"Enable or Disable Gamification?" – Analyzing the Impact of Choice in a Gamified Image Tagging Task

Pascal Lessel
Maximilian Altmeyer
DFKI
Saarland Informatics Campus
firstname.lastname@dfki.de

Lea Verena Schmeer
Saarland University
Saarland Informatics Campus
s8laschm@stud.uni-saarland.de

Antonio Krüger
DFKI
Saarland Informatics Campus
krueger@dfki.de

ABSTRACT

This paper investigates a simple form of customization: giving users the choice to enable or disable gamification. We present a study (N=77) in the context of image tagging, in which a gamification approach was shown to be effective in previous work. In our case, some participants could enable or disable gamification after they had experienced the task with and without it. Other participants had no choice and did the task with or without game elements. The results indicate that those who are not attracted by the elements can be motivated to tag more through this choice. In contrast, those that like the elements are not affected by it. This suggests that systems should provide the option to disable gamification in the absence of more sophisticated tailoring.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Customization; "bottom-up"; self-tailoring; decisions

ACM Reference Format:
Pascal Lessel, Maximilian Altmeyer, Lea Verena Schmeer, and Antonio Krüger. 2019. "Enable or Disable Gamification?" – Analyzing the Impact of Choice in a Gamified Image Tagging Task. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300380

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300380

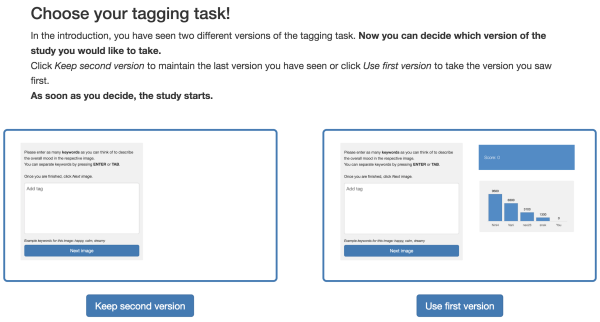

Figure 1: Participants could decide whether they wanted to continue the image tagging with or without gamification.

1 INTRODUCTION

Gamification, the use of game design elements in a non-game context [11], has been shown to be valuable in, for example, changing the behavior of people [4] or making activities more enjoyable [3]. It is already known that "one-size-fits-all" gamification solutions for users are problematic [13, 28, 37]: individuals perceive game design elements differently and when a system uses elements that are not appealing to a particular user, the actual power of the intervention is reduced [28]. The literature has identified aspects that impact perception, among them gender [17], age [2], player type [37], personality traits [15] or culture [1].

Tailoring is a way to overcome these issues and can be achieved through *personalization* (i.e., the system automatically adapts to the user) and *customization* (i.e., the system allows the user to adapt it) [27] or a combination of both [25]. While there are already approaches that investigate how to realize personalization in gamification (e.g., [3, 25]), to our knowledge, they often only consider specific factors. A universal solution incorporating all the above factors is not yet available (see also [6]). In parallel, empowering users to customize the gamification to their needs is investigated [16, 36], even to the extent that users have full control over the gamification at a system's runtime [20, 21], i.e., whether it is active and which game elements are available.

A drawback of customization is the effort a user needs to expend [27]. In this paper, we are thus interested in customization that imposes little effort on a user (as simple as a button click): allowing them to enable/disable the gamification in a system. As having choices was shown to be beneficial in various contexts already (see related work section), we expected to find similar effects. To validate this assumption, we based our work on a gamified image tagging platform concept which was already successfully used in gamification research [23, 24]. The authors showed that a gamification approach based on points and a leaderboard motivates users to create more tags compared to users tagging pictures without gamification. Based on this context, we added a condition where participants could decide to do the tagging task with or without gamification (see Figure 1).

Not only we were able to replicate the previous study results, but our findings also indicate that users who are not attracted by the game elements can be motivated by having a choice: participants making use of it to disable gamification provided significantly more tags compared to those who had no gamification and no choice. Through a range of supporting results, we were also able to relate this difference back to the provided choice, i.e., it was a conscious decision with respect to the provided game elements. In addition, participants that used their choice to enable gamification did not perform significantly differently compared to the (no choice) gamification condition, i.e., they did not perform better or worse. This shows that those participants were not demotivated by the need to enable gamification or the offered choice. Overall, this suggests that allowing users to enable/disable the gamification used in a system is a beneficial and easy-to-use form of tailoring that should be offered in systems, as it motivates those that dislike the available game elements and does not demotivate others.

2 RELATED WORK

Different degrees of customization options (and thus, different degrees of choices) are investigated in game and gamification literature. For example, Kim et al. [16] considered customization on the *aesthetic* (i.e., user can alter the audio-visual appearance of an element) and the *functional* levels (i.e., user can alter a mechanic of an element). In a study, they showed that both add to participants' enjoyment. Another example is the work done by Siu and Riedl [32]. In a game-based scenario, they allowed some of their participants to select the kind of reward (e.g., earning points, unlocking a short story or making progress on a global progress tracker) they would receive after having successfully done a task. Positive effects (e.g., higher task correction) were found for participants that could choose their reward type in comparison to those that could not. Lessel et al. [20, 21] investigated what they call "bottom-up" gamification. Here, users receive options to set up the gamification in a system during runtime: they could select and combine game elements and configure every element further. In the context of a task management application [20], it could be shown that such an idea is appreciated and that users self-report positive effects (e.g., having more fun doing tasks or doing them more conscious through the elements). In [21], in the context of a crowd-sourcing microtask platform, it was also shown that users who used their available "bottom-up" choices solved significantly more microtasks. All these works show that customization has positive effects on the user perception and behavior, but "having a choice" was not tested in isolation: with the choices, the form of gamification also changed; thus, these works tested for the combination of choice and gamification adaptions, which is a difference from the work presented in this paper, which focuses on the choice aspect alone.

It is an ongoing effort to pinpoint why customization in gamification is beneficial [19]. One explanation is based on the *Self-Determination Theory* (SDT) [8], "*one of the most established theoretical frameworks within gamification and game motivation research*" [24]. The theory focuses on three basic needs: *autonomy* (the need to engage in an activity under one's own volition), *relatedness* (the need to feel connected to others and be involved in a social context) and *competence* (the need to experience mastery and having an optimal challenge). According to the SDT, if these needs are satisfied by an activity, there is a higher chance to foster high-quality forms of motivation (i.e., intrinsic motivation; for more details see [9, 10, 30]). It can be seen that customization supports autonomy need satisfaction, as users can act on their own to alter/set up *their* gamification concept. Nicholson, following this theory, demands that users should be put in the loop and systems should allow users to "*create their own tools to track different aspects of the non-game activity, to create their own leveling systems and achievements, to develop their own game-based methods of engaging with the activity and to be able to share that content with other users*" [26]. Another explanation for why customization is beneficial might be that users experience higher levels of ownership [29] or that the created settings are particularly suitable for them (although [19] presents different results).

A further potentially motivational aspect (that is the driving factor for this paper) is "having choices". It has been investigated in isolation in a non-gamification context already: Corah and Boffa [7] exposed participants to white noise and were able to show that providing participants with a choice (to skip or continue the task as they liked), led to a reduced level of aversive equality of the stimulus. Stotland and Blumenthal [34] told all their participants that they had to solve ability tests and that the actual order of how the individual tests were done had no relevance to their score.

Half of their participants had the option to choose the order of the tests, and they were less anxious. Zuckerman et al. [39] considered a puzzle-based task and (separated) pairs of participants: one could decide which three (of six) puzzles to solve and how much time to spend on them. Those participants reported significantly higher feelings of control, spent more time on puzzle solving during a free choice phase and stated they would be likely to return to the laboratory for further puzzle solving. These examples show that providing people with choices has beneficial effects on perceptions and behaviors. Thus, we see it as reasonable to assume that this will be similar in gamification as well.

As discussed above, the degrees of customization, and thus the range of choices, are broad. In the literature, "choice overload" [14] has been investigated in general and was shown to be problematic: having too many choices can lead to negative effects. Schwartz calls this "*Tyranny of Freedom*" [31]. Similarly, works that compare customization and personalization in gamification often highlight the drawback that customization means more effort for users (e.g. [27]). Both aspects need to be considered when setting up user studies that aim to measure the effect of choice. Based on this, we decided to investigate a simple form of choice: the option to enable or disable gamification. While this is a rudimentary form of tailoring, it is unlikely to exert "choice overload" and minimizes potentially confounding variables such as the amount of game elements offered in a study.

3 IMAGE TAGGING PLATFORM

We decided to use a context and a system in which a gamification intervention has already been shown to have positive motivational effects. Mekler et al. [23, 24] used an online image tagging platform for abstract paintings (see Figure 2). The authors compared different non-customizable gamification interventions (points per tag; points and levels; points and leaderboards) with a condition in which no gamification was available. They were able to show that participants in a gamification condition provided significantly more tags without observing a change in the tag quality [24]. In addition, they also found differences between the gamified conditions, showing that the kind of gamification intervention matters as well. For our study, we used/kept the following aspects:

- The online platform was only available in German.
- We kept the tutorial to get participants familiar with the tagging task (but we adapted it slightly; see below).
- The same tutorial images and the same 15 images for the main part (presented in a random order) were used. An image was visible for five seconds; then a text input field appeared, and words describing the mood of the image were to be provided as tags.

Figure 2: Screenshot of the platform used in [23, 24].

- We kept the condition in which the image tagging was done **without** gamification.
- We kept the condition in which the image tagging was done **with** gamification. We used their gamification approach with points and leaderboards, in which participants generated significantly more tags compared to the baseline in [24]: participants receive 100 points per provided tag (independent of the tag quality) and they could compare their performance on a leaderboard (see Figure 2). The entries on the leaderboard relating to other players are always "fictive" and stable across participants, i.e., there is no real competition amongst participants, to rule this as out as an confounding variable. Our fictive players had 1300, 3100, 6300 and 9500 points. Considering these numbers, a participant providing 96 tags across the 15 images would be at the first position on the leaderboard.
- The user interface of their image tagging platform was adapted. We kept the concept that paintings were shown on the left and the game elements on the right (only the game elements were slightly changed visually; compare Figure 2 to Figure 3).

By re-using these aspects, we aimed to replicate the findings with respect to tag quality and quantity made in [24]. This also helps to set our own findings in relation to this and allow comparisons. Instead of using their platform, we re-created one from scratch as we planned to conduct further studies within this setup and could thereby, prepare for these situations properly. In addition, for this study, we extended the platform with these aspects:

- We added a guided tour (see Figure 4) to the tutorial that explained not only the task, but also the game elements involved (when in a condition with game elements), to ensure that participants understand these elements properly. In addition, the tutorial provides three example moods for every shown image, to illustrate what kind of tags we request, in order to make participants familiar with the task.

Figure 3: Our interface in a condition in which gamification was active.

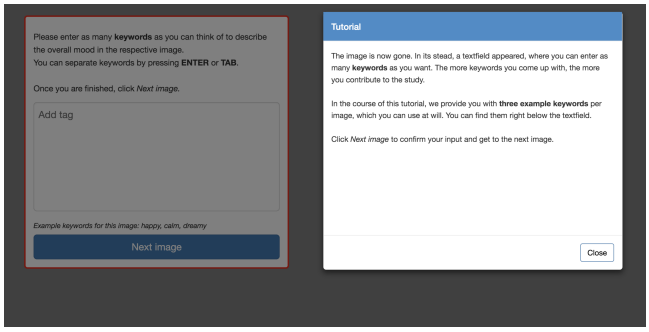

Figure 4: Excerpt of the guided tour in the tutorial explaining the tagging area.

- Depending on the experiment condition, we used questionnaires between and after the two tutorial images and after the 15 main image tagging tasks. The questionnaires were directly integrated into the platform.
- We added a condition in which participants had a choice: after the tutorial, in which they experienced the task with and without game elements (see below), participants could decide whether the main task should be done with or without gamification with just one button click, to minimize user effort (see Figure 1).

We will elaborate on the overall procedure from a participant's point of view in the method section of the study.

4 THE PRESENT STUDY

We had the following hypotheses:

H1 Participants who **have a choice** and use it to **enable gamification** perform subsequently better in a task in comparison to participants who did the task **with the same gamification but without having a choice**.

H2 Participants who **have a choice** and use it to **disable gamification** perform subsequently better in a task in comparison to participants who did the task **without gamification and without having a choice**.

Both hypotheses were based on the formerly presented related work showing that offering choices affects user's behavior positively. The context of this study was image tagging and we used the above-presented image tagging platform. Thus, our main dependent variable was the amount of generated tags. The independent variable was the condition the participant was assigned to (i.e., following a between-subject experimental design). To reason about the effects of choice, we used three conditions: **Choice**, **No Gamification** and **Gamification**. Participants in the **Choice** condition could decide, after the tutorial, whether they wanted to activate gamification (those participants will be considered as being in the condition $\textbf{Choice}_{\textbf{Gamification}}$ subsequently) or deactivate it ($\textbf{Choice}_{\textbf{No Gamification}}$). Then they experienced the main image tagging task similarly to the other conditions. The following relationships were expected:

- More tags in **Gamification** should be generated compared to **No Gamification**. As both conditions were exactly the same as in [24], we expected to replicate this finding. Consequently, we also expect that more tags in $\textbf{Choice}_{\textbf{Gamification}}$ are generated compared to the tag count in $\textbf{Choice}_{\textbf{No Gamification}}$.
- Following **H1**, compared to **Gamification**, more tags in $\textbf{Choice}_{\textbf{Gamification}}$ should be generated.
- Following **H2**, compared to **No Gamification**, more tags in $\textbf{Choice}_{\textbf{No Gamification}}$ should be generated.

As a secondary dependent variable, the tag quality was considered. As [24] did not find significant differences between having gamification or not, it will be interesting to see whether having a choice has an impact.

Method

The link to the above-introduced online image tagging platform was distributed via social media and social circles of the authors. For replication reasons, and as stated before, the platform and all questions were only available in German[1]. No hint towards a motivational study was given in the introduction or in our messages advertising the study. Instead, we highlighted that it would help science (which has led to a positive framing in this context before [23]) and provided the cover story that we wanted to investigate which moods images spark in people. Prospective participants would not receive any further incentive. When users decided to participate, they were assigned to one of the three conditions. We followed a 1:1:2 distribution approach in favor of the **Choice** condition. This was done to increase the number of

[1]The screenshots for this paper were translated. See the supplementary material for the German screenshots and a video showing the conditions.

participants with the goal to have more evenly distributed conditions (after the split of the **Choice** condition). Participants started with the tutorial: they saw two images, one after another, and the tour guided them through the process. We also provided examples to clarify what we meant by "moods". Condition-wise, the tutorials were different:

Choice: For half of the participants, the first image tagging task was presented with gamification and the second image was not (vice versa for the other half of the participants). After each image, the participants were asked to fill out the three items relating to the enjoyment scale of the validated German short scale of the *Intrinsic Motivation Inventory* (IMI) [38]. The IMI was oriented to the activity of the image tagging in general and not to the game elements in particular (e.g., "*Tagging pictures was enjoyable*"). This framing was kept constant across the study and conditions. After having tried out the gamified version, participants were additionally asked to assess the game elements seen, with six statements: "*[Receiving points for image tagging/To be able to compare myself with others on a leaderboard] has motivated me to provide more tags*"; "*I find [points/the comparison with others on a leaderboard] motivating in general*" and "*I like the game element: ['Receiving points for actions'/'Comparison with others on a leaderboard']*". These statements were to be rated on a 5-point scale with the labels *strongly disagree, somewhat disagree, neither agree nor disagree, somewhat agree, strongly agree.* After the second image, participants were led to a screen where they were asked whether they wanted to keep their last seen configuration for the main run by asking them "*Now you can decide which version you would like to use for the study*". We did not use the wording "game" or "gamification"; instead, we simply asked whether a participant wanted to keep the second version or whether they wanted to use the first one they had seen. To help participants remember both versions, we integrated screenshots of them. Figure 1 shows the screen that participants saw if they experienced the tagging task without gamification as the second version. For participants that experienced the gamification as the second version, the screenshots were swapped. The rest of the view (buttons, texts) remained the same.

Gamification: Both tutorial images were shown with gamification enabled. After the tagging, participants were asked to provide the answers to the enjoyment scale of the IMI and to the six game element questions.

No Gamification: Both tutorial images were shown without gamification. After the tagging, participants were asked to provide answers to the enjoyment scale of the IMI.

After the tutorial, participants were informed that 15 images would follow and that should be treated similarly to the images in the tutorial. Depending on the condition/their choice after the tutorial, they experienced this main part with or without gamification. After this, the *Gamification User Types Hexad Scale* [37] (to see whether the player type has an effect on the choice) and the complete short-scale IMI (with its four subscales, enjoyment, pressure, autonomy and competence) had to be filled out, so we could reason about the motivational effects. We closed the experiment with questions in relation to "having a choice": participants in **Gamification** were asked to rate the statement "*I would have liked it if I could have decided whether I wanted to use game elements or not*" on a 5-point scale (same labels as above), and participants in **Choice**, "*I liked that I was able to decide whether I want to use game elements or not*". Participants in **Gamification** and **Choice** were additionally asked to rate "*I would have liked to have had more choices*", and they could provide an answer as free text to explain which choices they would like to have additional. The study was approved by the Ethical Review Board of the Department of Computer Science at Saarland University (No. 18-1-9).

Coding Process

We contacted Mekler et al. and received access to their tag ratings to get familiar with their tagging approach [24] to code our tags. The coding process was done by two independent coders based on a set of rules which were initially derived and refined after a discussion of 100 controversial tags. For every tag given for a picture, they coded whether the tag was neither mood nor picture-related or just nonsense (value 1); only described the picture or was not a fitting mood (value 2) or was a suitable mood (value 3). Based on this, the inter-rater agreement was calculated and was κ=.86. This can be considered as "*almost perfect*" [33].

Participants

90 German-speaking participants tagged all images and completed all questionnaire parts. To identify and remove careless responses [22], we analyzed the answers to the questionnaires for participants that might have only clicked through them (and, for example, only selected extreme values), by calculating the standard deviation of their answers. We removed those who had a standard deviation of equal to or less than .5 in the Hexad or the IMI (exclusion happened across conditions: 1 **No Gamification**/2 $\textbf{Choice}_{\textbf{No Gamification}}$/1 $\textbf{Choice}_{\textbf{Gamification}}$/1 **Gamification**). We also considered the number of tags provided per participant that deviated from the condition mean ± 2.5× the respective standard deviation (148 tags in **No Gamification**, 172 in $\textbf{Choice}_{\textbf{No Gamification}}$ and 464 in $\textbf{Choice}_{\textbf{Gamification}}$ are examples of identified participants). As it is unclear where this lower/higher motivation

Table 1: Generated tags (independent of their quality) across conditions (n=number of participants, M=mean, SD=standard deviation, Mdn=median, Min=minimum, Max=maximum).

		Number of tags generated					
Condition	**n**	**M**	**SD**	**Mdn**	**Min**	**Max**	**Sum**
No Gamification	16	37.7	11.1	36	18	58	603
Gamification	19	62	17.9	63	33	90	1178
$\text{Choice}_{\text{Gamification}}$	19	70.9	24.8	70	18	121	1348
$\text{Choice}_{\text{No Gamification}}$	23	51.2	18.5	51	13	84	1178

comes from (an explanation could be that the study framing was more effective for these participants) and to not confound the data, we excluded these as well (3/3/1/1). In sum, we excluded 13 participants (4/5/2/2), with most exclusions happening in non-gamification conditions. The remaining 77 responses were analyzed (gender: 47 female, 26 male, 2 no answer, 2 other; age: 18–24: 41x, 25–31: 19x, 32–38: 5x, 39–45: 1x, 46–52: 5x, >52: 6x) and considered in the remainder of this paper.

Results

The following main results (MR) were found:

MR1: Gamification motivated participants to provide more tags in the image tagging task

We considered the amount of tags that were provided in the main experiment (i.e., without the tutorial phase) per condition (see Table 1). All tags, independent of their quality, were considered in this calculation (similarly, as was done in Mekler et al.'s work [24]). To compare the amount of tags generated in the conditions, we used an ANOVA, with which we found a significant difference (Welch's $F(3, 39.702)=13.535$, $p<.001$) with a large effect size (est. $\omega=.57$). As Levene's test for homogeneity of variance was almost significant ($p=.05$), we report the more robust Welch's F, instead of just the F statistics. Given the homogeneity aspect, we decided to use the Games-Howell post-hoc procedure, as it is robust even if the homogeneity of variance is questionable and when sample sizes are unequal [12]. It revealed that participants in the **Gamification** condition provided significantly more tags than participants in the **No Gamification** condition (M=62 vs. 37.7; $p<.001$). This replicates the findings reported in [24], i.e., our sample was also motivated by gamification to tag more in the image tagging task. The post-hoc test also revealed a significant difference between $\textbf{Choice}_{\textbf{Gamification}}$ and $\textbf{Choice}_{\textbf{No Gamification}}$ (M=70.9 vs. M=51.2; $p<.05$) that fits this as well.

MR2: Choice motivated those who disabled gamification to provide more tags in the image tagging task

The above described Games-Howell post-hoc test also revealed a significant effect between $\textbf{Choice}_{\textbf{No Gamification}}$ and **No Gamification** (M=51.2 vs. M=37.7, $p<.05$), i.e., participants who could decide to use gamification, but did not want to use it, performed better in the task than participants who had no option and needed to do the task without gamification. Thus, **H2** is supported. In contrast, no significant difference was found for participants in the $\textbf{Choice}_{\textbf{Gamification}}$ compared to participants in the **Gamification** condition (M=70.9 vs. M=62, $p=.585$). Thus, **H1** cannot be supported by this data.

Complementarily, we calculated a two-way ANOVA on the choice and gamification groups as a whole (the Levene's test for homogeneity of variance was borderline significant ($p=.0495$), thus, the following should be treated with caution). No interaction effect between choice and gamification was found ($F(1,73)=.277$, $p=.6$), but the main effects were significant (choice: $F(1,73)=6.663$, $p<.05$; gamification: $F(1,73)=25.583$, $p<.001$). This suggests that both aspects separately increase the number of tags, further supporting MR2.

MR3: The tag quality is not affected by gamification or choice in the image tagging task

Mekler et al. found no differences in the tag quality [24]. To replicate this finding as well, we compared the mean quality ratings across our conditions ($\textbf{Choice}_{\textbf{Gamification}}$: M=2.5; $\textbf{Choice}_{\textbf{No Gamification}}$: M=2.5; **Gamification**: M=2.4; **No Gamification**: M=2.6), but no significant effect was found (Welch's $F(3, 39.53)=2.438$, $p=.079$). Additionally, we only considered the amount of tags that received a high rating (i.e., 2 or 3) condition-wise and the same effects as reported in **MR1** and **MR2** were found. Overall, this hints that the tag quality might not be affected by gamification (in line with Mekler et al. [24]) or choice in the image tagging task.

We analyzed the data further to get more insights about how the choice was perceived in the corresponding groups. This was done to further support **MR2**, and especially to validate that the choice is the actual reason for why participants were motivated to tag more, compared to those that had no choice in the image tagging task. We will formulate these as supporting results (SR).

SR1: The presentation sequence of the two options did not impact the actual choice

To rule out an ordering bias, we checked if the sequence of seeing the gamification first or last in the tutorial had

Table 2: Ratings of the game element statements (5-point scale for each) presented in the different conditions.

Statement	$\text{Choice}_{\text{Gamification}}$ M	SD	Mdn	$\text{Choice}_{\text{No Gamification}}$ M	SD	Mdn	Gamification M	SD	Mdn
Receiving points for image tagging has motivated me to provide more tags	3.8	1	4	2.7	1.1	2	3.2	1.4	4
I find points motivating in general	3.9	.7	4	3.0	1.5	3	3.6	1.1	4
I like the game element: 'Receiving points for actions'	3.9	.7	4	3.1	1.3	3	3.7	.9	4
To be able to compare myself with others on a leaderboard has motivated me to provide more tags	3.5	.8	4	2.5	1.1	2	2.9	1.5	3
I find the comparison with others on a leaderboard motivating in general	3.8	.7	4	3.0	1.2	3	3.1	1.5	3
I like the game element: 'Comparison with others on a leaderboard'	3.6	.8	4	2.6	1.2	3	3.0	1.2	3

Table 3: Ratings on the IMI enjoyment subscale presented after the tutorial (consisting of three 5-point scale statements, which are summed up [38], i.e., value range from 3-15). In No Gamification and Gamification, the questions were only asked once, after both tutorial images had been tagged.

Condition	n	Without Gamification M	SD	Mdn	With Gamification M	SD	Mdn
No Gamification	16	9.8	2.9	10	-	-	-
Gamification	19	-	-	-	10.3	2.3	11
$\text{Choice}_{\text{Gamification}}$	19	9.6	2.9	9	9.5	2.7	9
$\text{Choice}_{\text{No Gamification}}$	23	10.4	3.2	11	9.2	3	10

an effect, and thus would be a potential confounding variable. Of the 23 participants in **Choice** that chose not to use gamification in the main run, eleven saw gamification first, twelve gamification last. Of the 19 participants that chose to use gamification in the main run, nine saw gamification first, ten gamification last. Considering how often the first presented version or last was chosen over all 42 participants in **Choice**, we see that this was evenly distributed (i.e., 50% of the participants in **Choice** selected the "condition" that they saw first). This is also supported by the non-significance of a Chi-square test ($\chi^2(1,42)=.001$, p=1). Overall, we reason that the sequence of presenting the options had no impact on the actual choice, hinting that other criteria were more dominant, for example not liking the game elements, as we will see in the next SR.

SR2: Participants in the Choice condition did not like the game elements equally, and chose accordingly

We focused on participants that were in **Choice** and saw the image tagging version with and without gamification to analyze whether their perceptions of the game elements differ. If true, it is likely that participants who did not particularly like them might have used the offered choice to disable gamification. Thus, this would mean that the choice would have been a deliberate, explainable action. To investigate this, we analyzed how participants who decided not to use gamification answered our game element questions compared to those who decided to use it. Table 2 shows their answers (columns 2 and 3). We calculated independent t-tests for every statement between these two groups and found that the differences were always statistically significant (all at p<.05), i.e., participants who decided to use gamification later in the study provided higher values to each of the points and the leaderboard-related question, compared to those that decided against gamification.

Further support for a deliberate choice which participants made can be derived from the values of the IMI enjoyment questions asked after every tutorial. Table 3 shows the results per condition: overall, the values are all in a similar range. Considering that three 5-point subscales were summed up, a value of 9 is a neutral answer. Thus, the image tagging task is neither particularly enjoyable nor unenjoyable on average in all conditions. To investigate whether those that had seen both versions of the task decided differently, we compared their IMI enjoyment reported in Table 3 condition-wise with paired t-tests: participants later in $\textbf{Choice}_{\textbf{Gamification}}$ assessed the enjoyment of the task not significantly differently (M=9.6 vs. M=9.5, t(18)=.195, p=.848), but those later in $\textbf{Choice}_{\textbf{No Gamification}}$ found the version without gamification significantly more enjoyable (M=10.4 vs. M=9.2, t(22)=2.859, p<.05, r=.52). Taking these results together, we conclude that it is reasonable to assume that participants disabled the gamification when they were not attracted to the game elements or gamification in general.

Table 4: Ratings on the subscales of the IMI after the main image tagging task (every subscale consists of three 5-point scale statements, which are summed up [38], i.e., values range from 3-15 on every subscale).

		Enjoyment			Competence			Autonomy			Pressure		
Condition	**n**	**M**	**SD**	**Mdn**	**M**	**SD**	**Mdn**	**M**	**SD**	**Mdn**	**M**	**SD**	**Mdn**
No Gamification	16	9.3	2.9	10	6.7	1.9	7	8.25	3.3	7.5	9.38	3.3	10
Gamification	19	10.4	3.1	11	8.8	2.4	9	9.7	3.4	10	8.6	3.1	8
$\text{Choice}_{\text{Gamification}}$	19	10.3	3.3	10	9.4	2.2	10	11.8	3.1	13	8.4	2.9	8
$\text{Choice}_{\text{No Gamification}}$	23	9.6	2.6	10	7	2.5	7	11.1	3.1	11	9.1	3.2	10

After the main image tagging task, we additionally asked how participants assessed the option to enable/disable game elements in this scenario and both **Choice** groups provided high ratings ($\textbf{Choice}_{\textbf{No Gamification}}$: M=4.3, SD=1, Mdn=4; $\textbf{Choice}_{\textbf{Gamification}}$: M=4.3, SD=1, Mdn=5), i.e., participants liked the offered choice. We also asked participants in the **Gamification** condition whether they wanted to have a choice to enable/disable the game elements, and received mixed answers with a tendency to disagreement (M=2.5, SD=1.3, Mdn=3). Potentially, the game elements were perceived better in **Gamification** than in $\textbf{Choice}_{\textbf{Gamification}}$, making a choice unnecessary, as participants are satisfied already. Again considering Table 2 and the responses participants provided in **Gamification** (column 4), no significant differences from the **Choice** groups (column 2 or 3) were found (Bonferroni corrections were applied to account for multiple t-tests), which makes this explanation unlikely. One reason for this can be seen in the non-uniformity of participants in **Gamification**: the standard deviations are higher (compared to $\textbf{Choice}_{\textbf{Gamification}}$) and the mean values are between the two other conditions; i.e., in this group not everyone liked gamification similarly, apparently. Another reason might be that it is necessary to experience the range of options in a situation to see value in such a choice. In light of this hypothesis, it is also not surprising that the three groups tended to disagree with the statement that they would like to have more choices ($\textbf{Choice}_{\textbf{No Gamification}}$: M=2.3, SD=1, Mdn=2; $\textbf{Choice}_{\textbf{Gamification}}$: M=2.3, SD=1.2, Mdn=2; **Gamification**: M=2.5, SD=1.2, Mdn=3).

SR3: Participants experienced higher levels of competence with gamification and those having choices experienced more autonomy in the image tagging task

After the main image tagging task, the complete IMI (not just the enjoyment scale as in the tutorial) needed to be answered by every participant. Table 4 shows the results per subscale and condition. We see that participants who tagged images with gamification provided higher values for enjoyment and competence compared to those who tagged without it. We also see that the autonomy is higher in the choice conditions. We compared the subscales separately using ANOVA tests. Levene's test for homogeneity of variance was not significant in all cases (all p>.7) and thus the assumption of homogeneity holds. While no significant differences for enjoyment (F(3,73)=.597, p=.619) and pressure (F(3,73)=.391, p=.76) were found, a significant one was found for competence (F(3,73)=6.505, p<.05, ω=.42) and autonomy (F(3,73)=4.365, p<.05, ω=.34). Because of the unequal sample sizes, we again used the Games-Howell post-hoc procedure:

Competence: Competence differs significantly between **No Gamification** and **Gamification** (M=6.7 vs. M=8.8, p<.05) and **No Gamification** and $\textbf{Choice}_{\textbf{Gamification}}$ (M=6.7 vs. M=9.4, p<.05). $\textbf{Choice}_{\textbf{No Gamification}}$ is also significantly different from $\textbf{Choice}_{\textbf{Gamification}}$ (M=7 vs. M=9.4, p<.05). The difference between $\textbf{Choice}_{\textbf{No Gamification}}$ and **Gamification** is only almost significant (M=7 vs. M=8.8, p=.088). These results show that gamification raises the feeling of competence. Based on the fact that other players were simulated, participants could climb the leaderboard easily. Work such as [5] has already shown that manipulating success perception through leaderboards increases the perception of competence, so this seems likely here as well. In addition, points represented a continuously increasing progress view. Both can be explanations for why gamification changed the feeling of competence in the image tagging task. Given that competence is one of the SDT needs and a higher feeling of competence should add to the intrinsic motivation [8], this could be an explanation for why participants in both gamification conditions provided more tags (see **MR1**).

Autonomy: Autonomy is significantly different between **No Gamification** and $\textbf{Choice}_{\textbf{No Gamification}}$ (M=8.3 vs. M=11.1, p<.05) and **No Gamification** and $\textbf{Choice}_{\textbf{Gamification}}$ (M=8.3 vs. M=11.8, p<.05). Again, according to the SDT, this should add to the intrinsic motivation and could explain why participants in the **Choice** condition provided more tags (given the descriptive data and the significant effect; see **MR2**). This also supports **SR2**, as it underlines that the options were perceived as an *actual choice*. The differences between **Gamification** and $\textbf{Choice}_{\textbf{No Gamification}}$ (M=9.7 vs. M=11.1,

p=.496) and **Gamification** and **Choice$_{\text{Gamification}}$** (M=9.7 vs. M=11.8, p=.196), although higher, are not statistically significant. One reason for this is that the autonomy in the **Gamification** condition is already higher (considering the descriptive data) compared to **No Gamification**. One explanation for this may be that participants in the **Gamification** condition might have a higher perceived feeling of autonomy (even without a clearly formulated choice) as they could decide how much effort they wanted to expend in the competitive task (e.g., people who did not want to engage in the competition could decide not to do so).

The Hexad player type questionnaire and the free text answer did not provide conclusive results; hence, we decided not to present these for readability reasons.

Discussion

Our study replicated the results of Mekler et al. [23, 24]: gamification positively influences the amount of tags generated in a tagging task without negatively affecting tag quality (**MR1** and **MR3**). Additionally, our results show that offering a choice is beneficial (**MR2** and **MR3**): while also not affecting tag quality negatively, we saw a significant increase in the number of tags produced when comparing both groups that experienced the task without gamification. Offering the choice of doing the task with or without game elements motivated participants who chose not to use them. They were more engaged compared to those who simply did the task without game elements. The supporting results showed that having a choice was appreciated and was not confounded by the presentation sequence of the options (**SR1, SR2**). Instead, it seemed to be a conscious decision because of game elements that were not appealing for these participants (**SR2**). Overall, this fits with the related work showing that having choices changes users' perception and behavior (e.g., [7, 39]; see related work section). Thus, based on these results, it seems advisable to provide a choice in a system.

In this respect, not finding a significant effect between the groups that used gamification was surprising. Given our data, it seems that having a choice did not affect participants who decided to use it compared to those who were given gamification without a choice. In both groups a similar amount of tags was generated, i.e., choice did not improve the situation here. An explanation for this could be a ceiling effect based on the IMI result that the task was not perceived as particularly enjoyable. Potentially, those who experienced gamification and were motivated by it could not be further motivated by other means, as the task remained uninteresting. If true, having an additional choice would thus not improve the situation further. To analyze this in more depth, different tasks need to be considered in the future. We reason that setups in which the (core) enjoyment of the task is particularly low or particularly high would be worthwhile to investigate. Here, it would be interesting to see whether in these situations choice on top of the gamification could also be beneficial for those who already like it. Another explanation is that the choice itself has a different value for those who like the gamification and those who dislike it. The former group would experience no drawback if no choice was provided and gamification was active by default (as they like gamification, and receive it). Both explanations lead to the result that providing a choice for those who already like gamification is not harmful for their motivation.

Considering the performance of those who had a choice and decided not to use gamification, we see that these values are below the performance of those who used gamification without having a choice. Although the difference between these two groups is not significant, the question arises whether it would be more advisable to always provide participants with gamification in a system. We do not think that this is a conclusion that should be derived:

First, this would be a form of "one-size-fits-all" gamification that has been shown to be problematic in the literature [13, 28, 37]. Clearly, some of our participants reported disliking the game elements used and decided against them, supporting the fact that our participants were not uniform. It can be assumed that at least some participants in **Gamification** (considering the game element statement ratings and standard deviations as reported in **SR2**) would have decided against gamification when having the chance to do so. While we could find a clear split based on this decision in **Choice**, we do not know who in **Gamification** would have decided against using it eventually. Thus, we cannot reason about this in more detail without being too speculative. However, it is also difficult to assess these in a study setting: if participants are asked before the task whether they want to enable/disable gamification, but then receive it regardless of the answer this can be problematic. Additionally, asking for a decision should be done at the same time in all conditions, and thus asking after the main task is also not an option.

Second, following on this, it is unclear how such participants would have performed. Potentially, a choice to disable gamification instead of experiencing non-optimal gamification would have led to more tags compared to their performance when urged to use non-optimal gamification.

Third, given the effects found in the IMI (**SR3**), higher values for autonomy are visible for those who had a choice. This should affect intrinsic motivation positively [9].

Fourth, as customization in general was shown to be beneficial from a user experience perspective (e.g., [16]) and considering the former aspect, we reason that also the perceived user experience will be better in systems when choices are offered.

In conclusion, considering the different aspects discussed here, we find that offering a choice in gamified system to enable or disable gamification is not harmful and can be particularly engaging for those who do not like the chosen game approach. In our case, this engagement led to an equal or significantly higher number of tags depending on the condition the participant was in.

Limitations

A clear limitation of this study is the small sample size (although the effect sizes were reasonably high) in the presence of three conditions, of which one was split. In future work, we will re-run the study on paid platforms (e.g., *Amazon Mechanical Turk*) to counteract this. Not conducting this study on such a platform was a conscious decision in order to obtain a valid baseline measure; in these future studies, it will not only be interesting to see how stable the results are given even more participants, but also whether these are comparable when the sample is paid.

Another limitation is the fact that we compared a group that *knew* of their choice with two groups that did not know what they "missed". As our results show that experiencing the options of a choice might be important to evaluate it, this has implications for future study designs as well. Besides having a **No Gamification** and a **Gamification** condition as a baseline to compare with, it seems to be reasonable to also integrate similar conditions in which participants also see both versions in the tutorial. Obviously, they would not know that other participants received a choice, but they would at least *know*, what the task looks like with and without gamification (e.g., by stating that they will continue the task with one of the previous options).

A further clear limitation is the fact that we just used a simple gamification approach consisting of points and leaderboards. This was done to replicate the setting of Mekler et al. [23, 24], but it is an open question what would have happened if the gamification had been even more sophisticated and had offered a richer set of game elements. In the same sense, it is questionable whether the same effects would occur by keeping the gamification but changing the task setting, away from an image tagging task. As already explained above, the task itself was not particularly rewarding. Additionally, it was potentially also difficult to provide fitting tags, based on the abstract nature of the images and the request for "moods", which might be a highly idiosyncratic perception. It is questionable how this affected the motivation of participants and their overall perception of the system. In addition, it also raises questions about the reliability of the "tag quality" metric. Again, for replication reasons, we did not want to deviate from the base study in this respect. Overall, similar studies in different settings should be conducted to gain further insights.

5 CONCLUSION

In this paper we considered an easy form of customization: enabling or disabling gamification. While some participants received either no gamification or a gamification approach consisting of points and leaderboards, others could decide whether they wanted to use this gamification or not. We based both the gamification and the task used in this paper on previous literature. In a user study, we were able to reproduce the core results of these former studies (a contribution in itself), and were able to show positive effects of having a choice in the context of image tagging. Participants who disliked the game elements were more likely to disable the gamification and subsequently performed better than those did the same task in the same configuration (i.e., without gamification) but without a choice. Participants who liked the game elements and made the choice to use gamification in the subsequent task did not perform significantly differently from participants who had no choice and needed to do the task with gamification, i.e., performance was not better or worse. Taken together, having a choice to enable or disable gamification seems to be a reasonable and potentially easy-to-realize option for customization that should be integrated into systems. In addition, it does not impose much effort on users, as they could interact with a simple button click, in comparison to more sophisticated customization approaches in which every aspect could be adjusted (e.g., [20, 21]).

We have already provided several directions for future work throughout the discussion and limitation sections. The clear next step will be to re-run the study on a crowdsourcing platform to allow more participants to experience the different conditions. Besides re-using the same conditions as were used in this study, it seems reasonable to also integrate further game concepts as well as offering further potential easy choices (e.g., allowing users to deactivate/activate the gamification during the main experiment as participants see fit). Additionally, the available/not available choices could be more specifically highlighted during the experiment to also consider framing effects, as was done in the choice literature previously (e.g., [18, 35]). It seems also interesting to investigate which further aspects moderate whether or not someone wants to experience gamification, besides the perception of the actual game elements used. Finally, different contexts should be evaluated to replicate our results, beyond image tagging alone. All this will provide more insights into the role of choice in customization.

6 ACKNOWLEDGMENTS

We thank all the participants who participated in our online study. Furthermore, we thank our two coders for rating the large amount of tags that were generated during the course of the study.

REFERENCES

[1] Malik Almaliki, Nan Jiang, Raian Ali, and Fabiano Dalpiaz. 2014. Gamified Culture-Aware Feedback Acquisition. In *Proceedings of the 7th International Conference on Utility and Cloud Computing (UCC '14)*. IEEE, 624–625.

[2] Max V. Birk, Maximilian A. Friehs, and Regan L. Mandryk. 2017. Age-Based Preferences and Player Experience: A Crowdsourced Cross-Sectional Study. In *Proceedings of the 4th Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, 157–170.

[3] Martin Böckle, Isabel Micheel, Markus Bick, and Jasminko Novak. 2018. A Design Framework for Adaptive Gamification Applications. In *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS '18)*. ScholarSpace, 1227–1236.

[4] Simone Borges, Vinicius Durelli, Helena Reis, Ig I. Bittencourt, Riichiro Mizoguchi, and Seiji Isotani. 2017. Selecting Effective Influence Principles for Tailoring Gamification-Based Strategies to Player Roles. In *Proceedings of the 28th Brazilian Symposium on Computers in Education (SBIE '17)*. Sociedade Brasileira de Computação, 857–866.

[5] Jason T. Bowey, Max V. Birk, and Regan L. Mandryk. 2015. Manipulating Leaderboards to Induce Player Experience. In *Proceedings of the 2nd Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. ACM, 115–120.

[6] Marc Busch, Elke E. Mattheiss, Wolfgang Hochleitner, Christina Hochleitner, Michael Lankes, Peter Fröhlich, Rita Orji, and Manfred Tscheligi. 2016. Using Player Type Models for Personalized Game Design – An Empirical Investigation. *Interaction Design and Architecture(s) – IxD&A Journal* 28 (2016), 145–163.

[7] Norman L. Corah and Joseph Boffa. 1970. Perceived Control, Self-Observation, and Response to Aversive Stimulation. *Journal of Personality and Social Psychology* 16, 1 (1970), 1–4.

[8] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 2001. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research* 71, 1 (2001), 1–27.

[9] Edward L. Deci and Richard M. Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (2000), 227–268.

[10] Edward L. Deci and Richard M. Ryan. 2002. Overview of Self-Determination Theory: An Organismic Dialectical Perspective. *Handbook of Self-Determination Research* (2002), 3–33.

[11] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart E. Nacke. 2011. From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th International Academic Mindtrek Conference (AcademicMindtrek '11)*. ACM, 9–15.

[12] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Sage Publications Ltd.

[13] Carrie Heeter, Brian Magerko, Ben Medler, and Yu-Hao Lee. 2011. Impacts of Forced Serious Game Play on Vulnerable Subgroups. *International Journal of Gaming and Computer-Mediated Simulations* 3, 3 (2011), 34–53.

[14] Sheena S. Iyengar and Mark R. Lepper. 2000. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology* 79, 6 (2000), 995–1006.

[15] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. 2016. Personality-Targeted Gamification: A Survey Study on Personality Traits and Motivational Affordances. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2001–2013.

[16] Keunyeong Kim, Michael G. Schmierbach, Saraswathi Bellur, Mun-Young Chung, Julia D. Fraustino, Frank Dardis, and Lee Ahern. 2015. Is It a Sense of Autonomy, Control, or Attachment? Exploring the Effects of In-Game Customization on Game Enjoyment. *Computers in Human Behavior* 48 (2015), 695–705.

[17] Jonna Koivisto and Juho Hamari. 2014. Demographic Differences in Perceived Benefits from Gamification. *Computers in Human Behavior* 35 (2014), 179–188.

[18] Ellen J. Langer and Judith Rodin. 1976. The Effects of Choice and Enhanced Personal Responsibility for the Aged: A Field Experiment in an Institutional Setting. *Journal of Personality and Social Psychology* 34, 2 (1976), 191.

[19] Pascal Lessel, Maximilian Altmeyer, and Antonio Krüger. 2018. Users As Game Designers: Analyzing Gamification Concepts in a "Bottom-Up" Setting. In *Proceedings of the 22nd International Academic Mindtrek Conference (AcademicMindtrek '18)*. ACM, 1–10.

[20] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. "Don't Whip Me With Your Games": Investigating "Bottom-Up" Gamification. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2026–2037.

[21] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2017. Measuring the Effect of "Bottom-Up" Gamification in a Microtask Setting. In *Proceedings of the 21st International Academic Mindtrek Conference (AcademicMindtrek '17)*. ACM, 63–72.

[22] Adam W. Meade and S. Bartholomew Craig. 2012. Identifying Careless Responses in Survey Data. *Psychological Methods* 17, 3 (2012), 1–19.

[23] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. In *Proceedings of the 31st Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, 1137–1142.

[24] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. 2017. Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior* 71 (2017), 525–534.

[25] Baptiste Monterrat, Elise Lavoué, and Sébastien George. 2014. Toward an Adaptive Gamification System for Learning Environments. In *Proceedings of the 6th International Conference on Computer-Supported Education (CSEDU '14)*. Springer, 115–129.

[26] Scott Nicholson. 2012. A User-Centered Theoretical Framework for Meaningful Gamification. In *Proceedings of the 8th International Conference on Games + Learning + Society (GLS '12)*. 1–7.

[27] Rita Orji, Kiemute Oyibo, and Gustavo F. Tondello. 2017. A Comparison of System-Controlled and User-Controlled Personalization Approaches. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)*. ACM, 413–418.

[28] Rita Orji, Gustavo F. Tondello, and Lennart E. Nacke. 2018. Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 435:1–435:14.

[29] Hector Postigo. 2007. Of Mods and Modders: Chasing Down the Value of Fan-Based Digital Game Modifications. *Games and Culture* 2, 4 (2007), 300–313.

[30] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67.

[31] Barry Schwartz. 2000. Self-Determination: The Tyranny of Freedom. *American Psychologist* 55, 1 (2000), 79–88.

[32] Kristin Siu and Mark O. Riedl. 2016. Reward Systems in Human Computation Games. In *Proceedings of the 3rd Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, 266–275.

[33] Steve Stemler. 2001. An Overview of Content Analysis. *Practical Assessment, Research & Evaluation* 7, 17 (2001), 137–146.

[34] Ezra Stotland and Arthur L. Blumenthal. 1964. The Reduction of Anxiety as a Result of the Expectation of Making a Choice. *Canadian Journal of Psychology* 18, 2 (1964), 139–145.

[35] William B. Swann Jr. and Thane S. Pittman. 1977. Initiating Play Activity of Children: The Moderating Influence of Verbal Cues on Intrinsic Motivation. *Child Development* 48 (1977), 1128–1132.

[36] Gustavo F. Tondello and Lennart E. Nacke. 2018. Towards Customizing Gameful Systems by Gameful Design Elements. In *Persuasive '18 Workshop on Personalization in Persuasive Technology*. 1–9.

[37] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 3rd Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, 229–243.

[38] Matthias Wilde, Katrin Bätz, Anastassiya Kovaleva, and Detleff Urhahne. 2009. Testing a Short Scale of Intrinsic Motivation. *Zeitschrift für Didaktik der Naturwissenschaften* 15 (2009), 31–35.

[39] Miron Zuckerman, Joseph Porac, Drew Lathin, and Edward L. Deci. 1978. On the Importance of Self-Determination for Intrinsically-Motivated Behavior. *Personality and Social Psychology Bulletin* 4, 3 (1978), 443–446.