

Tailoring a Points Scoring Mechanism for Crowd-Based Knowledge Pooling

Ganit Richter
University of Haifa
grichter@campus.haifa.ac.il

Daphne R. Raban
University of Haifa
draban@univ.haifa.ac.il

Sheizaf Rafaeli
University of Haifa
sheizaf@rafaeli.net

Abstract

We address the design of point scoring mechanisms in games for crowds, to promote user motivations to contribute knowledge. We measure the effectiveness of the scoring mechanism on users' performance across three types of crowd: general public, students in their field of study, general students. The conditions were: reward-free games (control group) and two reward-based systems differing in the algorithm applied (linear $y=3x$ vs. exponential $y=6e^x$). Results support the importance of the mathematical function of scores assignment as a motivator for knowledge contribution, and indicate that the effect of the scoring mechanism design should be tailored according to the type of crowd. These findings provide insights for designers of gamified systems on how to improve knowledge contributions in crowd-based systems.

1. Introduction

The success of knowledge pooling initiatives such as Wikipedia or Yahoo! Answers largely depends upon the motivation for participation of each contributor. The participation in crowd-based activities often follows a power law distribution, meaning that the crowd is heterogeneous in terms of motivation to contribute. However, the term 'crowd' in the literature often symbolizes a 'black box' of participants [1]. In this study we investigate the content of the 'black box' by probing the sensitivity of three types of crowd to various point scoring mechanism designs in a knowledge pooling assignment.

Gamification has emerged as a way to describe interactive online design that incorporates game-like approaches such as immediate feedback and virtual rewards (points, badges, gifts etc.), or status indicators (levels, progress bar, count of likes/friends/followers/retweets etc.). Gamification rapidly built momentum in industry [2] and academia creating a stream of research [3,4]. To date, gamification is applied in marketing as well as non-business contexts such as politics, health [5], problem

solving [6], work [7], online communities and social network [4]; computer science and engineering [8, 9] crowdsourcing [4, 6] and education [10]. Gamification brings opportunities but it is still unclear whether it is used properly in order to reach objectives [3, 4].

Gamification and crowdsourcing can be considered as interrelated as they share commonalities [11]: potentially large user numbers, requirements for scalability, potential for reusable functionality etc. Researchers see gamification as a way to incentivize people to voluntarily contribute to crowdsourcing tasks [3, 4, 11-13]. Still, these studies share several weaknesses: relying on a small sample size, simultaneous implementation of several gamification elements, lack of comparative examination of various types of crowd involved, disregard of individual characteristics. So far, scant research has been conducted to draw clear conclusions as to which specific game element would work better in certain situations and types of crowd. This dearth of understanding prevents organizations from adopting and designing effective gamification approaches.

The present study aims to: 1) shed light on the impact of the scoring mechanism design on players' behavior in a game designed for pooling knowledge. 2) examine the term crowds in regard to gamification while comparing different types of crowd in one gamification implementation 3) empirically evaluate scoring mechanism effectiveness as a tool for motivating and engaging users in non-entertainment contexts 4) strengthen the theoretical foundations of gamification.

After a brief introduction to gamification and crowds, the first part of this paper examines the motivational factors of scores and how scoring mechanism design can be associated with behavior. Following a description of Guess, a knowledge-pooling game for crowds, which serves as our research tool, we describe a repeated measures experiment to investigate how scoring patterns influence users' performance. We end this paper with a summary of findings, including suggestions for future research.

2. Background

2.1. Gamification and crowds

The development of social Web technologies enables harnessing large crowds of users for various tasks such as contribution of knowledge, data collection, and problem solving activities [14, 15]. The body of literature on crowdsourcing has been rapidly growing [11] offering four categories to describe crowdsourcing systems: crowdsolving, crowdcreation, crowd rating and crowdprocessing [16] and a conceptual framework of crowd capital [13, 17]. An active crowd of participants is crucial for such systems, raising the question of providing tailored incentives that will promote user motivation [18, 19]. Traditional approaches focus on economic approaches including tangible incentives such as monetary prizes [4, 11, 18, 19], algorithmic approaches such as improving software [20], encouraging social-psychological perspectives on the notion of gaining reputation [21], or expression of open-source ideology [22]. A recent direction is offering the enjoyable qualities of gameplay in non-game systems, coined as gamification [23]. Consequently, gamification can be seen as an aspect of incentives for people to voluntarily contribute to crowdsourcing tasks. Well-known examples in this regard are: Foldit¹ [24], Phrase Detective² [25], Phylo³, DARPA Formal Verification program⁴ and Games with a Purpose (GWAP) [26].

Gamification has become a popular strategic instrument to engage people in a given activity, to influence behavior and increase performance and productivity in various contexts [12, 27, 28]. It is one of the major instruments for driving users' motivation in crowdsourcing systems [4], i.e. "Transforming Homo Economicus into Homo Ludens" [29].

Interestingly, most studies combining crowdsourcing and gamification consider the term 'crowd' as a uniform concept that varies in amount rather than type, referring to an unspecified crowd [1]. Early empirical studies in this direction suggest to examine effects of gender and age on gamification effectiveness [4, 30]. Our study questions the overall crowd type rather than its particular composition. Given the gamified system, our focus is on differences in gamification effectiveness among three types of crowd: general public (GenPub), general

students (GenStu), and expert students (ExpStu), defined in Table 2.

2.2. Motivation theories

We test score mechanism design as motivator for sharing knowledge in a game for crowds, drawing on Csikszentmihalyi's Flow theory [31] and self-determination theory [32] to examine motivational effects in crowdsourcing [33] and gamification [3, 4, 12, 34]. Scores can be regarded as providing feedback which is an important antecedent to flow and engagement, and are mapped to the competence aspect of self-determination theory [4, 12, 27]. According to the game design elements taxonomy provided by Blohm and Leimeister, mechanisms such as scoring systems or badges create dynamics of collection which satisfy achievement motivation [27]. Thus, the motivational appeal of points is based upon their cumulative nature. By adding a certain number of points for completed actions to users' accounts the game keeps users engaged and encourages them to remain active [26, 35]. Although scoring mechanism design is a key component in making engaging games, only few empirical studies to date investigate their role on users' motivation and behavior [36-41]. The current study aims to address this gap.

Motivation is usually divided to intrinsic and extrinsic motivational factors. Intrinsic motivation applies to doing something because it is inherently interesting or enjoyable, while extrinsic motivation refers to doing something because it leads to a separable outcome [42]. In contrast, self-determination theory (SDT) defines intrinsic and subtypes of extrinsic motivation as falling along a continuum of internalization [32]. Cognitive evaluation theory (CET) and organismic integration theory (OIT) two sub-theories of SDT are of great importance in the study of gamification. According to CET, feedback, which is perceived as informing one's mastery, supports the need for competence and, subsequently, enhances intrinsic motivation. However, if feedback is perceived as controlling, then it reduces intrinsic motivation [42]. Within OIT, supports for autonomy and relatedness are critical to internalization; individuals can experience a sense of autonomy even when behaviors are extrinsically rewarded [42]. The idea of contributing to collective good; voluntary participation and full control on one's action; feedback messages and scores to develop competence and self-competition are the implementation of SDT in our study.

¹ <https://fold.it/portal/>

² <https://anawiki.essex.ac.uk/phrasedetectives/>

³ <http://phylo.cs.mcgill.ca/#1/EN>

⁴ <http://www.scientificamerican.com/citizen-science/darpa-verigames-crowdsourced-formal-verification-csfv-project/>

2.3. Motivational factors of points and scores

Within games, points play various roles: as a scoring system, a progression indicator, scale of rank, goal setting tool or even as a currency [12, 43]. Removal of a points-based incentive system can make a negative impact on players' activity; it may reduce overall participation [36]. Points add excitement to games by creating an artificial ladder from which players can fall if they make a mistake [39]. Points stimulate self-regulation and self-efficacy by providing direct input on performance, and thus afford regulating and monitoring performance more accurately [12]. The social effect of points ranges from status earned by performing certain actions up to reputation that is based on ratings received by others. Therefore, the motivational aspect of points is outlined with the help of social motivations as well as needs-based theories and rewards-based theory [12]. Gamification in crowdsourcing often uses a metric (scores) as a core reward for users' effort [11]. Further analyzing the scoring mechanism in regard to the four types of crowdsourcing systems based on [16] reveals that in most crowd-processing, crowd rating and crowd solving settings, the scoring mechanism design measures task fulfillment. Only few studies to date have provided an explanation or describe how the scoring mechanism is actually designed in crowd-created settings [11, 38, 40, 44]. Consequently it remains a question to explore.

3. Research questions

The lack of comprehensive understanding of scoring mechanism constrains the design of effective incentive systems for crowdsourcing. To evaluate the effect of the scoring system we designed a controlled study. In this context, we present three main research questions that summarize the goals of this study.

RQ1: Does assigning points affect performance in games for crowds? If so, how?

RQ2: How do different score mechanism designs affect performance in games for crowds?

Further, both crowdsourcing and gamification can take a variety of forms, and it would be myopic to assume that the same gamification implementations would function similarly across different types of crowd and different crowdsourcing approaches.

RQ3: Does the effect of the scoring design on users' performance depend on the type of crowd participating?

4. Method

4.1. GUESS: a game for crowds

We used a crowd- based knowledge pooling word association game called "Guess" which was developed by IBM⁵. Knowledge accumulates by prompting questions simultaneously to all users who, in turn, receive points for responses.

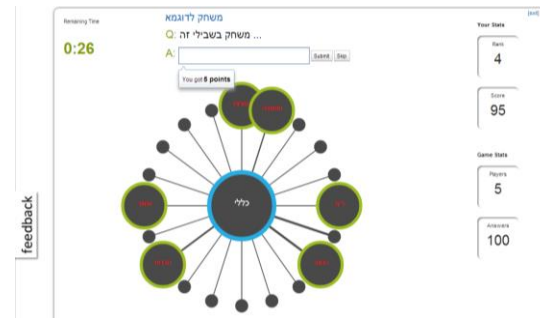


Figure 1. Main game interface

4.1.1. User interface. Users access a personalized home screen where they select a game to play from the available games. Figure 1 presents the main game interface. A question with an input box appears on top (example question: Name famous scientists). Users are encouraged to submit as many responses as they can by typing them in the input box within the time frame of 60 seconds per question (Example responses: Einstein, Newton, Curie). By pressing the SKIP button users can move on to the next question before the time expires or if they prefer not to respond. Responses are presented in the peripheral circles (see Figure 1). The black dots denote responses given by other users. Once the user types a response that had been already given by another user, the black dot opens and the response appears, however, there is no direct interaction between users [1]. 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). Once users enter a response they receive a feedback message. We used 3 types of text messages: (1) 'You got X points'- indicating how many points were gained for the current response; (2) 'you already mentioned this answer' to inform users when they type a response that they have already mentioned; (3) acknowledging submission by 'submitted ABC' (ABC= response) specifically for reward-free games.

⁵ GUESS- Gaming Umbrella for Enterprise Social Sourcing
<https://www.research.ibm.com/haifa/projects/imt/social/guess.shtm>
1

The game ends with a "game over" notification and a display of some game statistics and a leaderboard.

4.1.2. Designing the score mechanism. For this study a new scoring wizard was implemented in the game in order to allow fast adjustments of the algorithm applied. The goal was to compare reward-free gaming with external rewards in the form of points. Beyond that we compared usage of a monotonic and expected linear scoring mechanism with an irregular and incalculable (for the user) scoring mechanism. Three reward conditions were implemented: (1) reward-free condition (control); and two conditions that differ in the algorithm applied; (2) linear function: $y=3x$ (3) exponential function: $y=6e^x$. The linear score-keeping function provides the user a constant number of points for each submitted response. Because such functions are regular, users may become uninterested and quit the game. Part of the fun of games comes from their unexpected nature [45]. In order to make game scores less predictable, we introduced a new function using an exponential progression. Although both conditions present positive and ever growing scores, in the linear condition users can figure out fairly quickly how the scores are calculated and will know what to expect, while in the second condition the score is unexpected. Figure 2 illustrates point accumulations in both functions.

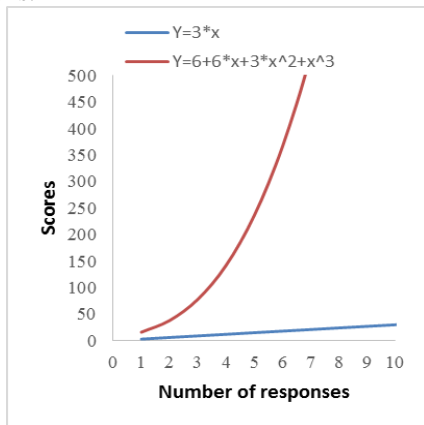


Figure 2. Point accumulation linear vs. exponential function

4.2. Operationalizing performance

Following common behavioral metrics of performance in gamification research we suggest to look at users' performance in two dimensional prism: the outcome and the process. Table 1 depicts these two dimensions along with some metrics GUESS produces for each user (log analysis).

Table 1. Performance metrics⁶

	Dependent variable	Metrics (calculation)
Outcome	Contribution	Average amount of content contributed (# of responses) per question
	Participation	Percentage of questions answered (proportion of answered questions out of all questions available in a game)
Process	Duration	Average gameplay time per question
	Engagement time	Utilization of playing time (proportion of gameplay time out of maximum possible time)

4.3. Procedure

GUESS was deployed implementing a control and two manipulations of scoring mechanism designs in two events of The European Researchers' Night which is a "popular science and fun learning" event open to the general public, and in five university courses, three at the Faculty of Management and two at the Faculty of Education. Participation was voluntary; yet we offered a modest prize (a pair of movie tickets) to the highest score in the Researchers' Night events; lecturers rewarded their students with 1 or 2 bonus points for participation.

Participants were asked to play one game in each of the three experimental conditions: reward-free games (users saw a pop-up message indicating their contribution), and 2 reward-based systems differing in the algorithm applied (linear $y=3x$ vs. exponential $y=6e^x$). Participants selected the order in which to play the games. A pop-up message indicated the number of points earned for the current response contribution. The data collected from this period includes 21 games, 7 games in each scoring condition. Group size (number of participants in a game) ranges from 13 to 53. A total of 17,752 responses were gathered, distributed as follow: reward-free games 3,681 responses, linear 6,697 responses and exponential 7,374 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.

⁶ We use the natural logarithm transformation for both contribution and duration in order to transform data to normal distribution

4.4. Participants

We identified three types of crowd: general public, expert students and general students as defined in Table 2. 576 participants played at least one game in one of three conditions (Table 3). 53 participants replayed at least one game; for them we calculated the average of contributed responses along the repeated games. The number of game replays for a user in a certain game varied between 1 to 10 times (Table 4).

Table 2. Types of crowd, size and definition

Types of crowd	N	%	Definition
GenPub general public (t1)	154	26.74	People attending The European Researchers' Night events, answering general questions that require common knowledge
ExpStu expert students (t2)	202	35.07	Students responding to professional questions in their field of study
GenStu general students(t3)	220	38.19	Students responding to general questions

Table 3. Number of participants per condition

Condition	Notation	N	%
Reward-free	c0	183	31.77
Linear score mechanism	c1	192	33.33
Exponential score mechanism	c2	201	34.90

We compare behavioral performance among the three scoring conditions in regard to three types of crowd. The independent variables are: the scoring mechanism design (c0, c1, c2), the types of crowd (t1, t2, t3), group size (d; number of participants in a certain game) and number of game rounds (k). The dependent variable, performance, was introduced in Table 1.

5. Results

To compare results from manipulation conditions with those from control condition an HLM⁷ model (Hierarchical Linear Modeling) was conducted. The

⁷ HLM can accommodate nested data structure with repeated observations on the same participant, with a lack of sphericity and missing data.

Table 4. Distribution of game-rounds

K=no. of game-rounds	Frequency	Percent	Frequencies according to respective types of crowd		
			t1	t2	t3
1	523	90.80	139	194	190
2	40	6.94	11	7	22
3	6	1.04	4		2
4	3	0.52		1	2
5	2	0.35			2
6	1	0.17			1
10	1	0.17			1

random variable was the user, the explanatory variables: scoring mechanism design(c), type of crowds (t), group size (d) and number of game rounds (k). The dependent variables were: contribution, participation, duration, and engagement time. A backward elimination procedure was used, whereby non-significant terms were dropped one by one and the model re-assessed to determine the significance of each of the remaining variables at every stage.

Contribution: Table 5.1 shows the final set of variables in the model. The analysis yielded that all effects were statistically significant at $p < .05$.

Table 5.1 Model results showing factors affecting mean change in contribution

Variables in the model	df	F	Sig
Condition (c)	2	6.99	0.0011
Group size (d)	1	23.10	<.0001
Type of crowds (t)	2	25.80	<.0001
c*t	4	2.68	0.0315

A Post hoc analysis using the Bonferroni correction further revealed that for GenPub (t1) the mean for the exponential scoring design (c2) was significantly higher than the means for both the linear and reward-free conditions ($M = 1.71$, $SD = 0.68$; $M = 1.52$, $SD = 0.6$; $M = 1.63$, $SD = 0.59$ respectively). No significant difference was found between the linear (c1) and reward-free conditions (c0). For the ExpStu (t2) the mean of the exponential condition (c2) was significantly higher than the mean for reward-free condition (c0) ($M = 1.45$, $SD = 0.72$, $M = 1.18$, $SD = 0.62$; respectively). No significant difference was found when the linear and reward-free conditions were compared or when the linear and exponential were compared. For the GenStu (t3) no significant differences were found among the three

conditions. The interaction can be seen in Figure 3. In addition, group size main effect was significant ($F_{(1,325)}=23.10, p<.001$) indicating that the larger the group, the lower the contribution of responses.

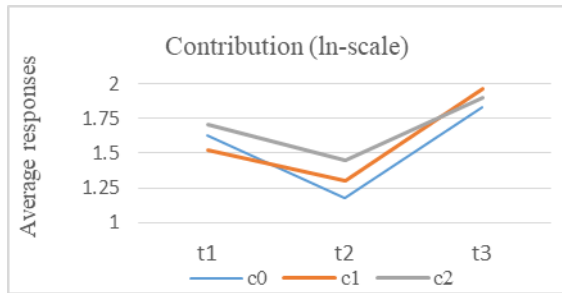


Figure 3. Interaction plot for contribution

Participation: a hierarchical logistic model was conducted to predict the probability to fulfill the task i.e. answering all the questions in the game (or logout from the game before it ends). Table 5.2 shows the results of the regression analysis.

Table 5.2 Model results showing factors affecting mean change in participation

Variables in the model	df	F Value	Sig
Condition (c)	2	5.24	0.0058
Group size (d)	1	5.61	0.0186
Type of crowds (t)	2	19.49	<.0001
c*t	4	6.27	<.0001

The analysis yielded that all effects were statistically significant at $p < .05$ significance level. Further analysis using the Bonferroni correction revealed that for the GenPub (t1) the probability of answering all questions and completing the task for the linear scoring design (c1) was significantly lower than both the exponential and reward-free conditions (45%, 59%, 61% respectively). No significant difference was found between the exponential (c2) and reward-free conditions (c0). For the ExpStu (t2) the exponential condition (c2) was significantly lower than reward-free and linear conditions (66%, 83% and 79% respectively). No significant difference was found when the linear and reward free conditions were compared. For the GenStu group (t3) the linear scoring design (c1) was significantly lower than both the exponential and reward-free conditions (77%, 83%, 82% respectively). No significant difference was found between the exponential (c2) and reward-free conditions (c0) (Figure 4). These results suggests that for both GenPub and GenStu it may be more

effective to use reward-free or exponential scoring design while for ExpStu exponential scoring design may be a less effective method to promote participation. In addition, there was a significant main effect of group size ($F_{(1,284)}=5.61, p < .05$) indicating that participation decreases the larger the group.

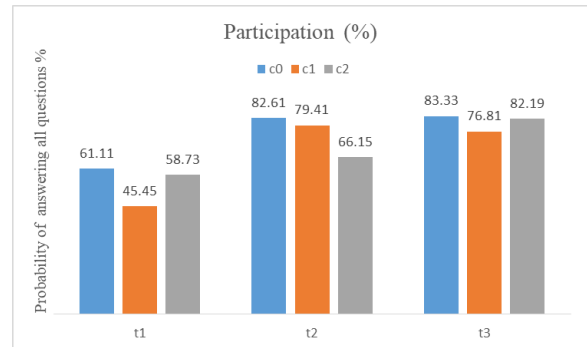


Figure 4. Participation per crowd type and scoring mechanism

Duration: Table 5.3 shows the final set of variables in the model in regard to duration.

Table 5.3 Model results showing factors affecting mean change in duration

Variables in the model	df	F Value	Sig
Condition (c)	2	3.07	0.0480
Group size (d)	1	5.67	0.0178
Type of crowds (t)	2	16.92	<.0001

The analysis yielded that all main effects were statistically significant at $p < .05$. For the main effect of score design ($F_{(2, 329)} = 3.07, p < .05$), indicates that there is an overall significant difference in means, post hoc tests using the Bonferroni correction revealed that duration was significantly higher for exponential condition ($M = 3.94, SD = 0.29$) than for reward-free ($M = 3.88, SD = 0.29$). There was no statistically significant difference between the linear and both reward free and exponential scoring. For the main effect of types of crowd ($F_{(2, 329)} = 16.92, p < .001$), post hoc tests using the Bonferroni correction revealed that duration was significantly higher for GenStu (t3) ($M = 4.01, SD = 0.18$) than for both GenPub (t1) and ExpStu (t2) ($M = 3.90, SD = 0.31; M = 3.79, SD = 0.38; respectively$). There was no statistically significant difference between t1 and t2. Figure 5 illustrates the findings.

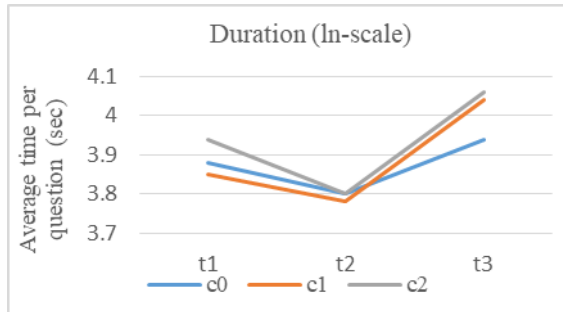


Figure 5. Duration per crowd type and scoring mechanism

In addition, group size main effect was significant ($F_{(1,329)}=5.67, p<.05$) indicating that duration decreases the larger the group.

Engagement time: The analysis yielded that all main effects were statistically significant at the $p < .05$ significance level as can be seen in Table 5.4.

Table 5.4 Model results showing factors affecting mean change in engagement time

Variables in the model	df	F Value	Sig
Condition (c)	2	4.16	0.0165
Group size (d)	1	10.51	0.0013
Type of crowds (t)	2	24.91	<.0001
Number of game rounds (k)	1	37.22	<.0001
c*t	4	2.41	0.0490

A Post hoc analysis using the Bonferroni correction further revealed that for the GenPub (t1) the mean for the exponential scoring design (c2) was significantly higher than the means for both the linear and reward-free conditions ($M = 0.76, SD = 0.30; M = 0.68, SD = 0.34; M = 0.72, SD = 0.31$ respectively). No significant difference was found between the linear (c1) and reward-free conditions (c0). For the ExpStu (t2) no significant difference

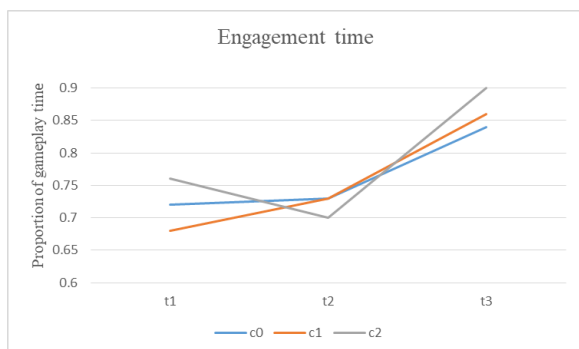


Figure 6. Interaction plot for engagement time

was found between the three conditions. For the GenStu(t3) the mean for the exponential scoring design (c2) was significantly higher than the mean for the reward free condition ($M = 0.9, SD = 0.19; M = 0.84, SD = 0.21$ respectively). No significant differences were found between the linear (c1) and both reward-free and exponential conditions. The interaction can be seen in Figure 6.

In addition there were main effects of group size (d) and of number of game rounds (k).

Generalizing the observations, the exponential reward function leads to higher contribution, participation, duration, and engagement time in the GenPub and GenStu groups. Exponential points reward catalyzes knowledge pooling. The GenPub responded similarly to the GenStu but at lower rates, i.e. it is more difficult to pool knowledge from the GenPub compared to the GenStu. The hardest challenge is to pool knowledge from the ExpStu accompanied by a surprisingly short duration. Possibly, less fun (serious questions) leads to less participation and contribution and less attention displayed by duration.

6. Discussion

Designing a game scoring mechanism to support achieving objectives is a relevant practical issue, as well as an academic interest particularly in light of the increased use of gamification in recent years [3, 4, 29, 46]. Most of the studies in gamification embedded a combination of motivational factors: points, leaderboards, and levels [41], dual point systems and rating [6, 37, 38], leaderboards, ranking, unlocking information [47], progress, points, ranking, network [48], leaderboard, badges [49], challenge, progress bar, theme, rewards [44], badges, levels[50], points, levels, avatar [44]. The distinct effect of point scores on users' behavior and motivation is still vague, a gap we wish to narrow.

The current study focuses on the effect of scoring design on each of three types of crowd in an implementation of a knowledge pooling game for crowds. We applied three scoring mechanisms: reward-free, linear and exponential in three types of crowd: GenPub, ExpStu and GenStu. We examined contribution of responses, participation, duration, and engagement time as dependent variables. Results showed a significant interaction between score design mechanism and types of crowd indicating that the groups behaved differently in reaction to the scoring condition. The analysis for duration yielded main effects for the scoring condition and type of crowd.

Overall, in answer to our research questions we can generalize the following assertions: 1.

implementing a point allocation mechanism promotes performance depending on the implementation. 2. linear point progression is detrimental to performance while exponential score allocation is favorable. 3. A crowd of experts is more difficult to mobilize than general crowds. Given general questions, students show better performance than the general public. In the following we unpack the particular observations that led to these generalizations.

Looking into the contribution behavior (amount of content) results suggests that for the GenPub and for the ExpStu exponential scoring design may be the more effective method when compared to linear and reward-free conditions. The GenStu appear to be indifferent toward these different scoring designs methods. Exponential scoring design may also be the more effective method to increase duration when compared to the reward-free condition, while the difference between the linear and both reward-free and exponential scoring was found to be insignificant.

Examining the types of crowd it appears that GenStu participants spend a longer time on each question compared to the other two groups. Examining engagement time more closely, it appears that exponential scoring design is more effective when compared to reward-free conditions for both the GenPub and the GenStu. In regard to the GenPub, it is also more effective when compared to linear condition. In contrast, the ExpStu appear to be indifferent about different scoring designs methods relating to engagement time.

Participation in the GenStu group was significantly higher than that of the GenPub, yet it is interesting to observe that the same pattern occurred for both GenPub and GenStu (t1 and t3). For these two groups both exponential scoring design and reward free condition seem to be more effective when compared to the linear condition. ExpStu seems to be indifferent for reward free and linear condition. Thus, it appears that linear scoring is least effective.

To conclude, the present study highlights the importance of point scoring mechanism design in games for crowds. Findings suggest that linear reward crowds out intrinsic motivation while exponential reward may strengthen motivation in relation to the reward free condition. Table 6 summarizes the complete matrix of the experiment results. Interestingly, the percentage of questions answered (participation) by the ExpStu (t2) was significantly lower in the exponential condition compared to the other conditions and ExpStu exploited less of the available game time (lower engagement time), yet the average responses per question answered was significantly higher. This

suggests that scores enhance intrinsic motivation, because mere “hunger” for points would lead to full usage of the available game time and questions. Participation for GenPub was almost the same for exponential and reward free conditions, yet engagement time and contribution were significantly higher in exponential scores. This suggests that scores serve as informative feedback generating motivation.

Table 6. Summary of the exponential condition

Exponential vs. reward free	Contribution	Duration	Engagement	Participation
t1 GenPub	↗	+ n.s	↗	- n.s
t2 ExpStu	↗	+ n.s	- n.s	↘
t3 GenStu	+ n.s	↗	↗	- n.s

6.1. Limitations and Future Research

The main limitation of the current study is the sample size which is modest considering we are interested in crowds. Another limitation is that we implemented just two score design mechanisms. This being an experiment, although a natural experiment as far as the general public goes, external validity may be questioned.

The findings suggest several points of departure for future research: 1) Group size – the effect of group size was significant in regard to our variables, indicating that the larger the group the lower the contribution of responses, participation and duration per question. This may imply diffusion of responsibility. It may be the expected emergence of a power law distribution, a well-known phenomenon of the web [51], which leads to higher diversity in levels of participation. 2) Replaying- repeated play (k) has a significant effect on engagement time. This raises the question of differences between highly-motivated users and average users and could also provide new insights into the design of effective gamified crowdsourcing systems for different target groups. 3) Quality- in the current study users were rewarded with points for every response. Our next step will be to explore the quality of responses.

6.2. Conclusions

This work contributes to the area of crowdsourcing and gamification, especially games

for crowds. Its main novelty lies in bringing together crowdsourcing and game capabilities for different scoring mechanism designs. When designing crowd-based knowledge pooling, gamification can help if implemented according to our findings. I.e., care should be taken in selecting whether and how to use points, and attention should focus on the type of crowd as well as the relation between the type of knowledge collected and the type of crowd.

7. Acknowledgement

Partial support for this research was provided by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation 1716/12.

8. References

- [1] J. Prpić and P. Shukla, "The Contours of Crowd Capability," in Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS), pp. 3461-3470, 2014.
- [2] IEEE, "Everyone's a Gamer – IEEE Experts Predict Gaming Will Be Integrated Into More than 85 Percent of Daily Tasks by 2020," Feb. 25, 2014.
- [3] J. Hamari, J. Koivisto and H. Sarsa, "Does Gamification Work?—A Literature Review of Empirical Studies on Gamification," in Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS), pp. 3025-3034, 2014.
- [4] K. Seaborn and D.I. Fels, "Gamification in theory and action: A survey," *International Journal of Human-Computer Studies*, vol. 74, pp. 14-31, 2015.
- [5] C. Lister, J.H. West, B. Cannon, T. Sax and D. Brodegard, "Just a Fad? Gamification in Health and Fitness Apps," *JMIR Serious Games*, vol. 2, e9, 2014.
- [6] M. Witt, C. Scheiner and S. Robra-Bissantz, "Gamification of online idea competitions: Insights from an explorative case," in *INFORMATIK 2011 - Informatik schafft Communities 41. Jahrestagung der Gesellschaft für Informatik*, pp. 192, 2011.
- [7] E.R. Mollick and N. Rothbard, "Mandatory Fun: Gamification and the Impact of Games at Work," *The Wharton School Research Paper Series*, vol. 1, pp. 1-51, 2013.
- [8] E.B. Passos, D.B. Medeiros, P.A.S. Neto and E.W.G. Clua, "Turning Real-World Software Development into a Game," in *Games and Digital Entertainment (SBGAMES'11)*, pp. 260-269, 2011.
- [9] J. Fernandes, D. Duarte, C. Ribeiro, C. Farinha, J.M. Pereira and M.M.d. Silva, "iThink: A Game-Based Approach Towards Improving Collaboration and Participation in Requirement Elicitation," *Procedia Computer Science*, vol. 15, pp. 66-77, 2012.
- [10] J.J. Lee and J. Hammer, "Gamification in Education: What, How, Why Bother?" *Academic Exchange Quarterly*, vol. 15, pp. 146, 2011.
- [11] B. Morschheuser, J. Hamari and J. Koivisto, "Gamification in crowdsourcing: A review," in Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), pp. 4375-4384, 2016.
- [12] G. Richter, D.R. Raban and S. Rafaeli, "Studying Gamification: The Effect of Rewards and Incentives on Motivation," in *Gamification in Education and Business*, Springer International Publishing, 2015, pp. 21-46.
- [13] J. Prpić and P. Shukla, "The Theory of Crowd Capital," in Proceedings of the 46th Hawaii International Conference on System Sciences, (HICSS), pp. 3505-3514, 2013.
- [14] M.K. Poetz and M. Schreier, "The value of crowdsourcing: Can users really compete with professionals in generating new product ideas?" *Journal of Product Innovation Management*, vol. 29, pp. 245-256, 2012.
- [15] P. Whitla, "Crowdsourcing and Its Application in Marketing Activities," *Contemporary Management Research*, vol. 5, pp. 15-28, 2009.
- [16] D. Geiger and M. Schader, "Personalized task recommendation in crowdsourcing information systems—Current state of the art," *Decis.Support Syst.*, vol. 65, pp. 3-16, 2014.
- [17] J. Prpić and P. Shukla, "Crowd Science: Measurements, Models, and Methods," in Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), pp. 4365, 2016.
- [18] C. Haythornthwaite, "Crowds and communities: Light and heavyweight models of peer production," in Proceedings of the 42nd Hawaii International Conference on System Sciences, HICSS'09, pp. 1-10, 2009.
- [19] M. Vukovic, J. Laredo and S. Rajagopal, "Challenges and experiences in deploying enterprise crowdsourcing service," in Proceedings of the 10th International Conference on Web Engineering, pp. 460-467, 2010.
- [20] A.J. Quinn and B.B. Bederson, "Human computation: a survey and taxonomy of a growing field," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1403-1412, 2011.
- [21] S. Rafaeli and Y. Ariel, "Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia," *Psychological Aspects of Cyberspace: Theory, Research, Applications*, pp. 243-267, 2008.
- [22] S. Oreg and O. Nov, "Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values," *Comput.Hum.Behav.*, vol. 24, pp. 2055-2073, 9. 2008.
- [23] S. Deterding, M. Sicart, L. Nacke, K. O'Hara and D. Dixon, "Gamification: Using Game Design Elements in Non-Gaming Contexts," in Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, ACM New York, NY, USA: ACM, 2011, pp. 2425-2428.
- [24] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker and Z. Popović, "Predicting protein structures with a multiplayer online game," *Nature*, vol. 466, pp. 756-760, 2010.
- [25] M. Poesio, J. Chamberlain, U. Kruschwitz, L. Robaldo and L. Ducceschi, "Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation,"

- ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 3, pp. 1-44, 2013.
- [26] L. von Ahn, "Games with a purpose," *Computer*, vol. 39, pp. 92-94, 2006.
- [27] I. Blohm and J.M. Leimeister, "Design of IT-Based Enhancing Services for Motivational Support and Behavioral Change," *Business & Information Systems Engineering*, vol. 5, pp. 275-278, 2013.
- [28] B. Morschheuser, J. Hamari, J. Koivisto and A. Maedche, "Gamified Crowdsourcing: Conceptualization, Literature Review, and Future Agenda," *International Journal of Human-Computer Studies*, vol. 106, pp. 26-43, 2017.
- [29] J. Hamari, "Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service," *Electronic Commerce Research and Applications*, vol. 12, pp. 236-245, 4. 2013.
- [30] T. Itoko, S. Arita, M. Kobayashi and H. Takagi, "Involving senior workers in crowdsourced proofreading," in *International Conference on Universal Access in Human-Computer Interaction*, pp. 106-117, 2014.
- [31] M. Csikszentmihalyi and M. Csikszentmihalyi, *Flow: The psychology of optimal experience*, HarperPerennial New York, 1991, .
- [32] R.M. Ryan and E.L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *Am.Psychol.*, vol. 55, pp. 68-78, 2000.
- [33] Y.C. Zhao and Q. Zhu, "Effects of extrinsic and intrinsic motivation on participation in crowdsourcing contest," *Online Information Review*, vol. 38, pp. 896-917, 2014.
- [34] J. Hamari and J. Koivisto, "Measuring flow in gamification: Dispositional Flow Scale-2," *Comput.Hum.Behav.*, vol. 40, pp. 133-143, 11. 2014.
- [35] T. Hiltbrand and M. Burke, "How Gamification will change Business Intelligence," *Business Intelligence Journal*, vol. 16, pp. 8-16, 2011.
- [36] J. Thom, D. Millen and J. DiMicco, "Removing gamification from an enterprise SNS," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp. 1067-1070, 2012.
- [37] D. Pothineni, P. Mishra, A. Rasheed and D. Sundararajan, "Incentive design to mould online behavior: a game mechanics perspective," in *Proceedings of the First International Workshop on Gamification for Information Retrieval*, pp. 27-32, 2014.
- [38] R. Snijders, F. Dalpiaz, S. Brinkkemper, M. Hosseini, R. Ali and A. Ozum, "REFine: A gamified platform for participatory requirements engineering," in *2015 IEEE 1st International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*, pp. 1-6, 2015.
- [39] S. Hacker and L. von Ahn, "Matchin: eliciting user preferences with an online game," in *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pp. 1207-1216, 2009.
- [40] I. Guy, A. Perer, T. Daniel, O. Greenshpan and I. Turbahn, "Guess Who?: Enriching the social graph through a crowdsourcing game," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, pp. 1373-1382, 2011.
- [41] E.D. Mekler, F. Brühlmann, K. Opwis and A.N. Tuch, "Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, pp. 66-73, 2013
- [42] R.M. Ryan and E.L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemp.Educ.Psychol.*, vol. 25, pp. 54-67, 2000.
- [43] J. Vassileva, "Motivating participation in social computing applications: a user modeling perspective," *User Modeling and User-Adapted Interaction*, vol. 22, pp. 177-201, 2012.
- [44] C. Prandi, M. Rocchetti, P. Salomoni, V. Nisi and N. Nunes, "Fighting exclusion: a multimedia mobile app with zombies and maps as a medium for civic engagement and design." *Multimedia Tools & Applications*, vol. 76, pp. 4951-4979, 2017.
- [45] P. Sweetser, D. Johnson, P. Wyeth and A. Ozdowska, "GameFlow heuristics for designing and evaluating real-time strategy games," in *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System*, pp. 1-24, 2012.
- [46] A. Kankanhalli, M. Taher, H. Cavusoglu and S.H. Kim, "Gamification: A new paradigm for online user engagement," in *the 33rd International Conference on Information Systems, ICIS 2012*, pp. 1-10, 2012.
- [47] M. Machnik, M. Riegler and S. Sen, "Crowdpinion: Motivating People to Share their Momentary Opinion." in *the 2nd GamifIR'15 Workshop*, pp. 44-51, 2015.
- [48] R. Farzan, J.M. DiMicco, D.R. Millen, C. Dugan, W. Geyer and E.A. Brownholtz, "Results from deploying a participation incentive mechanism within the enterprise," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 563-572, 2008.
- [49] A. Domínguez, J. Saenz-de-Navarrete, L. de-Marcos, L. Fernández-Sanz, C. Pagés and J. Martínez-Herráiz, "Gamifying learning experiences: Practical implications and outcomes," *Comput.Educ.*, vol. 63, pp. 380-392, 4. 2013.
- [50] J. Hamari and J. Koivisto, "Social motivations to use gamification: an empirical study of gamifying exercise," in *Proceedings of the 21st European Conference on Information Systems*, 2013.
- [51] D.R. Raban and F. Harper, "Motivations for answering questions online," in *New Media and Innovative Technologies*, D. Caspi and T. Azran, Israel: Ben-Gurion University press, 2008, pp. 73-97.