RE-PLAYING AND QUALITY CONTRIBUTION: THE ROLE OF THE SCORE MECHANISM DESIGN AS MOTIVATOR

The role of the score mechanism design as motivator

GANIT RICHTER, DAPHNE R. RABAN, AND SHEIZAF RAFAELI

Abstract

The current study investigates how the game points allocation function affects quality of responses and replaying behavior in a game-for-crowds implementation. In a series of experiments, we compare between two scoring designs across two types of users: students and public. Two reward-based mechanisms differing in the mathematical function were applied: linear (y=3x) vs. exponential (y=6ex). Findings highlight the nuances between repeated game rounds, types of users and quality contribution. We find support for the importance of the mathematical function of scores assignment as a motivator. Insights into the design of effective game-for-crowds for different target groups are discussed.

Introduction

Games-for-crowds and gamification

The rise of the digital game medium in entertainment has motivated its adoption for pursuits beyond entertainment (Gallagher, 2018; SIWEK, 2017). These games are known as serious games, and they are used in various aspects of everyday life such as training and knowledge sharing in many areas: defense, education, scientific exploration, health care, management, marketing, communication and politics (Lieberman, 2009; Ratan & Ritterfeld, 2009). Digital technology and the integration of the Internet into mainstream society made it possible for new serious-game environments to evolve. Of relevance to the present study are those computer environments associated with crowds, namely games-for-crowds.

Crowds are made up of independent individuals, each with their own objectives and behavior patterns. Games-for-crowds are a combination between the concept of "The Wisdom of Crowds[i]" and serious games. Consequently, games-for-crowds integrate computational tasks into networked games; people played the game to have fun and as a side effect of playing they helped to solve open problems, improve artificial intelligence, involve in projects on citizen science and more (Goh, Ang, Chua, & Lee, 2011; Khatib et al., 2011; Krause & Smeddinck, 2011; Quinn & Bederson, 2011). Similarly, to serious games, gamification is the application of game elements for purposes other than their expected use for entertainment (Deterding, Dixon, Khaled, & Nacke, 2011). Some researchers suggest that serious games are a subset of gamification (Kapp, 2012), others use the term to describe the addition of games into an existing non-game system, or converting a system into a game (Seaborn & Fels, 2015). Thus, the boundary between game and an artifact with game elements is blurry, subjective and social (Richter, Raban, & Rafaeli, 2015). Some well-known examples in this regard are: *Foldit* (Cooper, Khatib et al., 2010), *Phylo*[ii], and the series *Games With A Purpose* [iii](GWAP) (von Ahn, 2006). *Fold-it*, for example, exemplifies this blurriness as some reference it as a successful example of gamification in science, while others view it as a serious game for crowds in which players use a graphical interface to predict protein structures, a problem that computers cannot solve yet. The current research extends this direction utilizing an online game to enhance contribution of knowledge. For example, the game called *Guess*, introduced in this study, is an online knowledge pooling game for crowds. By using *Guess*, organizers can gauge their audience's sentiment or knowledge regarding the topic in question.

Games-for-crowds explore how games and games technology can increase participation to accomplish tasks at scale by engaging crowds (von Ahn, 2009). Arguably, the success of these system depends upon users' participation, quality of input, re-play and co-creation continuously (Majchrzak & Malhotra, 2013; Richter, Raban, & Rafaeli, 2018; Seaborn & Fels, 2015). So far, scant research has been conducted to link points allocation as motivators for quality and persistence of contribution. Our aim is to explore quality of contribution and re-playing behavior incentivized by two point-allocation functions implemented in two types of group-users (crowds): students and public. While the first group is fairly homogeneous, the second one consists of diverse users, as will be explained in the method section. We question the overall crowd type while discussing the effect of different point scoring systems on re-playing and quality of contribution. Understanding differences among target groups can provide new insights into the design of effective games-for-crowds and gamified systems.

The rest of the paper is structured as follows: after a brief introduction to the role of scores and points in games, we discuss studies in the domain of games-for-crowds and gamification that elaborate on the mechanisms to foster quality contribution. Following a description of *Guess*, a knowledge-pooling game for crowds, which serves as our research tool, we describe the experiment as well as the results of the study. Before concluding with a summary of findings, as well as highlight future research, we deliberate on key insights gained from this work.

Reward systems in games are generally tied to collecting points as a representation of progress. Next, we explain how rewards and incentives influence motivation in the special case of distributing points and scores.

Related Work

The role of scores and points

Scores are designed to deliver information to the player as part of an on-going motivational process. They are often used as a core reward for users' effort (Morschheuser, Hamari, & Koivisto, 2016); can facilitate competition; their motivational appeal is built upon their cumulative nature (Blohm & Leimeister, 2013). The scoring process offers numbers as a mirror of progression which, in turn, motivates the user to engage in self-evaluation (Sjöklint, Constantiou, & Trier, 2013). Yet, rewards can also have a negative impact on motivation: adversely affecting participation quality (Deci, Koestner, & Ryan, 1999) and lowering performance in complex and creative tasks (Toubia, 2006). Consequently, a positive and enjoyable experience can motivate and engage users in fulfilling tasks (Domínguez et al., 2013; Füller, 2010; Piller & Walcher, 2006; Zimmerling, Höllig, Sandner, & Welpe, 2018).

Computers can implement complex arithmetic and logical rules, and therefore can offer a large variety of scoring mechanisms. Combined with the ability to monitor game decisions and performance, the computerized environment is well-suited for developing behavior-based scoring mechanisms. Yet, the design of these systems is relatively seldom researched, a gap we wish to narrow (Richter et al., 2018).

Points and scoring mechanisms as design elements to foster quality of contribution

The most common purpose behind using a game for crowd or gamified approaches is to encourage behavior change, i.e. increase participation, improve performance, or quality of contribution (Seaborn & Fels, 2015). For example, Cooper et al. (2010) highlight the potential for encouraging quality contribution through online games to solve real-world problems; they used puzzles, ranking and accumulation of points to motivate users to find the best possible protein structures. Anderson et al.(2013) demonstrated that virtual rewards could motivate quantitative activities while using the question-answering site Stack Overflow; Von Ahn (2006) used human computation games to improve the accuracy of image search; participants evaluated gamification elements such as points and badges as motivating when used in an idea contest (Witt, Scheiner, & Robra-Bissantz, 2011). This approach rests on the assumption that game mechanics address user's psychological needs for relatedness, autonomy, and competence, and therefore positively influence motivation (Przybylski, Rigby, & Ryan, 2010).

Despite the growing use of games-for-crowds and gamification as a tool in various domains and contexts, there is still limited research on this topic, and developing a clear understanding of the positive or negative effects of the design of game elements is required. Furthermore, developing a stronger understanding for allocation of gamification elements, such as points, can contribute to the research regarding an optimal design of these platforms, to reach its objectives.

Previous research on points allocation refers to broad diversity of game elements: points, leaderboards, status, badges, and reports mixed results (Hamari, Koivisto, & Sarsa, 2014; Seaborn & Fels, 2015). It is difficult to isolate and measure the specific effect of scoring design (Richter et al., 2018). Liu et al. (2011) developed a mobile crowdsourcing application for image to-text translation. The application used several gamification features, including points and scores. They reported improvement in response speed and quality, yet how these results compared to a non-gamified version was unclear, and authors questioned the overall impact of gamification incentive on participation motivation.

Witt et al. (2011) examined the implementation of leaderboards and dual point systems: game points for completion of actions and social points for engaging in social behaviors (rating and commenting) in an online idea competition. Their results were mixed. Although gaining points motivated the introduction of further ideas, game elements were not consistently effective or well received, and questionnaire results evaluated their effect less strongly (they fell into a neither agree nor disagree category) (Witt et al., 2011). The authors attributed these results to the unsophisticated implementation of game elements that prevented its success.

Farzan et al. (2008) investigated the effect of points, levels, status and avatar on content contribution to a social networking site of an enterprise. A fixed number of points were awarded to each type of content contribution (photos, list of items, profile page). Researchers reported that the increase in contributions diminished shortly after the launch of the system. They attributed these results to the expected and no-dynamic nature of the system.

A scoring function that includes dividend points and dynamic update of player's contribution was introduced by Guy et al.(2011). Weekly update emails, indicating current ranking, and point missing for moving one ranking up, were used to support players' engagement. Yet, authors observed the same rapid decay of content contribution behavior. They suggested investigating refinements of the scoring function to obtain long- term engagement and quality contribution (Guy et al., 2011).

The above mentioned studies represent some of the attempts to understand the connection between points, scoring mechanism design and user behavior. They also pointed out several weaknesses: relying on a small sample size, implementation of several game elements simultaneously, not considering the varying response of participants to the incentives, count on one episode of play, missing a detailed description of the scoring mechanism, as well as suffer from lack of control group. It remains challenging to: 1) isolate and test different kinds of mechanical systems such as reward points, 2) discuss the design of game points allocation, 3) refine the link between point allocation and type of crowd, 4) investigate the scoring design to promote re-playing.

First steps of research in this direction can be seen in the work of Richter et al. (2018) who addressed the design of point scoring mechanisms to promote contribution of knowledge. Results indicated that implementing a point allocation mechanism promotes performance depending on the implementation, and that different types of crowds behaved differently in reaction to varying scoring conditions. The study involved three reward conditions: reward-free condition (control); and two conditions that differ in the algorithm applied across different crowds. In the current study, we expand this direction by looking at quality of information contributed and re-playing behavior as a function of game score design.

Points can be enhancers of quality, however, at the same time, they can catalyze generating spam just for the sake of earning points. Thus, we explore the balance between instrumental knowledge contributions and spam. To evaluate the effect of the scoring system we conducted a controlled study. In this context, we present three main research questions that summarize the goals of this study.

> RQ1: How do different score mechanism designs affect quality contribution in games-for-crowds?
>
> RQ2: What is the relation between different score mechanism designs and re-playing?
>
> RQ3: Do quality and re-play depend on the type of crowd participating?

Method

We compare two scoring designs (mathematical functions) across two types of crowds via a word association game for crowds called "*GUESS*" (originally developed by IBM). Following a description of the game we explain the experimental design and introduce the participants.

GUESS is an association game where knowledge accumulates by prompting questions simultaneously to all users who, in turn, receive points for responses. It is played in a 'Single Player' mode, where no direct interaction occurs between users. Yet the accumulated data is displayed visually, in real time, on users' screens. Game procedure:

- Users access a personalized home screen where they select a game to play from the available games.
- Each game consists of several short questions (example question: Name famous scientists).
- Each question appears for 60 seconds until the next question comes up.
- Users are encouraged to input as many responses as they can for each question. Responding is done by typing a response and pressing enter (Example responses: Einstein, Newton, Curie).
- Once users enter a response, they receive a feedback message: (1) 'You got X points'- indicating how many points were gained for the current response; or, (2) 'you already mentioned this answer' to inform users when they type a response that they have already mentioned.
- Players see responses graphically, on-screen, in the peripheral circles (see *Figure 1.*).
- Users can choose SKIP to move on to the next question (before the time expires or if they prefer not to respond a certain question).
- Previously given answers (by other users) are presented on screen yet they are hidden behind black dots. Once the user repeats such a response, the black dot opens, and the response appears.
- The game ends with a "game over" notification and a display of some game statistics and a leaderboard.

Figure 1. presents the main game interface: a question with an input box appears on top; responses are shown in the peripheral circles. The black dots denote responses given by other users. Additional on-screen information includes time remaining, user's statistics (dynamic update of current total points), basic game statistics (number of players, number of responses given by all players).

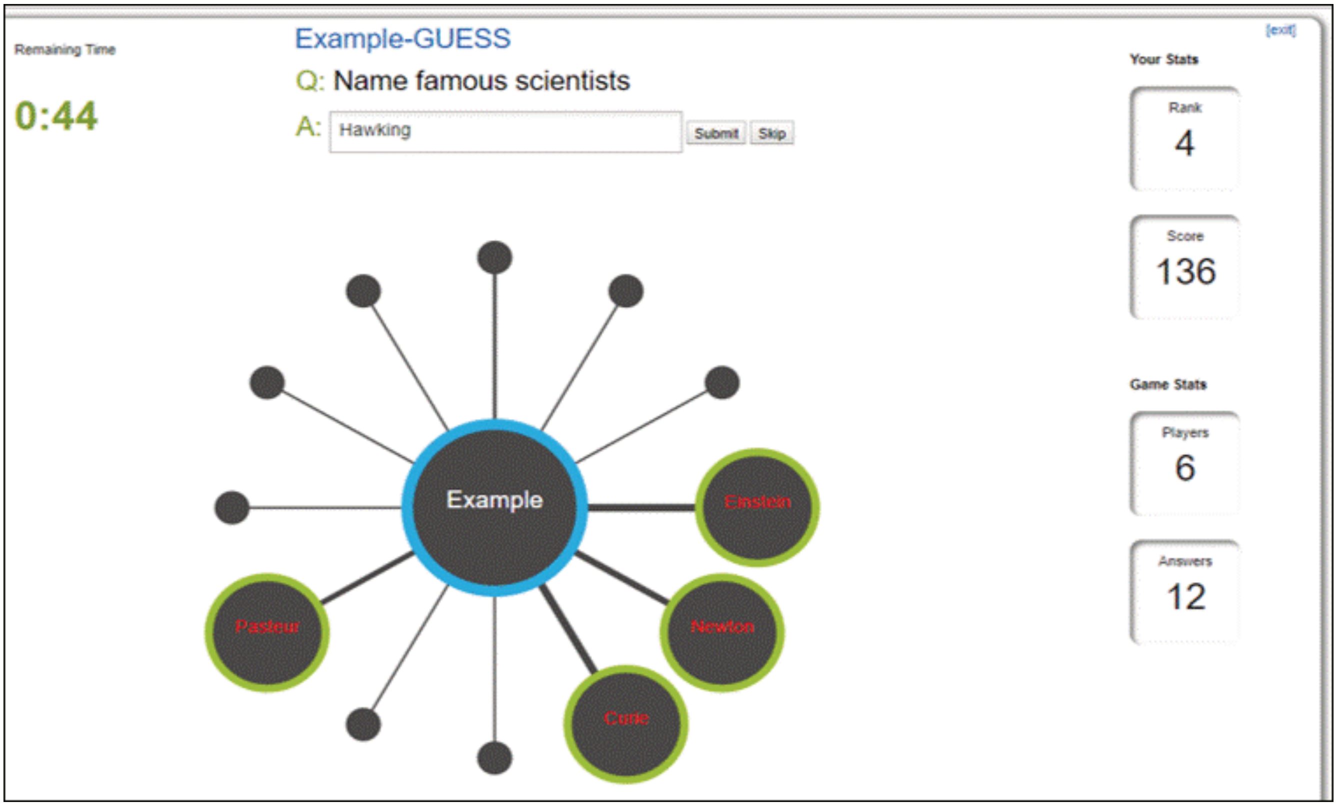

Figure 1. Main game interface

Designing the score mechanism

We implemented a new scoring wizard to allow fast adjustments of the algorithm applied. The goal was to compare two mathematical functions for the allocation of points. We compared usage of a monotonic and expected linear scoring mechanism with an irregular and incalculable (for the user) exponential scoring mechanism. The functions applied were: (1) linear function: $y=3x$ (2) exponential function: $y=6ex$. The linear function follows a simple arithmetic rule, while the exponential function introduces irregularity to the system. Rather than following formula strictly, we used an approximation with integer coefficients. Thus, the linear function y=3x, in which a player earns 3 points for each action, may be compared with the exponential function y=6ex (the series expansion at x=0 is y=6ex= 6+6x+3×2+x3+o(x4)). *Figure 2.* depicts both functions.

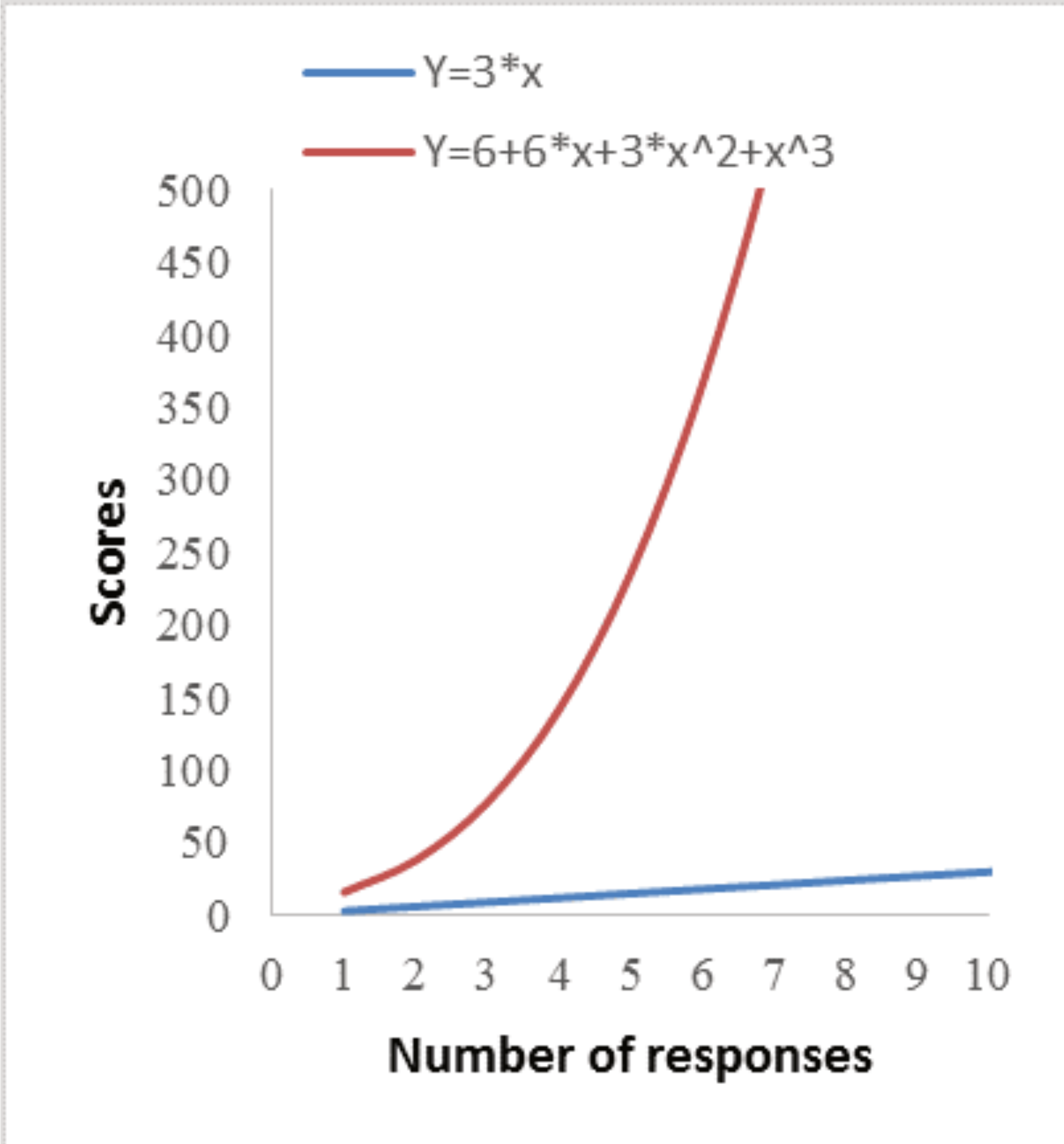

Figure 2. Point accumulation: linear vs. exponential function

Defining quality – Quality is calculated as the proportion of useful responses out of all responses. To clarify, 'useful responses' means plausible answers to a given question. For example, a plausible answer to a question about European capital cities would be Paris. Even typos or slight spelling errors were counted as useful, e.g. Pariz or Parris. However, writing 'abc' in response to the same question would count as a non-useful response.

Procedure– GUESS was deployed in two events of The European Researchers' Night which is a "popular science and fun learning event" open to the general public, and in three university courses at the Faculty of Management. Participation was voluntary. Participants were randomly assigned to one of the experimental conditions; two reward-based systems differing in the scoring function applied (linear vs. exponential). Questions were randomly displayed on users' screen.

The data collected includes 10 games, 5 games in each scoring condition. Group size (number of participants in a game) ranged from 10 to 57. A total of 7268 responses were gathered: linear score function 3318 responses, exponential score function 3950 responses. For the evaluation, we examined the server logs, which documented the details of each response provided by users in a game along with a time-stamp. Derogatory responses as well as meaningless letters and numbers were marked as spam.

Participants– A total of 352 participants were involved in the game; played at least once in one of the conditions. We identified two types of crowd: the general public and students distributed as defined in *Table 1.* We did not gather demographic data.

Types of crowd	**Definition**	**N**	**%**	Linear score mechanism N	Exponential score mechanism N
Public	Teenagers and older visitors attending The European Researchers' Night events, answering questions that require common knowledge	171	48.6	63	108
Students	Management students answering questions on management topics	181	51.4	97	84

Table 1. Participants types, size and distribution across condition

41 participants re-played the game without being asked to do that. The number of game rounds varied between 1 to 5 times. *Table 2.* shows the distribution of game rounds, across types of crowd and scoring design. The current study examined incentives for re-playing among the different types of crowd; therefore, in addition to analysis data of all users, we isolate participants' data across game-rounds, in order to look into participants who played the game the same number of times. Note that re-play was done voluntarily, not by design.

No. of game-rounds	N	%	Frequencies- types of crowd		Frequencies -scoring mechanism design	
			Public	Students	Linear score	Exponential score
1	311	88.35	155	156	142	169
2	33	9.37	15	18	16	17
3	5	1.42	1	4	1	4
4	2	0.57		2	1	1
5	1	0.28		1		1

Table 2. Participants frequencies according to type of crowd and scoring design

Results

We compare performance between point allocation conditions, linear or exponential, regarding the two types of users (crowds: public, students) across play rounds. First, we look at the 'big picture' as reflected in the collected data, then we focused on the first game round as compared to the others, later we investigate participants who played the game twice.

Quality**:** Since most responses were useful (58.02%) and variance was low, quality was defined as a dichotomy variable. A hierarchical logistic model was conducted to predict the probability for quality responses, i.e. all responses are useful. The analysis yielded that there was a main effect of score design ($F(1, 53) = 6.14$, $p < .05$). The probability that all responses are useful tends to be higher (in 0.54) in the linear condition than when the scoring function is exponential, across both types of crowd. There was no statistically significant difference between type of crowds, and there was no interaction effect of scoring condition*type of crowd.

Further analysis of responses that were not all useful revealed a main effect of type of crowd ($F(1, 17) = 8.39$, $p < .05$) indicating that there is an overall significant difference in means between the two crowds. The mean quality for students was significantly higher than the mean for public (M = 0.82, SD =0.24; M = 0.75, SD = 0.27; respectively). Analysis for game round revealed a main effect indicating that there is an overall significant difference in means between the first round and the other rounds ($F(1, 17) = 15.99$, $p < .001$). As the number of rounds increases, the quality decreases (M = 0.82, SD =0.22; M = 0.63, SD = 0.35; M = 0.58, SD =0.37; for the first, second and third rounds respectively).

Re-playing: A chi-square test of independence was performed to examine the relation between scoring design and replaying across the two type of crowds. The relation between these variables was significant in the student group, (*?*2(4)=12.12 , $p < .05$). For students, exponential point allocation design was more likely to promote re-playing than was linear scoring design.

To advance an understanding of replaying behavior and its relation to the scoring mechanism design, we investigate in isolation participants who played the game for the same number of times. However, since the sample size is very small with only one or two observations in some of the cells (*Table 2.*), we explore data of 41 participants from their second round. Due to the size of the data sets, we present descriptive statistics: 1) to provide basic information about the variables, and 2) to highlight potential relationships between variables. We add a basic performance metric, named contribution (# of responses per question) to get a more comprehensive description (Richter et al., 2018). *Table 3.* presents descriptive statistics along the two type of crowds and across the two-scoring design.

Scoring mechanism design	Variables	Public N=16 M (SD)	Students N=25 M (SD)
Linear scoring design (N=18)	Game rounds	2 (.00)	2.21 (.58)
	Contribution	3.25 (1.91)	5.26 (2.80)
	Quality	.98 (.05)	.85 (.28)
Exponential scoring design (N=23)	Game rounds	2.08 (.29)	2.73 (1.00)
	Contribution	6.49 (6.17)	5.46 (2.00)
	Quality	.69 (.42)	.90 (.15)

Table 3. Descriptive statistic according to type of crowd and scoring design

Discussion, Limitation and Future Research

Scores are collected in points that can be illustrated also in numbers of "likes", comments, shares, friends, and followers, or in the traditional way as plain numbers. Thus, designing a game scoring mechanism to support achieving objectives is a relevant practical issue, as well as an academic interest, not only in game environments (Guy et al., 2011; Hamari, 2013; Kankanhalli, Taher, Cavusoglu, & Kim, 2012). Current literature tends to explore game-systems, rather than elements in isolation (Hamari et al., 2014; Richter et al., 2018). This work narrows this gap, namely the need to isolate and test specific gamification elements.

Through a series of online experiments, we explore how different scoring mechanism designs have a differentiated effect on voluntary re-playing and quality items contribution. We compare usage of a monotonic and predictable linear function, with an exponential function which is more likely to be surprising and introduces irregularity to the system. We sampled from two different groups of users, the general public and students, exposing each to the different scoring conditions.

Findings offer first insights into the effect of scoring mechanism design, within a knowledge pooling game for crowds, on quality and re-playing. We used quantitative, unobtrusive data retrieved from game logs. Prior to discussing the implications of this work, it is necessary to recognize some inherent limitations of this study. With 41 participants who re-played the game, the sample is small, and the overall external validity may be questioned. In addition, re-playing was not the focus during the implementation of the system, yet it is an unexpected behavior worthy of attention relating to previous findings where we discovered a significant effect on engagement time, suggesting the exploring of quality of responses (Richter et al., 2018). Therefore, the study is exploratory, and findings are in tendency statements. Better understanding of re-playing behavior, awareness to differences between highly-motivated users and average users, may provide new insights into the design of effective games-for-crowds and gamified systems for different target groups and is much needed.

Overall, in answer to our research questions we can generalize the following assertions: 1. Implementing a linear point allocation mechanism promotes quality of responses independent of the type of crowds participating. 2. Re-playing tends to decrease quality. 3. Students show better performance than the general public regarding quality. 4. Scoring design affects re-playing behavior. 5. It is more difficult to mobilize students than the general public, yet, students generate high quality responses regardless of the point allocation mechanism. 6. The hardest challenge is to pool quality responses from the public. In the following we unpack the observations that led to these generalizations.

Analysis for quality yielded that quality was significantly higher for the linear condition than for exponential scoring. Investigating responses that were not all useful indicates a big difference among the two crowds, as students tend to contribute more quality responses. This means that it is more difficult to pool quality knowledge from the public compared to the students. Students appear to be indifferent toward the scoring conditions. In addition, they generated a low percentage of spam compare to the general public. Quality is also compromised due to re-playing. One explanation for that can be that the accuracy of responses may have decreased because players already know the questions and do not pay attention to typing, or by them deciding to spam.

Since there was a statistically significant association between scoring condition and game-round; that is, exponential scoring design boosts re-playing, it seems a good idea to look at all participants who re-played the game for the same number of times, in our case, twice.

It appears that exponential scoring design is more effective for students. They tend to re-play 1.2 times more in the exponential condition over the linear scoring, and 1.3 times higher than the general public in the exponential condition. Yet, a crowd of students is more difficult to mobilize than the general public in regard to content contribution, although students produce the higher quality in the exponential condition. Incentives in the form of points do not lead to quality impairment among highly motivated participants.

Working with heterogenous crowds (public) is challenging; incentives in the form of exponential points lead to an increasing in quantity yet also in generating spam. Contribution in the exponential condition was twice higher than in the linear scoring, but quality was higher for linear than for exponential scoring. This means that it is more difficult to pool quality knowledge from the public compared to the students, thus the linear condition seems to be the best option for public. Designers should take into consideration the low-quality rate and possibly rely on additional software to filter out spam. Further studies will take a deeper look and test on a larger scale game mechanics to confirm or disprove this effect.

To conclude, we test two different scoring mechanisms to determine which is better in driving quality of responses; furthermore, we suggest that exponential scoring encourages replaying behavior, and highlights the nuance between highly motivated users (students) and average users (public). To encourage knowledge contribution behavior via a game for crowds, care should be taken in selecting the scoring condition, and attention should focus on the type of crowd and kind of knowledge collected (common or specific knowledge). Caution must be exercised when using rewards (such as points) to encourage users' contribution in the long run, since the effect of rewarding on user behavior is complex.

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). Steering user behavior with badges. Paper presented at the *Proceedings of the 22nd International Conference on World Wide Web,* Rio de Janeiro, Brazil. 95-106.

Blohm, I., & Leimeister, J. M. (2013). Design of IT-based enhancing services for motivational support and behavioral change. *Business & Information Systems Engineering, 5*(4), 275-278.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., . . . Popović, Z. (2010). Predicting protein structures with a multiplayer online game. *Nature, 466*(7307), 756-760.

Cooper, S., Treuille, A., Barbero, J., Leaver-Fay, A., Tuite, K., Khatib, F., . . . Baker, D. (2010). The challenge of designing scientific discovery games. Paper presented at the *Proceedings of the Fifth International Conference on the Foundations of Digital Games,* 40-47.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627-668.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. Paper presented at the *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments,* Tampere, Finland. 9-15.

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education, 63*(0), 380-392. doi:http://dx.doi.org/10.1016/j.compedu.2012.12.020

Farzan, R., DiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W., & Brownholtz, E. A. (2008). Results from deploying a participation incentive mechanism within the enterprise. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Florence, Italy. 563-572. doi:10.1145/1357054.1357145

Füller, J. (2010). Refining virtual co-creation from a consumer perspective. *California Management Review, 52*(2), 98-122.

Gallagher, M. D. (2018). US video game industry revenue reaches $36 billion in 2017. Retrieved from http://www.theesa.com/article/us-video-game-industry-revenue-reaches-36-billion-2017/

Goh, D. H., Ang, R. P., Chua, A. Y. K., & Lee, C. S. (2011). Fight or unite: Investigating game genres for image tagging. *Journal of the American Society for Information Science and Technology, 62*(7), 1311-1324.

Guy, I., Perer, A., Daniel, T., Greenshpan, O., & Turbahn, I. (2011). Guess who?: Enriching the social graph through a crowdsourcing game. Paper presented at the *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems,* 1373-1382.

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236-245. doi:http://dx.doi.org/10.1016/j.elerap.2013.01.004

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?—A literature review of empirical studies on gamification. Paper presented at the *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS).* Hawaii, USA. 3025-3034.

Kankanhalli, A., Taher, M., Cavusoglu, H., & Kim, S. H. (2012). Gamification: A new paradigm for online user engagement. Paper presented at the *The 33rd International Conference on Information Systems, ICIS 2012,* Orlando. 1-10.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education* John Wiley & Sons.

Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popovic, Z., . . . Players, F. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences of the United States of America, 108*(47), 18949-18953. doi:10.1073/pnas.1115898108

Krause, M., & Smeddinck, J. (2011). Human computation games: A survey. Paper presented at the *Proceedings of 19th European Signal Processing Conference (EUSIPCO 2011),* Barcelona, Spain. 754-758.

Lieberman, D. A. (2009). Designing serious games for learning and health in informal and formal settings. In U. Ritterfeld, M. Cody & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 117-130) Routledge New York, NY.

Majchrzak, A., & Malhotra, A. (2013). *Towards an information systems perspective and research agenda on crowdsourcing for innovation* doi:https://doi.org/10.1016/j.jsis.2013.07.004

Morschheuser, B., Hamari, J., & Koivisto, J. (2016). Gamification in crowdsourcing: A review. Paper presented at the *49th Hawaii International Conference on System Sciences (HICSS), 2016,* 4375-4384.

Piller, F. T., & Walcher, D. (2006). Toolkits for idea competitions: A novel method to integrate users in new product development. *R&d Management, 36*(3), 307-318.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154-166.

Quinn, A. J., & Bederson, B. B. (2011). Human computation: A survey and taxonomy of a growing field. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Vancouver, BC, Canada. 1403-1412.

Ratan, R., & Ritterfeld, U. (2009). Classifying serious games. In U. Ritterfeld, M. J. Cody & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 10-24) New York/London: Routledge.

Richter, G., Raban, D. R., & Rafaeli, S. (2015). Studying gamification: The effect of rewards and incentives on motivation. *Gamification in education and business* (pp. 21-46) Springer International Publishing.

Richter, G., Raban, D. R., & Rafaeli, S. (2018). Tailoring a points scoring mechanism for crowd-based knowledge pooling. Paper presented at the *Proceedings of the 51st Hawaii International Conference on System Sciences,* Big Island, Hawaii. 1128-1137.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14-31.

SIWEK, S. E. (2017). *Video games in the 21st century: The 2017 report.* USA: Entertainment Software Association (ESA).

Sjöklint, M., Constantiou, I., & Trier, M. (2013). Numerical representations and user behaviour in social networking sites: Towards A multi-theoretical research framework. Paper presented at the *Proceedings of the 21st European Conference on Information Systems, ECIS 2013, Completed Research.* (144)

Toubia, O. (2006). Idea generation, creativity, and incentives. *Marketing Science, 25*(5), 411-425.

von Ahn, L. (2006). Games with a purpose. *Computer, 39*(6), 92-94.

von Ahn, L. (2009). Human computation. Paper presented at the *Proceedings of the 46th Annual Design Automation Conference,* San Francisco, California. 418-419.

Witt, M., Scheiner, C., & Robra-Bissantz, S. (2011). Gamification of online idea competitions: Insights from an explorative case. Paper presented at the *INFORMATIK 2011 – Informatik Schafft Communities 41. Jahrestagung Der Gesellschaft Für Informatik,* Technische Universität Berlin, Germany. 192.

112. Liu, T. Alexandrova, T. Nakajima, & V. Lehdonvirta. (2011). Mobile image search via local crowd: A user study. Paper presented at the *2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications, , 2* 109-112. doi:10.1109/RTCSA.2011.10

Zimmerling, E., Höllig, C. E., Sandner, P. G., & Welpe, I. M. (2018). Exploring the influence of common game elements on ideation output and motivation. *Journal of Business Research,* doi:https://doi.org/10.1016/j.jbusres.2018.02.030

[i] Some common examples of Wisdom of Crowds are information sites like Wikipedia and Yahoo! Answers.
[ii] http://phylo.cs.mcgill.ca/teaching/index.php
[iii] http://www.gwap.com/