. MDA: A formal approach to game design and game research. In Proceedings of the AAAI Workshop on Challenges in Game AI, San Jose, CA, USA, 26 July 2004; Volume 4, p. 1722.
9. Chou, Y.-K. Octalysis: Complete Gamification Framework. Yu-kai Chou: Gamification & Behavioral Design. 2021. Available online: https://yukaichou.com/gamification-examples/octalysis-complete-gamification-framework/ (accessed on 26 September 2022).
10. Marczewski, A. GAME: A Design Process Framework. Gamified U. 2014. Available online: https://www.gamified.uk/2014/05/07/game-design-process-framework/ (accessed on 26 September 2022).
11. Werbach, K.; Hunter, D. *For the Win: How Game Thinking Can Revolutionize Your Business*; Wharton Digital Press: Philadelphia, PA, USA, 2012.
12. Paiva, J.C.; Queirós, R.; Leal, J.P.; Swacha, J.; Miernik, F. An Open-Source Gamified Programming Learning Environment. In Proceedings of the Second International Computer Programming Education Conference (ICPEC 2021), Braga, Portugal, 27–28

May 2021; Henriques, P.R., Portela, F., Queirós, R., Simões, A., Eds.; Open Access Series in Informatics (OASIcs); Schloss Dagstuhl–Leibniz-Zentrum für Informatik: Dagstuhl, Germany, 2021, Volume 91; pp. 5:1–5:8. [CrossRef]
13. Yigitbas, E.; Schmidt, M.; Bucchiarone, A.; Gottschalk, S.; Engels, G. Gamification-Based UML Learning Environment in Virtual Reality. In Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings, Montreal, QB, Canada, 23–28 October 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 27–31. [CrossRef]
14. Lee, C.S.; Foo, J.J.; Sangar, V.J.; Chan, P.Y.; Hor, W.K.; Chan, E.K. A Knowledge Management-Extended Gamified Customer Relationship Management System. In Proceedings of the 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Bali, Indonesia, 26–29 September 2017; pp. 341–346. [CrossRef]
15. Heričko, M.; Rajšp, A.; Horng-Jyh, P.W.; Beranič, T. Using a Simulation Game Approach to Introduce ERP Concepts—A Case Study. In *Knowledge Management in Organizations*; Uden, L., Lu, W., Ting, I.H., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 119–132.
16. Herzig, P.; Strahringer, S.; Ameling, M. Gamification of ERP systems–Exploring gamification effects on user acceptance constructs. In *Multikonferenz Wirtschaftsinformatik*; GITO: Braunschweig, Germany, 2012; pp. 793–804. [CrossRef]
17. Alcivar, I.; Abad, A.G. Design and evaluation of a gamified system for ERP training. *Comput. Hum. Behav.* **2016**, *58*, 109–118. [CrossRef]
18. Jurado, J.L.; Fernandez, A.; Collazos, C.A. Applying Gamification in the Context of Knowledge Management. In Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business, Graz, Austria, 21–23 October 2015; Association for Computing Machinery: New York, NY, USA, 2015. [CrossRef]
19. Elm, D.; Kappen, D.L.; Tondello, G.F.; Nacke, L.E. CLEVER: Gamification and Enterprise Knowledge Learning. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Bremen, Germany, 2–5 November 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 141–148. [CrossRef]
20. Martí, I.G.; Rodríguez, L.E.; Benedito, M.; Trilles, S.; Beltrán, A.; Díaz, L.; Huerta, J. Mobile Application for Noise Pollution Monitoring through Gamification Techniques. In Proceedings of the Entertainment Computing—ICEC 2012, Bremen, Germany, 26–29 September 2012; Herrlich, M., Malaka, R., Masuch, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 562–571.
21. Hsia, B.C.; Singh, A.K.; Njeze, O.; Cosar, E.; Mowrey, W.B.; Feldman, J.; Reznik, M.; Jariwala, S.P. Developing and evaluating ASTHMAXcel adventures: A novel gamified mobile application for pediatric patients with asthma. *Ann. Allergy Asthma Immunol.* **2020**, *125*, 581–588. [CrossRef]
22. Shortt, M.; Tilak, S.; Kuznetcova, I.; Martens, B.; Akinkuolie, B. Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Comput. Assist. Lang. Learn.* **2021**, 1–38. [CrossRef]
23. Emam, H.Y.; Abdel Aziz, G.S. Investigating the Effect of Gamification on Website Features in E-Banking Sector: An Empirical Research. *Acad. J. Contemp. Commer. Res.* **2021**, *1*, 24–37. [CrossRef]
24. Jang, Y.T.; Hsieh, P.S. Understanding consumer behavior in the multimedia context: Incorporating gamification in VR-enhanced web system for tourism e-commerce. *Multimed. Tools Appl.* **2021**, *80*, 29339–29365. [CrossRef]
25. Morschheuser, B.; Hamari, J.; Werder, K.; Abe, J. How to Gamify? A Method For Designing Gamification. In Proceedings of the 50th Hawaii International Conference on System Sciences, HICSS, Waikoloa Village, HI, USA, 4–7 January 2017. [CrossRef]
26. Klock, A.C.T.; Gasparini, I.; Pimenta, M.S. Designing, Developing and Evaluating Gamification: An Overview and Conceptual Approach. In *Data Analytics Approaches in Educational Games and Gamification Systems*; Tlili, A., Chang, M., Eds.; Springer: Singapore, 2019; pp. 227–246. [CrossRef]
27. Dubois, D.J.; Tamburrelli, G. Understanding gamification mechanisms for software development. In Proceedings of the 9th Joint Meeting on Foundations of Software Engineering, Saint Petersburg, Russia, 18–26 August 2013; ACM: Saint Petersburg, Russia, 2013; pp. 659–662.
28. Dal Sasso, T.; Mocci, A.; Lanza, M.; Mastrodicasa, E. How to gamify software engineering. In Proceedings of the 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER), Klagenfurt, Austria, 20–24 February 2017; pp. 261–271. [CrossRef]
29. Fitz-Walter, Z.J. Achievement Unlocked: Investigating the Design of Effective Gamification Experiences for Mobile Applications Furthermore, Devices. Ph.D. Thesis, Queensland University of Technology, Brisbane, QLD, Australia, 2015.
30. Monteiro, R.H.B.; Souza, M.R.D.A.; Oliveira, S.R.B.; Portela, C.D.S.; Lobato, C.E.D.C. The Diversity of Gamification Evaluation in the Software Engineering Education and Industry: Trends, Comparisons and Gaps. *arXiv* **2021**, arXiv:2102.05089.
31. Klock, A.C.T.; Ogawa, A.N.; Gasparini, I.; Pimenta, M.S. Does gamification matter?: A systematic mapping about the evaluation of gamification in educational environments. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Pau, France, 9–13 April 2018; pp. 2006–2012. [CrossRef]
32. Nacke, L.E.; Deterding, S. The maturing of gamification research. *Comput. Hum. Behav.* **2017**, *71*, 450–454. [CrossRef]
33. Gasca-Hurtado, G.P.; Gómez-Álvarez, M.C.; Muñoz, M.; Mejía, J. Proposal of an assessment framework for gamified environments: A case study. *IET Softw.* **2019**, *13*, 122–128. [CrossRef]
34. Ren, W.; Barrett, S.; Das, S. Toward Gamification to Software Engineering and Contribution of Software Engineer. In Proceedings of the 2020 4th International Conference on Management Engineering, Software Engineering and Service Sciences, Wuhan, China, 17–19 January 2020; pp. 1–5. [CrossRef]

35. Inocencio, F. Using gamification in education: A systematic literature review. In Proceedings of the Thirty Ninth International Conference on Information Systems, San Francisco, CA, USA, 13–16 December 2018.
36. Liu, D.; Santhanam, R.; Webster, J. Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Q.* **2017**, *41*, 1011–1034. [CrossRef]
37. Marczewski, A. Defining Gamification—What Do People Really Think? Gamified UK. 2014. Available online: http://www.gamified.uk/2014/04/16/defining-gamification-people-really-think (accessed on 26 September 2022).
38. Jo, J.; Jun, H.; Lim, H. A comparative study on gamification of the flipped classroom in engineering education to enhance the effects of learning. *Comput. Appl. Eng. Educ.* **2018**, *26*, 1626–1640. [CrossRef]
39. Jayalath, J.; Esichaikul, V. Gamification to Enhance Motivation and Engagement in Blended eLearning for Technical and Vocational Education and Training. *Technol. Knowl. Learn.* **2022**, *27*, 91–118. [CrossRef]
40. Díaz-Ramírez, J. Gamification in engineering education—An empirical assessment on learning and game performance. *Heliyon* **2020**, *6*, e04972. [CrossRef]
41. Sarangi, S.; Shah, S. Individuals, teams and organizations score with gamification. *Hum. Resour. Manag. Int. Dig.* **2015**, *23*, 24–27. [CrossRef]
42. Medeiros, D.B.; Neto, P.D.A.D.S.; Passos, E.B.; De Souza Araújo, W. Working and Playing with Scrum. *Int. J. Softw. Eng. Knowl. Eng.* **2015**, *25*, 993–1015. [CrossRef]
43. Landers, R.N.; Bauer, K.N.; Callan, R.C. Gamification of task performance with leaderboards: A goal setting experiment. *Comput. Hum. Behav.* **2017**, *71*, 508–515. [CrossRef]
44. Eisingerich, A.B.; Marchand, A.; Fritze, M.P.; Dong, L. Hook vs. hope: How to enhance customer engagement through gamification. *Int. J. Res. Mark.* **2019**, *36*, 200–215. [CrossRef]
45. Fathian, M.; Sharifi, H.; Solat, F. Investigating the Effect of Gamification Mechanics on Customer Loyalty in Online Stores. *J. Inf. Technol. Manag.* **2019**, *11*, 1–23. [CrossRef]
46. Hu, S.C.; Chen, I.C. A Gamified Online Forum Inspiring Group Intelligence Distillation for Policy Making. In Proceedings of the Advances in Swarm and Computational Intelligence, Beijing, China, 25–28 June 2015; Tan, Y., Shi, Y., Buarque, F., Gelbukh, A., Das, S., Engelbrecht, A., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 423–430.
47. Ding, L. Applying gamifications to asynchronous online discussions: A mixed methods study. *Comput. Hum. Behav.* **2019**, *91*, 1–11. [CrossRef]
48. Christmann, C.A.; Hoffmann, A.; Zolynski, G.; Bleser, G. Stress-Mentor: Linking Gamification and Behavior Change Theory in a Stress Management Application. In Proceedings of the HCI International 2018—Posters' Extended Abstracts, Las Vegas, NV, USA, 15–20 July 2018; Stephanidis, C., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 387–393.
49. Muroi, F.; Tao, X.; Han, T. A Study on the Effect of Gamification on Alleviation Anxiety Levels of the Elderly in China. In Proceedings of the Human Aspects of IT for the Aged Population. Healthy and Active Aging, Copenhagen, Denmark, 19–24 July 2020; Gao, Q., Zhou, J., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 329–342.
50. *ISO/IEC 9126*; Software Engineering—Product Quality. ISO: Geneva, Switzerland, 2001.
51. Prisacaru, D. Performance, Reliability, Availability, Scalability. IASA. Available online: https://itabok.iasaglobal.org/itabok3_0-2/performance-reliability-availability-scalability/ (accessed on 26 September 2022).
52. Högberg, J.; Hamari, J.; Wästlund, E. Gameful Experience Questionnaire (GAMEFULQUEST): An instrument for measuring the perceived gamefulness of system use. *User Model. User Adapt. Interact.* **2019**, *29*, 619–660. [CrossRef]
53. Lewis, J.R. Comparison of Four TAM Item Formats: Effect of Response Option Labels and Order. *J. Usability Stud.* **2019**, *14*, 224–236.
54. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef]
55. Reeve, J. *Understanding Motivation and Emotion*, 7th ed.; John Wiley & Sons: Nashville, TN, USA, 2017.
56. Guay, F.; Vallerand, R.J.; Blanchard, C. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motiv. Emot.* **2000**, *24*, 175–213. [CrossRef]
57. Ryan, R.M.; Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **2000**, *55*, 68–78. [CrossRef]
58. Krath, J.; Schürmann, L.; von Korflesch, H.F. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Comput. Hum. Behav.* **2021**, *125*, 106963. [CrossRef]
59. Kahn, W.A. Psychological Conditions of Personal Engagement and Disengagement at Work. *Acad. Manag. J.* **1990**, *33*, 692–724. [CrossRef]
60. Hamari, J.; Shernoff, D.J.; Rowe, E.; Coller, B.; Asbell-Clarke, J.; Edwards, T. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Comput. Hum. Behav.* **2016**, *54*, 170–179. [CrossRef]
61. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*; Harper & Row: San Francisco, CA, USA, 1990.
62. Rogers, R.; Dillman Carpentier, F.R.; Barnard, L. Media enjoyment as a function of control over characters. *Entertain. Comput.* **2016**, *12*, 29–39. [CrossRef]
63. McKee, A. *FUN!: What Entertainment Tells Us about Living a Good Life*; Springer: Berlin/Heidelberg, Germany, 2016. [CrossRef]
64. McManus, I.C.; Furnham, A. "Fun, Fun, Fun": Types of Fun, Attitudes to Fun, and their Relation to Personality and Biographical Factors. *Psychology* **2010**, *1*, 159–168. [CrossRef]

65. Read, J.C. Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. *Cogn. Technol. Work* **2008**, *10*, 119–128. [CrossRef]
66. van der Sluis, F.; van Dijk, B.; Perloy, B. Measuring fun and enjoyment of children in a museum: Evaluating the Smileyometer. In Proceedings of Measuring Behavior 2012, Utrecht, The Netherlands, 28–31 August 2012; pp. 86–89.
67. Tisza, G.; Markopoulos, P. FunQ: Measuring the fun experience of a learning activity with adolescents. *Curr. Psychol.* **2021**. [CrossRef]
68. Ebbinghaus, H. Memory: A contribution to experimental psychology. *Ann. Neurosci.* **2013**, *20*, 155–156. [CrossRef]
69. Thorndike, L.; Bruce, D. *Animal Intelligence: Experimental Studies*; Routledge: New York, NY, USA, 2017.
70. Hilgard, E.R. *Introduction to Psychology*; Harcourt Brace: San Diego, CA, USA, 1953.
71. Roediger, H.L.; Capaldi, E.D.; Paris, S.G.; Polivy, J.; Herman, C.P. *Psychology*, 4th ed.; Wadsworth Publishing: Belmont, CA, USA, 1995.
72. Delucchi, M. Measuring Student Learning in Social Statistics: A Pretest-Posttest Study of Knowledge Gain. *Teach. Sociol.* **2014**, *42*, 231–239. [CrossRef]
73. Campbell, J.; McCloy, R.; Oppler, S.; Sager, C. *A Theory of Performance*; Jossey-Bass: San Francisco, CA, USA, 1993; pp. 35–70.
74. Institute, C.F. Sales Revenue. 2022. Available online: https://corporatefinanceinstitute.com/resources/knowledge/accounting/sales-revenue/ (accessed on 28 October 2022).
75. Berkman, L.F.; Kawachi, I.; Theorell, T. Working Conditions and Health. In *Social Epidemiology*; Oxford University Press: Oxford, UK, 2014. [CrossRef]
76. Stansfeld, S.A.; Bosma, H.; Hemingway, H.; Marmot, M.G. Psychosocial Work Characteristics and Social Support as Predictors of SF-36 Health Functioning: The Whitehall II Study. *Psychosom. Med.* **1998**, *60*, 247–255. [CrossRef]
77. World Health Organization. Stress. 2021. Available online: https://www.who.int/news-room/questions-and-answers/item/stress (accessed on 28 October 2022).
78. Goldberg, R.C. Exploring the Relationships among Bio-Psycho-Social Measures of Stress: A Multifactorial Approach Towards the Evaluation and Reduction of Stress. Ph.D. Thesis, California Institute of Integral Studies, San Francisco, CA, USA, 2022.
79. Levenstein, S.; Prantera, C.; Varvo, V.; Scribano, M.; Berto, E.; Luzi, C.; Andreoli, A. Development of the perceived stress questionnaire: A new tool for psychosomatic research. *J. Psychosom. Res.* **1993**, *37*, 19–32. [CrossRef]
80. *ISO/IEC 25010*; Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—System and Software Quality Models. ISO: Geneva, Switzerland, 2011.
81. Kulpa, A.; Swacha, J. Design and Implementation of Rule Execution Mechanism for an eGuide Gamification Web Service. In Proceedings of the Information and Software Technologies, Vilnius, Lithuania, 10–12 October 2019; Damaševičius, R., Vasiljevienė, G., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 173–181.
82. Zhang, Q.; Kang, R.; Wen, M. Belief reliability for uncertain random systems. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 3605–3614. [CrossRef]
83. Oyebisi, T. On reliability and maintenance management of electronic equipment in the tropics. *Technovation* **2000**, *20*, 517–522. [CrossRef]
84. Iyer, R.K.; Kalbarczyk, Z.; Kalyanakrishnan, M. Measurement-Based Analysis of Networked System Availability. In *Performance Evaluation: Origins and Directions*; Haring, G., Lindemann, C., Reiser, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2000; pp. 161–199. [CrossRef]
85. Bondi, A.B. Characteristics of scalability and their impact on performance. In Proceedings of the 2nd International Workshop on Software and Performance, Ottawa, OT, Canada, 17–20 September 2000; pp. 195–203.
86. Kenton, W. Rate of Adoption. Investopedia. 2022. Available online: https://www.investopedia.com/terms/r/rate-of-adoption.asp (accessed on 28 October 2022).
87. Santhanamery, T.; Ramayah, T. Explaining and Predicting Users' Continuance Usage Intention Toward E-Filing Utilizing Technology Continuance Theory. In *Encyclopedia of Information Science and Technology*, 4th ed.; IGI Global: Hershey, PA, USA, 2018; pp. 831–846. [CrossRef]
88. Bhattacherjee, A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Q.* **2001**, *25*, 351–370. [CrossRef]
89. Yan, M.; Filieri, R.; Gorton, M. Continuance intention of online technologies: A systematic literature review. *Int. J. Inf. Manag.* **2021**, *58*, 102315. [CrossRef]
90. Mathieson, K. Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. *Inf. Syst. Res.* **1991**, *2*, 173–191. [CrossRef]
91. King, R.C.; Schilhavy, R.A.; Chowa, C.; Chin, W.W. Do Customers Identify with Our Website? The Effects of Website Identification on Repeat Purchase Intention. *Int. J. Electron. Commer.* **2016**, *20*, 319–354. [CrossRef]
92. Csikszentmihalyi, M. *Beyond Boredom and Anxiety*; Jossey-Bass: San Francisco, CA, USA, 2000.
93. Sherry, J.; Greenberg, B.; Lucas, K.; Lachlan, K. *Video Game Uses and Gratifications as Predictors of Use and Game Preference*; American Psychological Association: Washington, DC, USA, 2006; Volume 8, pp. 213–224.
94. Ijsselsteijn, W.; Hoogen, W.; Klimmt, C.; de Kort, Y.; Lindley, C.; Mathiak, K.; Poels, K.; Ravaja, N.; Turpeinen, M.; Vorderer, P. Measuring the Experience of Digital Game Enjoyment. In Proceedings of the International Conference on Methods and Techniques in Behavioral Research, Maastricht, The Netherlands, 26–29 August 2008.

95. van Lankveld, G.; Spronck, P.; van den Herik, H.J.; Rauterberg, M. Incongruity-Based Adaptive Game Balancing. In *Advances in Computer Games, 12th International Conference, ACG 2009, Pamplona, Spain, 11–13 May 2010*; van den Herik, H.J., Spronck, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 208–220.
96. Sirlin, D. Balancing Multiplayer Competitive Games. Game Developer's Conference. 2009. Available online: https://www.gdcvault.com/play/1570/Balancing-Multiplayer-Competitive (accessed on 28 October 2022).
97. Newheiser, M. Playing fair: A look at competition in gaming. *Strange Horizons*, 9 March 2009; p. 1.
98. Zuo, W. Why Are You Satisfied with an Online Game? Exploring Game Attractiveness and Gaming Climate from a Socio-Technical Perspective. Master's Thesis, Department of Information Processing Science, University of Oulu, Oulu, Finland, 2016.
99. Cairns, P.; Cox, A.; Nordin, A.I. Immersion in Digital Games: Review of Gaming Experience Research. In *Handbook of Digital Games*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014; pp. 337–361. [CrossRef]
100. Murray, J.H. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*; The Free Press: Washington, DC, USA, 1997.
101. Patrick, E.; Cosgrove, D.; Slavkovic, A.; Rode, J.A.; Verratti, T.; Chiselko, G. Using a Large Projection Screen as an Alternative to Head-Mounted Displays for Virtual Environments. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, The Hague, The Netherlands, 1–6 April 2000; pp. 478–485. [CrossRef]
102. Jennett, C.; Cox, A.L.; Cairns, P.; Dhoparee, S.; Epps, A.; Tijs, T.; Walton, A. Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* **2008**, *66*, 641–661. [CrossRef]
103. Cox, A.L.; Cairns, P.; Berthouze, N.; Jennett, C. The use of eyetracking for measuring immersion. In *Proceedings of the CogSci 2006 Workshop: What Have Eye Movements Told Us So Far, and What Is Next*; Cognitive Science Society: Vancouver, BC, Canada, 2006.
104. Davis, F.D. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1985.
105. Komischke, T. Usability Testing: 7 Metrics to Assess Ease of Use. CMSWire. 2021. Available online: https://www.cmswire.com/digital-experience/usability-testing-7-metrics-to-assess-ease-of-use/ (accessed on 29 December 2022).
106. MacDonald, C.M.; Atwood, M.E. What Does It Mean for a System to Be Useful? An Exploratory Study of Usefulness. In Proceedings of the 2014 Conference on Designing Interactive Systems, Vancouver, BC, Canada, 21–25 June 2014; pp. 885–894. [CrossRef]
107. Hwang, J.; Choi, L. Having fun while receiving rewards?: Exploration of gamification in loyalty programs for consumer loyalty. *J. Bus. Res.* **2020**, *106*, 365–376. [CrossRef]
108. Barna, B.; Fodor, S. Gamification's Impact on Employee Engagement: Enhancing Employee Well-Being with a Cloud Based Gamified Team-Building Application. In Proceedings of the 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Barcelona, Spain, 6–8 August 2018; pp. 203–208. [CrossRef]
109. Silic, M.; Marzi, G.; Caputo, A.; Bal, P.M. The effects of a gamified human resource management system on job satisfaction and engagement. *Hum. Resour. Manag. J.* **2020**, *30*, 260–277. [CrossRef]
110. Jantke, K.P.; Krebs, J.; Santoso, M. Game Amusement & CRM: Castle Scharfenstein AR case study. In Proceedings of the 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), Tokyo, Japan, 7–10 October 2014; pp. 488–491. [CrossRef]