Metaverse, Virtual Environments and Game Design, Vol. 217, 2026, 174–183
https://doi.org/10.54941/ahfe1007664

Game On? Exploring Subgroup Differences in the Effects of Gamification on Task Motivation in Crowdsourcing

Edwin Gamboa, Fatemeh Niayeshnia, Soroush Daneshi, and Matthias Hirth

User-centric Analysis of Multimedia Data Group, TU Ilmenau, Ilmenau, 98693, Germany

ABSTRACT

Crowdsourcing enables organizations to assign tasks to anonymous crowdworkers, whose output depends on task complexity, instructions, and worker proficiency and motivation. Gamification, i.e., the use of game elements in non-ludic context, has been proposed to boost motivation, but prior studies often modify the task when gamifying, rely on qualitative feedback or have mixed results, limiting clear conclusions. We examined gamification in a car image annotation microtask by adding game elements using a player-centered design based on HEXAD user types, while keeping the task unchanged. For this, we conducted two studies: an exploratory study to identify potential subgroup effects, and a confirmatory replication with a larger sample. Enjoyment, engagement, and performance were measured for two versions of the task: a baseline and a gamified. Across both studies, gamification did not yield consistent improvements. While the exploratory study indicated potential trends, these were not confirmed by the second study. These results suggest that for straightforward microtasks, task clarity and design outweigh additional gamified elements in supporting engagement and performance.

Keywords: Crowdsourcing, Gamification, Motivation, Task performance

INTRODUCTION

Crowdsourcing allows individuals and organizations to outsource tasks to an anonymous group of individuals called crowdworkers, who are paid after successful completion (Estellés-Arolas & Guevara, 2012). The quality of crowdworkers' contributions depends on factors like task complexity, but also their proficiency and motivation towards the task. Gamification, i.e., the use of game elements in non-ludic contexts (Deterding et al., 2011), has been explored as one option to foster crowdworkers' motivation (Morschheuser et al., 2017) or as an alternative to monetary rewards (Carlier et al., 2016). However, existing studies often modify the intended microtask, creating a gamified version that differs from the original, or rely on qualitative feedback to assess workers' attitudes, making it difficult to isolate its actual impact on task outcomes.

To address this gap, we adopt a player-centered approach to gamification based on the HEXAD framework (Tondello et al., 2016). This framework identifies different user types and their preferred game elements, enabling

Received February 4, 2026; Revised April 8, 2026; Accepted April 23, 2026; Available online July 20, 2026

the design of gamification tailored to workers' motivations. We use previous findings on crowdworkers' HEXAD types (Gamboa et al., 2022) to implement a set of gamification elements without altering the car annotation task itself.

In this paper, we present two studies conducted under identical conditions. The first study (*Study 1*) is exploratory, aimed at identifying potential subgroup-specific effects of gamification. The second study (*Study 2*) is a confirmatory replication designed to assess whether the trends observed in *Study 1* persist when tested with a larger sample.

The studies investigate the impact of gamification on three key outcomes, i.e., workers' enjoyment, engagement, and performance. Specifically, we address the following research questions:

RQ1: Does gamification increase workers' enjoyment of the task?
RQ2: Does gamification affect workers' engagement, measured as voluntary completion of multiple sessions?
RQ3: Does gamification influence workers' performance on the same task?

The rest of the paper is organized as follows. First, we briefly outline the concept of crowdsourcing and related works. Then, we detail the methodology and design of our studies. After that, we present and discuss our results regarding the research questions. Finally, we summarize and conclude the paper.

BACKGROUND AND REALTED WORK

In crowdsourcing, non-expert crowdworkers complete online microtasks in exchange for monetary rewards via dedicated platforms (Christoforou et al., 2025; Estellés-Arolas & Guevara, 2012). Microtasks are small, well-defined units of work designed to be completed within minutes, reducing task complexity and enabling focused contributions (Daniel et al., 2018). Typical microtasks include content annotation, surveys, providing of subjective feedback, or data completion and correction. The quality of the results depends on several factors, including task difficulty, the quality of instructions, task duration, and crowdworkers' abilities and motivation (Daniel et al., 2018). Consequently, effective crowdsourcing requires, the inclusion of reliability checks, the design of unambiguous instructions and the optimization of the user interface to meet crowdworkers' needs (Hirth et al., 2020; Rahmanian & Davis, 2014).

In this context, motivation plays a central role in how workers interact with a task. Accordingly, prior work has examined motivation and performance as related outcomes when evaluating crowdsourcing approaches. One such approach is gamification which has been applied in various forms. A comprehensive review by Morschheuser et al. (2017) reports different applications including object recognition, OCR text validation, and video transcription. They also document common game elements including points, leaderboards, badges, levels, progress indicators, and feedback. The review further reports that in most studies, gamification served as the main reward mechanism, which contrasts with evidence that payment remains the dominant

motivator for crowdworkers (Soprano et al., 2024). Overall, the review by Morschheuser et al. (2017) highlights effects of gamification, ranging from increased motivation to higher contributions, while emphasizing the role of contextual factors and worker characteristics. Moreover, they argue that only few studies rely on validated psychometric instruments to assess subjective outcomes such as enjoyment.

Other works suggest rather mixed effects of gamification in crowdsourcing. For example, Walter et al. (2022) study whether gamification could increase the performance of paid crowdworkers in a 3D mesh labeling task. Their results showed that gamification led workers to collect more data compared to a non-gamified version, but this effect decreased for workers who performed the task repeatedly. At the same time, correctness was higher in the non-gamified condition. However, the study does not report a user-centered rationale for the selection of the applied gamification elements

Furthermore, Carlier et al. (2016) compared a standard object segmentation task with a game-based version. The results showed a substantial loss in quality in the game-based approach. The authors attribute this gamification loss to differences in task nature, as workers focused on playing the game rather than performing the segmentation task itself. In contrast, Jenadeleh et al. (2024) showed that immediate feedback improved user experience and reduced response time in a subjective image quality assessment task, while not affecting workers performance.

In summary, prior work indicates that the effects of gamification in crowdsourcing are highly context-dependent and vary across outcomes. While gamification can improve aspects of user experience or short-term engagement, its impact on performance remains inconsistent. In addition, methodological limitations such as non-equivalent task designs, limited use of validated measures, and a lack of replication make it difficult to draw robust conclusions.

STUDIES DESCRIPTION

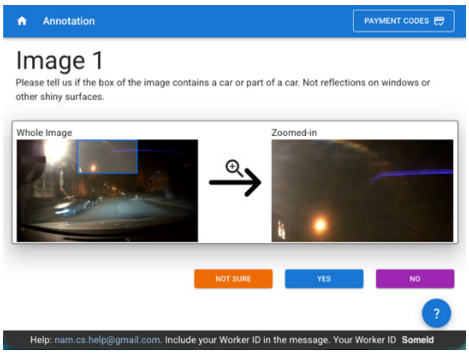
To examine differences in the effects of using gamification in a crowdsourced image annotation task, we conducted two user studies. *Study 1* was designed as an exploratory investigation to identify potential subgroup-specific effects of gamification. The exploratory analyses revealed several trends in unadjusted results that did not remain significant after correcting for multiple testing, indicating possible but inconclusive effects. Therefore, to assess whether these trends would persist with increased statistical power, we conducted *Study 2* as a confirmatory replication under identical conditions with a larger sample.

TASK VERSIONS

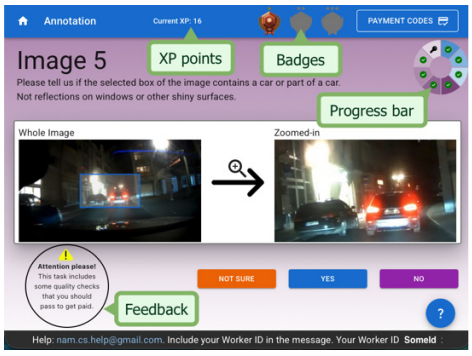
The task used in this study is an image annotation task, in which workers are shown five images, each containing a single bounding box, and are asked to decide whether the box contains a car or part of a car. In the *baseline* version (see Figure 1a), first, the workers receive general information about the task.

Then, they fill a demographic questionnaire out to report their location, crowdworking habits, and experience with image annotation. Next, they complete a training phase, in which they learn how to do the annotation by validating 3 images. After successful completion, they should validate 5 images. Then, they should answer 7 questions to report their enjoyment. Finally, they receive a payment code. The workers can complete at most 3 sessions of the task.

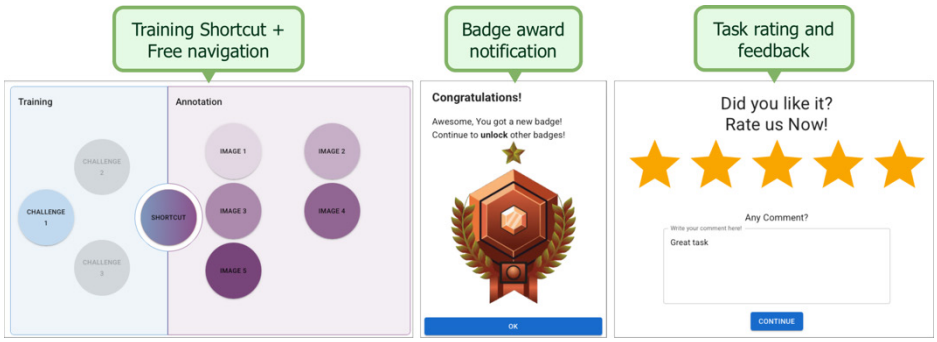
The *gamified* version (see Figure 1b and c) has essentially the same components as the baseline task but includes additional gamification elements. It was designed in a player-centered way based on the HEXAD framework (Tondello et al., 2016), and user types identified for crowdworkers in a previous study (Gamboa et al., 2022). To support *mastery* and *reward*, workers receive experience points after each validation, which later allow them to get badges. Both elements are always visible at the top of the task. *Autonomy* is supported in the training phase by allowing workers to choose between validating three images or completing a more complex shortcut validation to access the main task. They also can decide in which order they want to complete that task. Finally, the possibility to *provide feedback* is promoted by allowing workers to rate the task on a five-star scale and optionally submit a comment.



a) Baseline version



b) Gamified version



c) Additional elements in the gamified version

Figure 1: Car annotation task performed by crowdworkers. The gamified version includes elements like experience points, badges, on-task feedback, and a progress bar.

MEASURES

To assess the effect of using gamification on the car annotation task, three measures were selected. First, *enjoyment (RQ1)* was measured using the Enjoyment sub-scale of the Intrinsic Motivation Inventory (Green, 1972). Second, *engagement (RQ2)* was measured as the number of task sessions that each worker completed. Third, *performance (RQ3)* was calculated using the F1-Score, which summarizes accuracy by balancing correct and missed cases in a single value.

Analysis Strategy

Both *Study 1* and *Study 2* were analyzed using the same procedure. First, the measures for the baseline and the gamified task versions were compared using Wilcoxon rank-sum tests, as the data was not normally distributed and each worker interacted with only one task version. Subsequently, subgroup analyses were conducted using separate linear regression models fitted within each subgroup level (i.e., gender, annotation experience, and crowdworking frequency) to explore whether gamification might differentially affect specific groups of workers.

Because subgroup analyses substantially increase the number of statistical tests and the risk of false-positive findings, p-values were adjusted using the Benjamini–Hochberg (BH) false discovery rate procedure. Adjusted BH p-values were used as the primary criterion for statistical significance, whereas unadjusted p-values were reported to transparently document exploratory trends identified in *Study 1* and evaluated in *Study 2*.

RESULTS

The task was published on the Microworkers[1] platform for both studies and workers were compensated 0.10 USD per session. A total of 126 workers accessed the task in *Study 1*. After identifying potential trends in the collected data, we conducted *Study 2*, which was completed by 630 workers. In both studies, only workers who passed pre-defined validity checks were included in the analyses. These checks required consistency in per- and post-task questionaries regarding the workers demographics, agreement in self-reported and platform-reported country, and coherent responses to the reverse-coded items of the enjoyment scale. For *Study 1*, 89 workers (70.63%) passed the validity checks, while 470 (74.60%) did it for the *Study 2*. Table 1 reports the demographics of the valid participants included in the analyses. The data collected from the crowdworkers is available Online (Gamboa et al., 2026).

[1]https://microworkers.com - Accessed Feb. 2026

Table 1: Demographic overview of participants in both studies, separated by task version.

Variable	Value	Study 1 (n = 89)		Study 2 (n = 470)	
		Baseline	Gamified	Baseline	Gamified
Gender	Male	29	33	179	107
	Female	13	14	114	70
Task familiarity	Some or none	30	34	192	117
	Completely familiar	12	13	101	60
Crowd-working frequency	Several times a day	28	30	167	102
	Once a day	7	7	52	35
	A few times a month	7	10	74	40

Overall Effects

After applying Wilcoxon rank-sum tests in the exploratory *Study 1*, no significant differences between the *baseline* and the *gamified* task were identified for any of the measures, as shown in Table 2 and depicted in Figure 2. These findings were subsequently replicated in the confirmatory *Study 2* (Table 2).

Table 2: Wilcoxon rank-sum test results comparing task versions in *Studies 1 and 2.*

Compared Variable	Median (IQR)		W	p
	Baseline	Gamified		
Study 1				
Enjoyment (RQ1)	5.71 (1.32)	6.00 (1.24)	774	0.08
Engagement (RQ2)	3.00 (1.75)	3.00 (1.00)	1008	0.80
Performance (RQ3)	0.89 (0.16)	0.89 (0.26)	938	0.70
Study 2				
Enjoyment (RQ1)	5.71 (1.43)	5.86 (1.60)	25511	0.80
Engagement (RQ2)	3.00 (1.00)	3.00 (2.00)	24112	0.10
Performance (RQ3)	0.89 (0.25)	0.89 (0.33)	26750	0.50

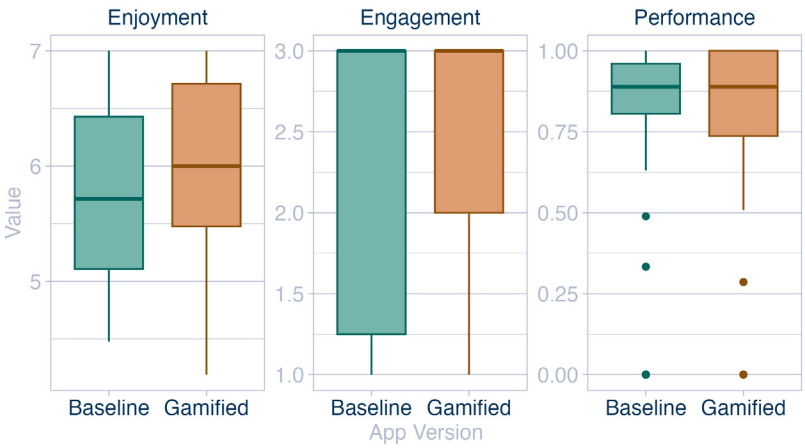


Figure 2: Boxplots of enjoyment, engagement, and performance for Study 1. Plots from Study 2 are omitted due to similar patterns.

Subgroup Analyses

To examine potential differential effects of gamification, subgroup analyses were first conducted in the exploratory *Study 1*. The same analyses were then applied in the confirmatory *Study 2*, allowing us to evaluate whether the trends observed in Study 1 were statistically robust.

As shown in Table 3, the exploratory results for *Study 1* suggested potential positive effects of gamification on workers perceived enjoyment (RQ1) in some subgroups, specifically female workers and participants with little or no experience in image annotation. The identified subgroup trends were subsequently tested in *Study 2* but were not confirmed.

A similar pattern was observed for task engagement (RQ2), as shown in Table 4. In the exploratory *Study 1*, female workers and participants who typically worked once per day showed trends toward completing more sessions in the gamified version of the task. These exploratory trends were subsequently tested in *Study 2* but were not confirmed.

Table 3: Subgroup linear regression results for perceived Enjoyment (RQ1). Estimates reflect the effect of the *Gamified* version relative to the *Baseline* (*p < .05).

Subgroup		Est. (β)	St. Err.	t	Raw p	Adj. BH p
Study 1						
Gender	Male	0.08	0.19	0.41	0.68	0.83
	Female	0.65	0.29	2.36	0.03*	0.14
Task familiarity	Some or none	0.39	0.18	2.11	0.04*	0.14
	Completely familiar	–0.07	0.30	–0.22	0.83	0.83
Crowd-working frequency	Several times a day	0.35	0.20	1.74	0.09	0.20
	Once a day	0.30	0.38	0.79	0.45	0.78
	A few times a month	–0.08	0.36	-0.23	0.82	0.83
Study 2						
Gender	Male	0.19	0.13	1.50	0.14	0.47
	Female	–0.32	0.19	–1.73	0.09	0.47
Task familiarity	Some or none	0.00	0.13	0.03	0.97	0.97
	Completely familiar	–0.03	0.18	–0.17	0.86	0.97
Crowd-working frequency	Several times a day	–0.06	0.14	–0.46	0.65	0.97
	Once a day	0.20	0.24	0.84	0.40	0.94
	A few times a month	–0.04	0.23	–0.19	0.85	0.97

Table 4: Subgroup linear regression results for Engagement (RQ2). Estimates reflect the effect of the *Gamified* version relative to the *Baseline* (*p < .05).

Subgroup		Est. (β)	St. Err.	t	Raw p	Adj. BH p
Study 1						
Gender	Male	–0.58	0.50	–1.16	0.25	0.58
	Female	1.13	0.48	2.34	**0.03***	0.15

(Continued)

Table 4: Continued.

Subgroup		Est. (β)	St. Err.	t	Raw p	Adj. BH p
Task familiarity	Some or none	–0.14	0.53	–0.27	0.79	0.79
	Completely familiar	0.15	0.22	0.69	0.50	0.64
Crowd-working frequency	Several times a day	–0.44	0.50	–0.88	0.38	0.64
	Once a day	2.14	0.96	2.23	**0.04***	0.15
	A few times a month	–0.29	0.47	–0.61	0.55	0.64
Study 2						
Gender	Male	0.06	0.09	0.65	0.51	0.51
	Female	0.24	0.13	1.78	0.08	0.32
Task familiarity	Some or none	0.11	0.10	1.13	0.26	0.46
	Completely familiar	0.18	0.13	1.34	0.18	0.42
Crowd-working frequency	Several times a day	0.07	0.10	0.69	0.49	0.51
	Once a day	0.30	0.18	1.70	0.09	0.32
	A few times a month	0.13	0.16	0.85	0.40	0.51

Finally, performance (RQ3) outcomes in the exploratory *Study 1* did not suggest any effect of gamification. However, in the confirmatory *Study 2*, female crowdworkers showed a significant negative effect (Estimate = –0.14, r = 0.21), indicating a small-to-moderate decrease in performance in the gamified version. The results are reported in Table 5.

Table 5: Subgroup linear regression results for Performance (RQ3). Estimates reflect the effect of the *Gamified* version relative to the *Baseline* (*p < .05).

Subgroup		Est. (β)	St. Err.	t	Raw p	Adj. BH p
Study 1						
Gender	Male	0.01	0.06	0.25	0.81	0.94
	Female	0.15	0.13	1.11	0.28	0.60
Task familiarity	Some or none	0.10	0.07	1.37	0.18	0.60
	Completely familiar	–0.08	0.05	–1.73	0.10	0.60
Crowd-working frequency	Several times a day	0.07	0.07	0.96	0.34	0.60
	Once a day	0.07	0.18	0.37	0.71	0.94
	A few times a month	0.00	0.12	0.02	0.99	0.99
Study 2						
Gender	Male	0.01	0.03	0.24	0.81	0.81
	Female	–0.14	0.05	–2.94	**0.01***	**0.03***
Task familiarity	Some or none	–0.07	0.04	–2.05	**0.04***	0.14
	Completely familiar	–0.02	0.04	–0.40	0.69	0.80
Crowd-working frequency	Several times a day	–0.03	0.03	–1.07	0.29	0.44
	Once a day	–0.07	0.07	–1.01	0.32	0.44
	A few times a month	–0.09	0.06	–1.40	0.13	0.17

DISCUSSION

Although the exploratory results of *Study 1* suggested that gamification might benefit certain groups, such as female or inexperienced participants, these trends were not confirmed in *Study 2*. The only identified effect of gamification, i.e., a slightly lower performance among female workers, was small. These findings indicate that the inclusion of gamification in the examined crowdsourcing task does not lead to robust or reliable improvements in enjoyment, engagement, or performance—either in the overall sample or within demographic subgroups—suggesting that the effects observed in *Study 1* were likely sample-specific or smaller than initially estimated. This contrasts with prior work reporting mixed effects: for example, Walter et al. (2022) reported increased contributions in a gamified labeling task, while Carlier et al. (2016) observed substantial losses in task quality in a gamified segmentation task.

An important limitation is that we only used a single task, i.e., a car annotation task. This choice was motivated by the fact that recognizing, identifying, and tagging objects are among the most common types of microtasks in crowdsourcing (Christoforou et al., 2025; Gadiraju et al., 2014; Morschheuser et al., 2017). However, the simplicity of this task may have outweighed any potential effects of gamification: even workers excluded from the main analysis due to questionnaire inconsistencies—rather than task performance—performed accurately (Median = 0.82, IQR = 0.20), indicating that the task was already straightforward. Therefore, additional gamification elements may have been unnecessary or even distracting, highlighting that task clarity and design may be more important for workers than external game elements.

Another limitation is that engagement was measured only in the short term. Although we observed no effect of gamification in the current tasks, results might differ in continuously available crowdsourcing tasks, where sustained motivation and retention could play a larger role. Future research could explore whether these findings generalize to more complex or tedious tasks.

CONCLUSION

In this work, we investigated the effects of gamification on crowdworkers' enjoyment, engagement, and performance in an image annotation microtask. Across two studies, we found that gamification did not lead to robust improvements, suggesting that the task's simplicity and clarity may be more important for worker outcomes than additional gamified elements. These findings indicate that well-designed, straightforward microtasks can sustain high performance and engagement without requiring extrinsic motivational features, providing guidance for the design of future crowdsourcing tasks.

REFERENCES

Carlier, A., Salvador, A., Cabezas, F., Giro-i-Nieto, X., Charvillat, V., & Marques, O. (2016). Assessment of crowdsourcing and gamification loss in user-assisted object segmentation. *Multimedia Tools and Applications*, *75*(23), 15901–15928. https://doi.org/10.1007/s11042-015-2897-6

Christoforou, E., Demartini, G., & Otterbacher, J. (2025). Crowdsourcing or AI Sourcing? *Communications of the ACM*, *68*(4), 24–27. https://doi.org/10.1145/3706101

Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., & Allahbakhsh, M. (2018). Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys*, *51*(1). https://doi.org/10.1145/3148148

Deterding, S., Khaled, R., Nacke, L., & Dixon, D. (2011). Gamification: Toward a Definition. *Proceedings of the CHI 2011 Gamification Workshop*, 12–15.

Estellés-Arolas, E., & L. Guevara, F. G. (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science*, *38*. https://doi.org/10.1177/0165551512437638

Gadiraju, U., Kawase, R., & Dietze, S. (2014). A Taxonomy of Microtasks on the Web. *Proceedings of the 25th ACM Conference on Hypertext and Social Media, HT '14*, 218–223. https://doi.org/10.1145/2631775.2631819

Gamboa, E., Bauer, M., Khawaja, M. R., & Hirth, M. (2022). Rather Multifaceted than Disruptors: Exploring Gamification User Types of Crowdworkers. *Human Factors in Virtual Environments and Game Design*. https://doi.org/10.54941/ahfe1002069

Gamboa, E., Niayeshnia, F., Daneshi, S., & Hirth, M. (2026). *Crowdworker Gamification Study Data – Game On? Exploring Subgroup Differences (Study 1 & Study 2)*. Zenodo. https://doi.org/10.5281/ZENODO.18481254

Green, K. (1972). Intrinsic Motivation. *Journal of Philosophy of Education*, *6*(1), 73–96. https://doi.org/10.1111/j.1467-9752.1972.tb00461.x

Hirth, M., Borchert, K., de Moor, K., Borst, V., & Hoßfeld, T. (2020). Personal task design preferences of crowdworkers. *Proceedings of the 12th International Conference on Quality of Multimedia Experience*, 1–6. https://doi.org/10.1109/QoMEX48832.2020.9123094

Jenadeleh, M., Heß, A., Del Pin, S. H., Gamboa, E., Hirth, M., & Saupe, D. (2024). Impact of feedback on crowdsourced visual quality assessment with paired comparisons. *Proceedings of the 16th International Conference on Quality of Multimedia Experience, QoMEX 2024*, 125–131. https://doi.org/10.1109/QOMEX61742.2024.10598256

Morschheuser, B., Hamari, J., Koivisto, J., & Maedche, A. (2017). Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human Computer Studies*, *106*, 26–43. https://doi.org/10.1016/j.ijhcs.2017.04.005

Rahmanian, B., & Davis, J. G. (2014). User interface design for crowdsourcing systems. *Proceedings of the Workshop on Advanced Visual Interfaces*, 405–408. https://doi.org/10.1145/2598153.2602248

Soprano, M., Roitero, K., Gadiraju, U., Maddalena, E., & Demartini, G. (2024). Longitudinal Loyalty: Understanding The Barriers To Running Longitudinal Studies On Crowdsourcing Platforms. *ACM Transactions on Social Computing*, *7*(1–4), 1–49. https://doi.org/10.1145/3674884

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The gamification user types Hexad scale. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 229–243. https://doi.org/10.1145/2967934.2968082

Walter, V., Kölle, M., & Collmar, D. (2022). A Gamification Approach for the Improvement of Paid Crowd-based Labelling of Geospatial Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *V–4–2022*(June), 113–120. https://doi.org/10.5194/isprs-annals-V-4-2022-113-2022