“Absolute or Relative?” – Exploring the Choice Between Leaderboard Types in an Image Tagging Task

Pascal Lessel
Maximilian Altmeyer
Marc Schubhan
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
pascal.lessel@dfki.de
maximilian.altmeyer@dfki.de
marc.schubhan@dfki.de

Dennis Gudea
Saarland University
Saabrücken, Germany
s8degude@stud.uni-saarland.de

Antonio Krüger
German Research Center for Artificial Intelligence (DFKI)
Saabrücken, Germany
krueger@dfki.de

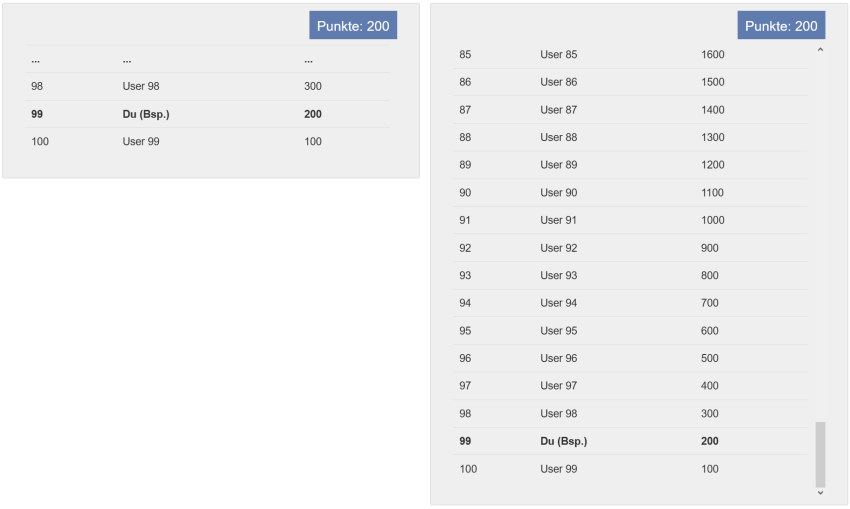

Figure 1: Leaderboards used in the second study (100 participants in the tutorial). Left: Relative leaderboard. Right: Absolute leaderboard.

ABSTRACT

Allowing users to customize gamification has been shown to be promising, but how much freedom and which types of options users should receive is still unclear. In this paper, focusing on low-effort customization, we considered giving users the choice between an absolute (seeing all opponents) and a relative (seeing opponents that are just below and above their own rank) leaderboard in the context of image tagging. Results of two studies showed that, although the task performance is not affected significantly, offering this choice is appreciated, and users’ preferences for the type of leaderboard are moderated by the number of opponents on it.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHI PLAY ’22 EA, November 2–5, 2022, Bremen, Germany
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9211-2/22/11.
https://doi.org/10.1145/3505270.3558346

CCS CONCEPTS

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

KEYWORDS

Autonomy; Self-Determination Theory; Gamification; Competition

ACM Reference Format:
Pascal Lessel, Maximilian Altmeyer, Marc Schubhan, Dennis Gudea, and Antonio Krüger. 2022. “Absolute or Relative?” – Exploring the Choice Between Leaderboard Types in an Image Tagging Task. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY ’22 EA), November 2–5, 2022, Bremen, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3505270.3558346

1 INTRODUCTION & RELATED WORK

Gamification, the use of game design elements in a non-game context [8], often follows an “one-size-fits-all” approach in which every user receives the same set of game elements. This has been shown to be suboptimal (e.g., [4]). Researchers are investigating tailored

gamification approaches through personalization (e.g., [1]) or customization (e.g., [21]). While in the former, the system adapts to the user automatically based on a user model, the latter allows users to actively set up the gamification to their preferences. Concerning self-determination theory [7], the latter also directly addresses to the need for autonomy. Different levels of gamification customization have already been investigated, spanning a spectrum: on the lower end, users could only enable or disable gamification (e.g., [13]); on the upper end, users could decide on which game elements to use in a given task and modify aspects for each as they like (e.g., [11]). Positive effects through the customization options, such as an improved user experience or improved task performance, were reported for these different versions (e.g., [9, 11–13, 21, 23, 25]). However, Orji et al. [18] did a study on the perception of personalization and customization and found that participants think the latter is, for instance, more time consuming and difficult. Similarly, these also require more resources during development (e.g., if ten game elements need to be implemented to provide a rich selection set). Approaches being on the lower end of the spectrum likely do not have these drawbacks, as they probably require less cognitive resources, less interface interactions and will less likely lead to a decision paralysis, as the options are limited. For example, in [13] only a binary decision was involved that could be completed through one button click [13]. However, this lower end of the spectrum is not fully understood yet, as questions such as 1) how many gamification-related choices are needed, 2) which kinds of choices are worthwhile and 3) how many choices are considered too many, are still open. It also seems worthwhile to investigate which customization offers the best compromise between user autonomy, user effort, meaningfulness [17] and effectiveness.

With this paper, we contribute to the second question, as we focused on whether the choice for modifying a fixed game element (e.g., "*Do you want to have version A or B of game element X*") would also lead to positive effects. To our knowledge, this has not been investigated so far, but would fit into low-effort customization and would complement our previous work [13], in which users could only decide to enable gamification, not adapt it. Here, in contrast, while users cannot disable the gamification, adjusting game elements to their own preferences is possible. For the sake of comparability, we situated our work also in the area of image tagging, as here many (tailored and non-tailored) gamification studies have been carried out in the last years and have shown that it is a context in which effects can be measured (e.g., [13, 15, 16, 21, 25]). Concerning the game elements used in these studies, points and leaderboards were typically involved. This also fits other contexts in which points, badges and leaderboards are frequently used (e.g., [10, 22]). Although successful gamification should incorporate more than just these game elements (compare, e.g, [6]), for a controlled study, it seems acceptable to consider these elements alone.

In this work, we focused on leaderboards (and thereby also points), as the literature discusses different realization options (e.g., [2, 3, 5, 19, 20, 27]) with two general approaches that are often called *relative* and *absolute leaderboards* (see Figure 1). Absolute leaderboards are those that show every opponent, or the top X only, while relative leaderboards only show a few "neighboring" opponents around the player's placement. Not only has it been shown that leaderboards in the context of image tagging lead to significant changes in task performance (e.g., [16]), but also much work has been done to understand leaderboard attributes (see [5] for a literature review), including the different leaderboard types. For example, Pedersen et al. [20] reported experiences derived in a *Game With a Purpose* context, with a top 5 leaderboard, a relative leaderboard and without a leaderboard. They found no clear results that the usage of a leaderboard was beneficial in their context. However, they did not analyze whether player preferences differ or subgroups in their samples would profit from specific leaderboard types. Ortiz-Rojas et al. [19] showed that a gamified intervention using an absolute (showing only the first quartile of students) together with a relative leaderboard led to improvements in learning performance compared to a condition with no gamification. However, they did not compare how the leaderboard types would perform when used separately in their context. In the (within-subject) study of Bai et al. [2], participants experienced an absolute and a relative leaderboard (showing five neighbors, but also the top 5 students overall) in an educational context over several months (five weeks per type). Although no performance differences were measurable, through interviews and surveys, it became clear that students have different preferences (e.g., relative was preferred, because of less peer pressure). In follow-up work by Bai et al. [3], again, different preferences for the leaderboards were expressed, making it reasonable to investigate giving users a choice for this.

To investigate the effect of choice in the present work, we conducted two user studies (before exclusions: n=96 and n=41). In study 1, we gave some participants the choice, after having experienced the absolute and relative leaderboard, to select which one they want to use for the main image tagging task; others experienced the complete task with one of the leaderboards only (without a choice). Similarly to [13], if a choice was available the decision could be made with a simple button click. In the second study, we showed significantly more opponents in the tutorial (instead of six, we showed 99) to see whether this had a moderating effect on the choice. Through both studies, we found that participants appreciate having the choice between these two leaderboard types, although this does not lead to (significant) effects on the task performance. In addition, we learned that a task tutorial, with a number of fictive users to explain the game elements, itself moderated participants' decision as to which leaderboard type they wanted to use (although they could not know how many opponents would be presented in the main task). However, this also indicates that the number of opponents has a strong impact on which kind of leaderboard should be offered even in "one-size-fits-all" solutions, as an absolute leaderboard is not the best solution in every case. Effectively, the type of leaderboard might be switched dynamically, if no customization options are available for a user in the system.

2 USER STUDY

We conducted two online studies. In this work-in-progress paper, we will only focus on the main results of both studies. We mirrored the task and context used in our previous study [13], in which, after an tutorial explaining the task and game elements, 15 abstract images were presented one by one and needed to be tagged in terms of which moods the images convey. More information on the setup is given next.

2.1 Apparatus

The main change compared to the online platform used in our low-level customization study [13] was that we did not allow participants to disable gamification, and used two types of leaderboards: the relative leaderboard shows the rank of the user and up to two opponents, one below and one above the user (see Figure 1, left). If the user is in the last place, only one opponent above will be shown; if the user is in the first place, only one below. The absolute leaderboard (see Figure 1, right) shows all opponents at once (for study 1 we ensured that no scrolling was necessary; for study 2 a scrollbar was added, but we set the scrolling position, after an image has been tagged, back to the participant's rank). We needed to switch the visual presentation from a bar chart (as it was done in the former studies) to a list view to ensure comparability of the leaderboards, as more than five ranks were shown. Points were provided after a tag was entered (or subtracted if a tag was deleted again) and the point counter and the user placement on the leaderboard were directly updated. Each entered tag provided 100 points.

2.2 Method

In this section, study 1 (S1) is explained and the changes made for study 2 (S2). Following [13], the study was advertised with the information that participation would help the field of affective image classification (i.e., no hint toward a motivational study was given). We received ethical approval by the Ethical Review Board of the Faculty of Mathematics and Computer Science at Saarland University (No. 21-10-7). After agreement to a privacy statement, the study started.

2.2.1 Conditions. Participants were assigned to a condition before the tutorial started: **NoChoice**$_{absolute}$: the absolute leaderboard was active in the tutorial and the main task; **NoChoice**$_{relative}$: the relative leaderboard was active in tutorial/main task; **Choice**: they received the (one-time) choice, after experiencing both leaderboards in the tutorial (see below), whether to do the main task with the absolute (**Choice**$_{absolute}$) or the relative leaderboard (**Choice**$_{relative}$). We used a 1:1:2 distribution to condition, with the idea that, if participants in **Choice** would choose each leaderboard equally often, we would have an equal number of participants in every condition (before exclusions).

2.2.2 Tutorial. The tutorial explained the task through a guided tour in which two images should be tagged. We gave example tags per tutorial image to help participants understand which kind of tags are of interested (i.e., "tags describing the mood of the image"). The tutorial also explained the active leaderboard. Depending on the condition, the same or different leaderboards (then counterbalanced) were active for the two images. To help participants understand that the tutorial shows only example entries on the leaderboard (i.e., it was unclear how many were active in the main task), the entries showed "User 1", "User 2", etc. as names and their own score with the entry "You (Example)". In total, six opponents and the user's own placement were shown in the tutorial. To reach first place, users need to gain 800 points.

2.2.3 Main task. After the tutorial, participants in **Choice** were able to select the leaderbaord for the main task: we showed a screenshot of each, and showed buttons below the screenshots ("use first version" and "use second version"), i.e., participants could decide with a button click. All participants were prompted to enter a nickname before the main tasks started. As in the main task nicknames (chosen in a way that they could be other players) of opponents were shown on the leaderboard, we needed to ensure that participants did not doubt that these were real users by not being able to enter their own nickname. We also highlighted that they should not use nicknames they typically use, to maintain anonymity. The opponents had the same scores and names for each participant. The main task was carried out, besides the different game elements, as it was in previous works [13, 16, 21], including the same 15 images to tag. In the main task, we displayed 15 opponents and a participant needed 9,800 points to reach first place. After the main task, participants were presented with demographics questions, (free text) questions (answers were coded qualitatively [24]) and statements to answer on 5-point scales (labels: *disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree*), covering aspects of the task, choice and leaderboard perception (will be presented in the result section). Finally, participants were debriefed and the real purpose of the study was disclosed.

2.2.4 Changes for S2. Based on the results of study 1, we set up a complementary second study. It had the same setup as S1. However, we only kept **Choice** as a condition and adapted the questions partially. The two main changes, besides the former aspects, were: 1) we changed the number of opponents shown on the tutorial leaderboards. Instead of showing six, we showed 99 opponents (the user at rank 1 had 10,100 points) and 2) after the choice page, we told participants, that they would need to fill out a questionnaire just before the main task would start, and showed the adapted questions. However, on the last page of the questionnaire, we debriefed them and stated that they did not need to do the main task as we were only interested in their perceptions and how they would decide.

2.3 Participants

The platform, similar to the previous work, was only available in German. The link for S1 was distributed via the social circle of one of the authors, via a mailing list including all students at the University of Applied Sciences Kaiserslautern, as well as the author's professional environment. S2 was distributed via Facebook in groups dedicated for sharing online studies. There was no incentive involved for participating in the study. In both studies, we analyzed the data to detect potential careless responses: we removed those who asked us not to use their data after they knew the true purpose of the study (S1: 3 exclusions, S2: 1); those who answered "no" to the question "*in your honest opinion, should we use your data in the data analysis of the study*" [14] (S1: 9, S2: 1); those who are identified as outliers (+-1.5 * Inter-Quartile Range; Tukey's Fences [26]) in either the number of tags (S1 only: 4) or the survey duration (e.g., one participant needed more than 5 hours, with a long pause in the middle of the task; S1: 4, S2: 1); those who had a tag quality mean below 1.5 (S1 only: 3)[1]; or those who reported something in the free text answers that led to an exclusion (e.g., using offensive language or always using one tag for all images; S1: 7, S2: 0). In S1, 66 participants remained after exclusions

[1] As in [13], two coders assessed the tag quality per tag on a 3-point scale with 1 being the worst rating. The tag quality per tag was the average of the two.

(Age: <18: 8x; 18–24: 38x; 25–31: 15x; 39–45: 2x; 46–52: 1x; >60: 2x; Gender: female: 33x; male: 32x; not answered: 1x). Participants assessed themselves as gaming-affine (mean M=3.9, standard deviation SD=1, median Mdn=4, agreement rate[2] AR=67%) and half the participants reported playing video games/board games frequently (M=3.2, SD=1.5, Mdn=4, AR=51%; M=3.1, SD=1.2, Mdn=4, AR=52%). Finally, the majority of the participants agreed with the statement that it is important for them to perform well in games (M=3.8, SD=1.1, Mdn=4, AR=72%). 49 participants, after being given a definition, reported to have already experienced gamified systems before (74%). In S2, 38 participants remained after exclusions (Age: 18–24: 18x; 25–31: 9x; 32–38: 5x; 39–45: 1x; 46–52: 3x; 53–59=1x, >60 : 1x; Gender: female: 20; male: 18), with other gaming characteristics compared to S1 (gaming-affine: M=3.3, SD=1.5, Mdn=4, AR=53%; frequently playing video games/board games: M=2.6, SD=1.6, Mdn=2, AR=37%/M=2.9, SD=1.1, Mdn=3, AR=37%; wanting to perform well in games: M=3.6, SD=1.2, Mdn=4, AR=68%; 16 had prior experience with gamified systems (42%)).

2.4 Results

We present main results (MR) by considering both studies. We are only interested in the effects of choice, i.e., statistical comparisons will only being made between the respective condition with choice and its non-choice counterpart condition.

MR1: The choice was a purposeful decision in both studies
Every participant with a choice received a free text-based question to explain why they selected the absolute or relative leaderboard. Only 3 of the 32 participants in S1 (compare Table 1) provided an answer indicating that there was no clear reason. The other 29 participants indicated that the choice was in general a purposeful decision. Answers for why the absolute leaderboard was chosen were: 12x better overview (exemplified with the following quote[3]: "*Seeing the entire leaderboard gives you a better sense of ranking in the group than just noticing the next better/worse one*"); 7x a better motivational effect ("*So I know how many tags I have to specify to be at the top*"); 2x less pressure ("*I don't like the 'uncertainty' and that would have put me under even more pressure.*"). For choosing the relative leaderboard, reasons were: 3x less pressure ("*Because I thought that it would trigger less stress with me*"); 3x less distraction ("*More relaxed concentrated work*"); 2x more motivating ("*Since the first place was probably out of reach, I thought it would be more motivating to just target the next person as the next milestone.*") and 1x providing tension. While 4 of the 38 participants in S2 indicated no clear reason, those who selected the absolute leaderboard (16x) did so based on the following reasons: 8x better overview; 7x more motivating and 1x curiosity. Those who selected the relative leaderboard (22x) provided the following reasons: 12x better overview and that it would not overwhelm them; 5x less distraction; 2x that it is more motivating; 2x the absence of scrolling and 1x the surprise effect. Overall, these qualitative answers show that in both studies, for the majority of participants, selecting the leaderboard type was a purposeful decision.

MR2: The choice was perceived as non-overwhelming and was appreciated
In both studies, the majority of participants tended to agree with the statement that they understood their choice through the tutorial (S1: M=3.8, SD=1.3, Mdn=4, AR=69%/S2: M=4.3, SD=.7, Mdn=4, AR=89%) and were not overwhelmed by deciding on a version, as they disagreed with this statement (S1: M=2, SD=1.2, Mdn=2, AR=13%/S2: M=2, SD=1.1, Mdn=2, AR=5%), i.e., the low-effort customization can be seen as an easy to handle choice. In addition, the majority of participants in both studies agreed with the statement that they liked being able to decide which leaderboard version they wanted to use (S1: M=4, SD=1.2, Mdn=4, AR=72%/S2: M=4, SD=1.1, Mdn=4, AR=66%). Interestingly, in S2 participants choosing the relative leaderboard had a much higher, although non-significant, agreement rate compared to those who selected the absolute one (73% vs. 56%; Mann-Whitney test: p=0.258). The majority of participants in S1[4] agreed with the statement that they were satisfied with their choice (M=4.3, SD=.7, Mdn=4, AR=84%); in fact nobody disagreed or somewhat disagreed with this. We also asked (depending on the leaderboard used in the main task) whether they liked that all players were shown on the leaderboard at once ($\textbf{Choice}_{absolute}$: M=4.1, SD=1.1, Mdn=4, AR=74%), or only the players below and above them ($\textbf{Choice}_{relative}$: M=4.1, SD=.8, Mdn=4, AR=78%). Comparing these answers to the respective conditions without a choice ($\textbf{NoChoice}_{absolute}$: M=3.6, SD=1.2, Mdn=4, AR=61%: Mann-Whitney test: U=258.5, z=1.42, p=.156, r=.22 and $\textbf{NoChoice}_{relative}$: M=2.8, SD=1.3, Mdn=3, AR=25%: Mann-Whitney test: U=116, z=2.553, p=.012, r=.5), it became obvious that participants without a choice were less satisfied with the assigned leaderboard. For the statement "*Comparing myself with others on the leaderboard has motivated me to generate more tags*" those who had a choice consistently agreed with it more: $\textbf{NoChoice}_{absolute}$: M=3.3, SD=1.4, Mdn=4, AR=67% vs. $\textbf{Choice}_{absolute}$: M=3.9, SD=1.1, Mdn=4, AR=74% (Mann-Whitney test: U=255, z=1.353, $p_{adj.}$=.352, r=.21) and $\textbf{NoChoice}_{relative}$: M=3.2, SD=1.2, Mdn=4, AR=63% vs. $\textbf{Choice}_{relative}$: M=4.2, SD=1, Mdn=4, AR=89% (Mann-Whitney test: U=111.5, z=2.474, $p_{adj.}$=.046, r=.49). Overall, this shows that offering a choice is reasonable, increases the perceived satisfaction with the game element and has effects (especially for those who use the relative leaderboard) on self-reported measures.

MR3: The choice did not lead to significant effects on the task performance in S1
The choice did not result in a significant performance improvement (see Table 2). Although descriptively, participants who had a choice provided more tags than participants in the same game condition who did not have a choice, these differences were not significant (Student's t-test for $\textbf{NoChoice}_{absolute}$ vs. $\textbf{Choice}_{absolute}$: t(39)=-.62, $p_{adj.}$=1.078, Cohen's d=-.195; $\textbf{NoChoice}_{relative}$ vs. $\textbf{Choice}_{relative}$: t(23)=-1.22, $p_{adj.}$=.472, Cohen's d=-.507). Nearly 2/3 of the participants reported liking the offered gamification (M=3.6, SD=.9, Mdn=4, AR=65%), with no significant changes in the answer to this statement when only comparing conditions with and without a choice. We also checked whether significant differences

[2]The number of participants answering somewhat agree or agree.
[3]Translated from German.
[4]As participants in S2 did not need to execute the image tagging main task, we did not ask them to rate the following statements.

Table 1: Choice distribution in S1 and S2.

Study (n in Choice condition)	Absolute leaderboard chosen	Relative leaderboard chosen
S1 – 6 opponents in tutorial (32)	23 (72%)	9 (28%)
S2 – 99 opponents in tutorial (38)	16 (42%)	22 (58%)

Table 2: Tag quality and quantity in S1 across conditions (M=Mean, SD=Standard deviation, Mdn=Median).

		Tag Quantity			Tag Quality		
Condition	**n**	**M**	**SD**	**Mdn**	**M**	**SD**	**Mdn**
NoChoice$_{absolute}$	18	65.9	31.6	59	2.6	.3	2.7
Choice$_{absolute}$	23	72.1	31.4	65	2.6	.3	2.7
NoChoice$_{relative}$	16	58.1	25.0	57	2.6	.3	2.6
Choice$_{relative}$	9	70	20.1	66	2.6	.3	2.6

were found when considering only those who at least somewhat agreed with the statement that they liked the gamification, but here too, no significant differences in the tag quality or quantity could be found. Finally, we asked participants whether the choice had motivated them to do well on the task, to which only a minority (**Choice**$_{absolute}$: M=3.1, SD, 1.2, Mdn=3, AR=35%) or roughly half the sample agreed (**Choice**$_{relative}$: M=3.7, SD=1.3, Mdn=4, AR=56%). Overall, the choice did not lead to significant task performance changes.

MR4: The number of opponents on the leaderboard moderates the choice and impacts preferences

Table 1 shows that the number of participants who chose the relative or absolute leaderboard differed significantly between both studies, with a medium effect size (as a Fisher's exact test showed: p=.016, effect size: Cramer's V=.3). This could either be explained by the differences in the two samples (see Participants section) or be attributed to the fact that the tutorial showed more opponents on the leaderboard in S2 than S1, although we never explicitly stated that the number of opponents in the tutorial was representative for the main task (and made preparations in the setup so as not to suggest this, see Method section). We asked participants in S2 whether they would have decided differently, if instead of the 100 users, only seven would have been on the leaderboard (as it was in S1): 9 of the 22 participants who selected the relative leaderboard would have picked the absolute in this case. Reasons (provided as free text answers) were that the list would provide enough overview in this case (9x) and that it would not need to be scrolled (2x). None of the participants who picked the absolute leaderboard in S2 would have decided differently. Assuming that the tutorial in S2 would have looked this way, in sum, 25 participants would have selected the absolute (66%) and 13 the relative leaderboard (34%). This would have been a similar distribution as in S1. We also asked participants whether they would have changed their decision if the leaderboard in the tutorial had shown 1000 instead of 100 users. Interestingly, only two participants – one who chose absolute and one who chose relative – would have changed their decision. The former stated that the list would then be too crowded and cluttered and would involve too much scrolling; the latter provided an argument that does not clarify why a switch was deemed as reasonable ("*100 or 1000 participants are not believable for me*"). In sum, two aspects can be derived: first, it seems that the number of users in the tutorial in S1 and S2, although independent of the actual user count in the main task, moderated the choice, suggesting that participants would not think deeply about different circumstances when opting for a game element. Second, independent of the choice, absolute leaderboards become less appealing when more opponents are present, although there seems to be a cut-off point where participants who are in favor of absolute leaderboards stay "invested" in this type of leaderboard.

MR5: Choices are not deemed necessary

We asked those in S1 who had a choice whether they would like to have more choices, to which only a minority agreed (**Choice**$_{absolute}$: M=2.3, SD=1.2, Mdn=2, AR=13% and **Choice**$_{relative}$: M=3, SD=1.2, Mdn=3, AR=33%). The six participants who (somewhat) agreed to this provided inconclusive answers (although two in **Choice**$_{relative}$ stated wanting to disable it completely). Those who had no choice were shown the other leaderboard type and asked whether they would have liked the option to chose between them, to which only a low percentage agreed (**NoChoice**$_{absolute}$: M=2.7, SD=1.6, Mdn=2.5, AR=33% and **NoChoice**$_{relative}$: M=3.1, SD=1.3, Mdn=3, AR=44%). Although the choice is appreciated by users (**M2**), it is apparently not something that is deemed necessary when absent.

3 DISCUSSION AND CONCLUSION

The results of the first study are surprising: Although the choice was appreciated, did not lead to choice overload (both **MR2**) and nearly all participants had clear reasons to choose one over the other leaderboard (**MR1**), no significant effects on the task performance were found (**MR3**). Considering the choice literature (e.g., [7, 28]) and previous gamification choice work (e.g., [13, 21]), we would have expected that the low-effort customization of selecting a leaderboard type would lead to objectively measurable positive effects. However, the study had limitations, on the one hand in respect to the low number of participants overall, but on the other hand also in the condition-wise distribution. Although

we distributed participants in a 1:1:2 manner to the three conditions, with the notion that participants in the **Choice** condition would select the absolute and relative leaderboard equally often, we saw that in S1 the absolute leaderboard choice dominated. As the amount of (fictive) opponents was low in the tutorial and the main task, we hypothesized that maybe another explanation for the non-significant results in the performance measures is that the choice was not perceived as meaningful enough. Therefore, we performed S2 to investigate whether a significantly higher number of opponents in the tutorial would have an impact on the choice, even though in both studies we did not suggest that the number in the tutorial is representative for the main task. With **MR4** we could show that this indeed has an effect. Based on these findings, we see this contribution only as a work in progress: with the low number of opponents in the tutorial in S1 and the one-time choice before seeing the actual number of opponents in the main task, it is difficult to derive whether the choice indeed has no effect on task performance. Setup changes are needed to explore this topic further, such as having more tutorial images showing different (extreme) numbers of opponents on the leaderboard and adding a clear statement that this number is not representative for the main task, or even allowing users to switch during the task itself. This is especially important as **MR5** indicates that choices are not something that people have directly in mind when confronted with such a task.

Still, this work already contributes to the body of knowledge: we showed that users have different preferences in respect to the type of leaderboard, underlining that "one-size-fits-all" gamification is not optimal. While there seem to be general criteria affecting whether users want to have the absolute or relative leaderboard (mentioned in **MR1**), certain users would take one or the other depending on the number of opponents (which is thus a relevant context factor). Even in settings in which a user would not receive customization options in the system, this is a factor that should be considered. We also clearly showed that participants would not think about different ways in which a game element could unfold during the task. Apparently, the tutorial had a significantly moderating effect on the choice, which needs to be considered for future gamification-choice studies. Finally, this work showed that the choice to select a leaderboard type is perceived as non-overwhelming and is appreciated, which are two core aspects for further investigations of this kind of low-effort customization option.

ACKNOWLEDGMENTS

This work is partially supported by the German Federal Ministry of Education and Research (16SV8364).

REFERENCES

[1] Maximilian Altmeyer, Pascal Lessel, Subhashini Jantwal, Linda Muller, Florian Daiber, and Antonio Krüger. 2021. Potential and Effects of Personalizing Gameful Fitness Applications Using Behavior Change Intentions and Hexad User Types. *User Modeling and User-Adapted Interaction* 31, 4 (2021), 675–712.

[2] Shurui Bai, Khe Foon Hew, and Donn Emmanuel Gonda. 2021. Examining Effects of Different Leaderboards on Students' Learning Performance, Intrinsic Motivation, and Perception in Gamified Online Learning Setting. In *2021 IEEE International Conference on Educational Technology (ICET)*. IEEE, 36–41.

[3] Shurui Bai, Khe Foon Hew, Michael Sailer, and Chengyuan Jia. 2021. From Top to Bottom: How Positions on Different Types of Leaderboard May Affect Fully Online Student Learning Performance, Intrinsic Motivation, and Course Engagement. *Computers & Education* 173 (2021), 104297.

[4] 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.

[5] Yang Cao, Shao-Ying Gong, Zhen Wang, Yang Cheng, and Yan-Qing Wang. 2022. More Challenging or More Achievable? The Impacts of Difficulty and Dominant Goal Orientation in Leaderboards Within Educational Gamification. *Journal of Computer Assisted Learning* 38, 3 (2022), 845–860.

[6] Yu-kai Chou. 2019. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Packt Publishing Ltd.

[7] 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.

[8] 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.

[9] Ido Guy, Anat Hashavit, and Yaniv Corem. 2015. Games for Crowds: A Crowdsourcing Game Platform for the Enterprise. In *Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '15)*. ACM, 1860–1871.

[10] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14)*. IEEE, 3025–3034.

[11] 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.

[12] 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.

[13] 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 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 1–12.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Margarita Ortiz-Rojas, Katherine Chiluiza, and Martin Valcke. 2019. Gamification Through Leaderboards: An Empirical Study in Engineering Education. *Computer Applications in Engineering Education* 27, 4 (2019), 777–788.

[20] Mads Kock Pedersen, Nanna Ravn Rasmussen, Jacob F Sherson, and Rajiv Vaid Basaiawmoit. 2017. Leaderboard Effects on Player Performance in a Citizen Science Game. *arXiv preprint arXiv:1707.03704* (2017).

[21] Marc Schubhan, Maximilian Altmeyer, Dominic Buchheit, and Pascal Lessel. 2020. Investigating User-Created Gamification in an Image Tagging Task. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–12.

[22] Katie Seaborn and Deborah I. Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31.

[23] 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.

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

[25] Gustavo F. Tondello and Lennart E. Nacke. 2020. Validation of User Preferences and Effects of Personalized Gamification on Task Performance. *Frontiers in Computer Science* 2 (2020), 29.

[26] John W Tukey. 1977. *Exploratory Data Analysis*. Vol. 1. Pearson Education.

[27] Gabe Zichermann and Christopher Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps.* O'Reilly Media, Inc.

[28] 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.