Human Factors in Virtual Environments and Game Design, Vol. 50, 2022, 125–133
https://doi.org/10.54941/ahfe1002069

Rather Multifaceted than Disruptors: Exploring Gamification User Types of Crowdworkers

Edwin Gamboa, Mathias Bauer, Muhammad Reham Khawaja, and Matthias Hirth

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

ABSTRACT

Crowdsourcing allows individuals and organizations to outsource tasks to an anonymous group of individuals called crowdworkers, who are paid when completing the tasks. The quality of task results depends on factors like task complexity, task instruction quality, and crowdworker-related aspects, like their motivation towards the task. In this context, gamification, i.e., the use of game elements in non-ludic contexts, could foster the crowdworkers' motivation besides monetary incentives. Nevertheless, identifying the audience's gamification preferences is important to maintain their motivation in the long term. Although gamification has been used in crowdsourcing, to the best of our knowledge, it has not been based on crowdworkers' preferences. The User Types HEXAD scale is suitable to identify those preferences and is designed explicitly for gamification. In our research, we investigated which HEXAD user types characterize crowdworkers. To aim this, we conducted a large-scale user study on two well-known crowdsourcing platforms, Amazon Mechanical Turk and Microworkers. Crowdworkers completed a demographic questionnaire, an image annotation task, and the HEXAD scale. The results show that crowdworkers are a multifaceted audience since all user types are exhibited. Therefore, we argue that one gamification approach might already satisfy a broad range for crowdworkers.

Keywords: Crowdworkers motivation, HEXAD user types, Gamification

INTRODUCTION

Gamification is defined as the use of game elements in non-game contexts, such as education or work (Deterding et al., 2011). There is evidence that gamification might foster crowdworkers' motivation (Morschheuser et al., 2017). However, to the best of our knowledge, gamification has not been designed based on crowdworkers' characteristics, needs, and sources of motivation. A first step to aim this is to identify what kind of user types the crowdworkers are, so that tailored gamification can be applied (Hallifax et al., 2019). In this context, the User Types HEXAD scale (Tondello et al., 2016) is suitable to characterize the crowdworkers' sources of motivation in gamification since it is built upon the Self Determination Theory (SDT)

(Ryan & Deci, 2000). The scale allows to identify to what extent people exhibit traits of Philanthropists, Socializers, Free spirits, Achievers, Disruptors or Players.

Therefore, we employ the HEXAD scale to identify the user types that characterize crowdworkers regarding their gamification preferences in two well-known crowdsourcing platforms. To aim this, we first analyze the HEXAD user types within each platform. Then, we compare the platforms to investigate whether each platform present different patterns regarding the user types. Finally, we study whether there is a relationship between the demographics of the crowdworkers and the user types.

The remainder of this paper is structured as follows. First, we briefly present the SDT, the HEXAD user types, and the concept of crowdsourcing. Then, we detail the methodological setup of our study. After this, we describe the demographics of the participants, the observed user types, and the relation of demographics and user types. Finally, we discuss our findings, outline future research, and conclude the paper.

BACKGROUND

Self-Determination Theory

In SDT the individual self is understood as the central control mechanism of autonomous and self-determined behavior (Heckhausen & Heckhausen, 2018). The theory suggests three basic needs. First, autonomy corresponds to the self-determination: a person has an individual experience as being his or her own center of action with own goals and resulting activities. Second, competence refers to the perceived efficacy of actions. Third, relatedness corresponds to the acceptance of own actions in interaction with other people (Adams et al., 2017). Furthermore, the organismic integration theory as part of SDT states the grade of perceived autonomy within a given situation, and results in different motivation types of intrinsic and extrinsic regulation (Ryan & Deci, 2000). Intrinsic regulation refers to the inherent need for personal growth and is associated with merging in an activity. Autonomy and competence are ideally addressed in this regulation style. The different regulation types of extrinsically motivated behavior, i.e., external, introjected, identified and integrated, refer to actions under a specific goal with an instrumental function, e.g. successfully completing a study program for better career perspectives (Heckhausen & Heckhausen, 2018). The four types of extrinsic regulation differ in their degree of perceived control and autonomy. The result is a continuum with external and introjected regulation as more controlled types, and identified, integrated and intrinsic regulation as more self-determined types (Ryan & Deci, 2000).

The HEXAD User Types

The User Types HEXAD scale was presented in (Tondello et al., 2016) and validated in (Tondello et al., 2019). It is based on the SDT and allows identifying to what extent a person exhibits six user types. First, philanthropists, who are motivated by purpose and meaning. They are altruistic and want to help others without expecting an external reward. Second, socializers, who are motivated by relatedness since they want to interact with others to create social connections. Third, free spirits, who are motivated by autonomy and self-expression. Their goal is to create and explore new things. Fourth, achievers, whose motivation is mastery, i.e., learn new things and improve themselves. Fifth, players, who look for external rewards and will do the necessary to collect them in a system. Finally, disruptors, who want to initiate positive or negative change in systems directly or indirectly.

Crowdsourcing

Crowdsourcing allows an individual or an organization to propose a heterogeneous and anonymous group of individuals, called crowdworkers, to complete tasks via the internet. After successful completion, the crowdworkers receive an economic reward (Estellés-Arolas & L. Guevara, 2012).

In crowdsourcing, and particularly in micro-tasking, workers are often asked to annotate content, fill out surveys, give subjective feedback, and complete or correct data. The quality of the results depends on different aspects like task difficulty, quality of instructions, task duration, and crowdworkers' abilities and motivation (Daniel et al., 2018). Thus, crowdsourcing implies, among others, including reliability checks to filter unreliable crowdworkers (Hossfeld et al., 2014), designing unambiguous instructions (Khanna et al., 2010), and optimizing the user interface (Hirth et al., 2020; Rahmanian & Davis, 2014) to meet crowdworkers' needs. Our work contributes to this by providing insights on which user types are most frequent among the crowdworkers, and, consequently, provides insights which gamification methods might increase the motivation of the workers.

STUDY DESCRIPTION

Our goal is to identify crowdworkers' user types via the HEXAD scale. Although a survey is the straightforward method for collecting the data, research has shown that surveys attract a dedicated group of crowdworkers and leads to biased results. Thus, we attached the questions to a simple image annotation task since it is known to the crowdworkers (Yuen et al., 2012). The task includes five images from the BDD100K database (Yu et al., 2020). Each image is divided into a grid of boxes as shown in Figure 1. The crowdworkers should select the boxes containing cars or parts of it. The task includes a training session, in which crowdworkers should annotate 3 images. The system explains when and why an annotation was wrong. After completing the image task, the crowdworkers fill out the HEXAD scale (Tondello et al., 2016). Also, we ask the crowdworkers to complete a demographic questionnaire.

Since unreliable submissions are frequent in crowdsourcing (Hossfeld et al., 2014), we perform checks to filter unreliable crowdworkers. First, we exclude those who selected a location country that does not match the selected continent in the demographic questionnaire. Also, we exclude crowdworkers who report to be under 18 years old. Moreover, we exclude crowdworkers, who selected the same answer option, e.g., left-most option,

Figure 1: Image grid presented to crowdworkers, who should select boxes containing cars or car parts. The image belongs to the BDD100K database (Yu et al., 2020).

for all questions in the HEXAD scale. Finally, we also exclude crowdworkers who started the study, but did not finish it.

EVALUATION

Participants

Our task was published on Microworkers[1] (MW) and MTurk[2] (MT) platforms in November 2021 and paid 0.50 USD. In total, 322 crowdworkers from MW and 226 from MT participated in the study. We identified 276 (85.71%) valid crowdworkers from MW and excluded 46 (14.29%) participants since 16 (4.97%) selected an inconsistent country, 3 (0.93%) reported to be under 18, 23 (7.14%) did not pass the check in the HEXAD scale, and 7 (2.17%) did not finish the study. Some crowdworkers failed more than one check, thus, there is an overlapping in the quantities. In the case of MT, 204 (90.27%) crowdworkers were reliable. We excluded 22 (9,73%) crowdworkers, 16 (7.08%) due to the country check, and 6 (2.65%) who did not finish the study. Table 1 shows the demographics of the reliable participants for both platforms. The data collected from the crowdworkers is available online (Gamboa et al., 2022).

Dominant User Types

We calculated the HEXAD user type scores for each participant as in (Tondello et al., 2016). The score for each user type ranges from 4 to 28. To determine whether there is a dominant user type, we compared the user type

[1] www.microworkers.com Accessed February 2022

[2] www.mturk.com Accessed February 2022

Table 1. Overview of participants' demographic data for both platforms.

Variable	Value	No. of participants MW, n = 276	No. of participants MT, n = 204
Gender	Female	147 (53.26%)	59 (28.92%)
	Male	128 (46.38%)	144 (70.59%)
	Other	0 (0.00%)	0 (0.00%)
	Not said	1 (0.36%)	1 (0.49%)
Age	18-25	116 (42.00%)	52 (25.49%)
	26-40	142 (51.4%)	131 (64.21%)
	Over 40	18 (6.50%)	21 (10.29%)
Continent	Africa	25 (9.10%)	0 (0.00%)
	America	50 (18.10%)	94 (46.10%)
	Asia	180 (65.20%)	105 (51.50%)
	Australia	0 (0.00%)	0 (0.00%)
	Europe	21 (7.60%)	5 (2.50%)
Highest degree	< High school	2 (0.72%)	0 (0.00%)
	High school	54 (19.57%)	10 (4.90%)
	Bachelor	162 (58.70%)	155 (76.00%)
	Master	57 (20.65%)	39 (19.10%)
	Ph.D.	1 (0.36%)	0 (0.00%)
Employment	Unemployed	78 (28.26%)	4 (1.96%)
	Part time	117 (42.39%)	12 (5.88%)
	Full time	79 (28.62%)	187 (91.67%)
	Retired	2 (0.72%)	1 (0.49%)
Frequency of platform check	Hardly ever	21 (7.60%)	3 (1.50%)
	A few times a week	30 (10.90%)	48 (23.50%)
	Once a day	54 (19.60%)	93 (45.60%)
	Several times a day	171 (62.00%)	60 (29.40%)

scores within each platform. Then, we compared the scores between the platforms to investigate whether the platforms show different score patterns. Figure 2 presents an overview of the results.

For the within-platform comparisons, we use a Friedman rank sum test since the data is not normally distributed and is dependent within the platform itself. We found that there are significant differences in the scores per participant within both platforms (MW: $\chi^2=441,\ df = 5,\ p<0.001$, MT: $\chi^2=101,\ df = 5,\ p<0.001$). The Friedman post-hoc tests show a similar pattern within both platforms, i.e., the Philanthropist score per participant is significantly higher than the other scores (MW: *Median* = *26.00*, MT: *Median* = *24.00*). In contrast, the Disruptor score per participant was significantly lower than the other scores (MW: *Median* = *17.00*, MT: *Median* = *21.00*).

To compare the scores between the platforms, we use Wilcoxon rank-sum tests since the data is independent but not normally distributed. We found that the scores per participant are significantly higher in MW ($p<0.001$) for Philanthropist (MW*: Median* = *26.00*, MT: *Median* = *24.00, W* = *41060*), Socializer (MW: *Median* = *25.00*, MT: *Median* = *23.00, W*=*40698*), Free

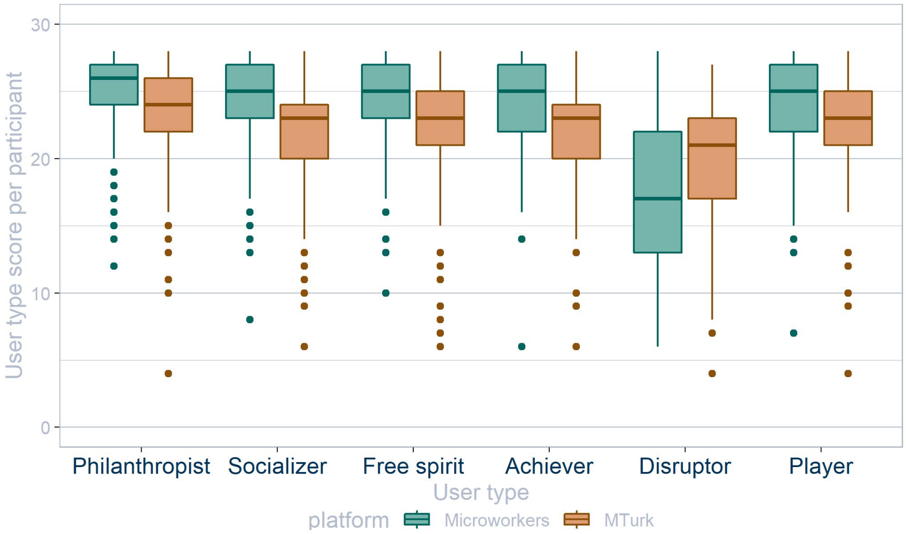

Figure 2: Overview of the HEXAD user types per participant identified for both platforms. The boxplot shows the median, interquartile range, and maximum and minimum values.

spirit (MW: $Median = 25.00$, MT: $Median = 23.00$, $W = 39317$), Achiever (MW: $Median = 25.00$, MT: $Median = 23.00$, $W = 39651$), and Player (MW: $Median = 25.00$, MT: $Median = 23.00$, $W = 37812$). In contrast, the Disruptor score was significantly lower ($W = 19498$, $p<0.001$) in MW ($Median = 17.00$) than in MT ($Median = 21.00$). This suggests that although the user type scores have a similar pattern in both platforms, the pattern is slightly more pronounced in the MW platform.

User Types and Demgographic Characteristics

Next, we performed a best subset regression test to determine whether the demographic variables (see Table 1) are significant predictors of the user type scores. We use the for Schwarz Bayesian Criteria to select the best model to avoid including insignificant predictors in the regression, and we did not include platform as a possible predictor since the previous results already suggest a significant difference between the scores of the platforms. These results are summarized in Table 2.

Both the continent in which crowdworkers are located, and the frequency in which they check the platform explained 11% of variance in the Philanthropist score, and 9% in the Free spirit score. Being in America significantly predicted and increase in the score of both user types. Nevertheless, being in Europe and checking the platform once a day significantly predicted lower scores.

The continent alone explained 7% and 5% of the variance of Socializer and Disruptor scores respectively. This time, being in America predicted a decrease in the Socializer score, but an increase of the Disruptor score. Additionally, being in Asia predicted higher Disruptor scores, while being in Europe predicted lower Socializer scores.

Table 2. Regressions of demographic predictors of user type scores for both platforms, n = 480.

	PH	SO	FS	AC	DI	PL
Constant	26.37	24.76	25.09	24.13	14.60	24.96
	(0.90)	(0.73)	(0.97)	(0.74)	(1.0)	(0.73)
Continent $_{America}$	−2.30***	−2.46**	−1.62*		2.90**	
	(0.69)	(0.79)	(0.74)		(1.10)	
Continent $_{Asia}$	−0.76	−0.63	0.01		4.50***	
	(0.66)	(0.76)	(0.71)		(1.10)	
Continent $_{Europe}$	−2.92**	−3.88***	−2.45*		1.60	
	(0.89)	(1.02)	(0.95)		(1.50)	
Platform check	−1.36		−1.18	−1.84*		−2.42**
A few times a week	(0.74)		(0.80)	(0.85)		(0.83)
Platform check	−1.64*		−2.02**	−1.80*		−2.44**
Once a day	(0.70)		(0.75)	(0.80)		(0.79)
Platform check			−0.38	−0.02		−0.88
Several times a day			(0.73)	(0.78)		(0.77)
F (6, 473)	10.00***	14.00***	9.00***	9.70***	8.90***	8.90***
R^2	0.12	0.08	0.10	0.06	0.05	0.05
Adjusted R^2	0.11	0.07	0.09	0.05	0.05	0.05

PH: Philanthropist, SO: Socializer, FS: Free Spirit, AC: Achiever, DI: Disruptor, PL: Player.
Standard errors are reported in parenthesis. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Finally, the frequency of platform check accounted for 5% of de variance of Achiever and Player scores. Checking the platform a few times or once a day predicted a decrease in the scores of these user types.

DISCUSSION AND CONCLUSION

The results about crowdworkers' exhibited user types suggest that there are small differences among the scores, being the Philanthropist score the most prominent and the Disruptor score the lowest one. These findings are in line with previous studies outside the crowdsourcing context (Tondello et al., 2016, 2019). Moreover, we found that there are small significant differences between the scores of the studied platforms. The regression results using demographics as predictors of the user type scores indicate that the location and frequency of checking a crowdsourcing platform might explain the differences between the platforms. To understand this, we can analyze the share of participants for the identified significant demographics predictors, i.e., continent and frequency of platform check. As shown in Table 1, the share is similar for both platforms except for *participants who are in America* and *check the platform once a day*. In the case of MT, almost half of the participants have these two demographics, which is high compared to MW, where this occurs in less than 20% of the participants. Considering that these two demographics significantly predict higher Disruptor scores and lower scores for the other user types, we argue that these demographics might account for the differences between the platforms.

Although the identified differences are statistically significant, except for the disruptor score in MW, the user type scores differ marginally from each other, all being above a median of 20. Therefore, we argue that crowdworkers exhibit characteristics of all HEXAD user types. This is contrary to what might be intuitively expected in a crowdsourcing context, in which crowdworkers are normally rewarded extrinsically. Hence confirming Mason and Watts' findings (2010), who argue that pure monetary compensation is not the only way to motivate crowdworkers. Our findings imply that applying gamification in crowdsourcing requires fostering intrinsic motivation sources such as relatedness, autonomy, mastery, and purpose.

Finally, for exploratory purposes, we studied the correlation between crowdworkers' HEXAD scores and their performance on the task. We did not find any significant result using Kendall's tests. In the future, we plan to study whether using gamification to address the user types, as proposed in (Tondello et al., 2016), influences the quality of crowdworkers' performance and their motivation towards a task.

In this work, we studied crowdworkers' gamification preferences using the User Types HEXAD scale with two crowdsourcing platforms. Our results show that crowdworkers are a multifaceted audience in which no user type is clearly predominant because all resulted in high scores. Thus, adding any gamification elements can already be a motivation source for a broad range of crowdworkers. Our findings can be used as a base to design gamified crowdsourcing tasks targeted at crowdworkers' preferences according to the HEXAD user types.

ACKNOWLEDGMENT

The authors would like to thank Muhammad Nouman Khan for his valuable contribution to the tool employed for collecting the data of this study.

REFERENCES

Adams, N., Little, T. D., & Ryan, R. M. (2017). Self-Determination Theory. In M. L. Wehmeyer, K. A. Shogren, T. D. Little, & S. J. Lopez (Eds.), *Development of Self-Determination Through the Life-Course* (pp. 47–54). Springer.

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), 1–40.

Deterding, S., Khaled, R., Nacke, L., & Dixon, D. (2011). Gamification: toward a definition. *Conference on Human Factors in Computing Systems*, 12–15.

Estellés-Arolas, E., & L. Guevara, F. G. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*.

Gamboa, E., Bauer, M., Khawaja, M. R., & Hirth, M. (2022). *Dataset for analysis of paper: Rather Multifaceted than Disruptors: Exploring Gamification User Types of Crowdworkers*. Zenodo.

Hallifax, S., Serna, A., Marty, J.-C., Lavoué, G., & Lavoué, E. (2019). Factors to consider for tailored Gamification. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 559–572.

Heckhausen, J., & Heckhausen, H. (Eds.). (2018). *Motivation and action*. Springer.

Hirth, M., Borchert, K., de Moor, K., Borst, V., & Hoßfeld, T. (2020). Personal task design preferences of crowdworkers. *International Conference on Quality of Multimedia Experience*.

Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., & Tran-Gia, P. (2014). Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing. *IEEE Transactions on Multimedia*, *16*(2), 541–558.

Khanna, S., Ratan, A., Davis, J., & Thies, W. (2010). Evaluating and improving the usability of Mechanical Turk for low-income workers in India. *ACM Symposium on Computing for Development*, 1–10.

Mason, W., & Watts, D. J. (2010). Financial incentives and the "performance of crowds." *ACM SIGKDD Explorations Newsletter*, *11*(2), 100–108.

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.

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

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Tondello, G. F., Mora, A., Marczewski, A., & Nacke, L. E. (2019). Empirical validation of the gamification user types HEXAD scale in English and Spanish. *International Journal of Human-Computer Studies*, *127*, 95–111.

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.

Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). BDD100K: A diverse driving dataset for heterogeneous multitask learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2636–2645.

Yuen, M.-C., King, I., & Leung, K.-S. (2012). Task recommendation in crowdsourcing systems. *Proceedings of the International Workshop on Crowdsourcing and Data Mining*, 22–26.