Gamification and the Trough of Disillusionment

Jan Broer

Institut für Informationsmanagement, Universität Bremen

Abstract

Consulting agencies predict a drop in the visibility of gamification as studies report more and more negative results. We re-evaluate studies with positive results from a recent literature review with regards to the effects of gamification reported therein. We find that most reviewed studies do not actually report net-positive results and discuss reasons as to why gamification seemingly fails to deliver. One major Potenzial reason identified is the variance in motivations to play games.

1   Introduction

Many introductions to gamification start with a reference to the Gartner Hype Cycle (Gartner 2013), usually pointing out that gamification is an up-and-coming technology. The Hype Cycle also predicts a strong fall in the visibility of hyped technologies, however. Gartner calls this period the trough of disillusionment. While one can argue the validity of the Hype Cycle (due to a lack of scientific validation), gamification has clearly seen a meteoric rise in expectations in the past few years, as projected by Gartner's model. A recent literature review by Hamari et al. (2014) also seems to paint a positive picture at first glance. Out of 24 peer reviewed empirical research papers reviewed, 22 included quantitative results. 2 of those reported purely positive results, 13 are listed as partially positive and 7 as only presenting descriptive statistics. Hamari et al. discuss a variety of caveats, however, that put the results in a less positive light. The authors mention 8 methodological shortcomings in the reviewed works: small sample sizes, lacking use of validated instruments, lack of control groups, multiple affordances studied at once, use of purely descriptive statistics, very short experiment timeframes, lack of clarity, and no use of multi-level measurement models (Hamari et al. 2014). In this paper, we review the same papers again with the aim of identifying which of them can actually be considered evidence of positive effects of gamification. We then discuss what this means for gamification as a discipline, as well as possible reasons why gamification does not perform as well as anticipated.

2   Methods

In order to get a better understanding of the state of gamification, we have reviewed the papers listed by Hamari et al. as having positive results, focusing on four major questions:

1. Pertinence: Does the paper actually deal with gamification? Are the instruments employed appropriate to give insight into the effectiveness of gamification?

2. Positive results: Does the paper report significant positive outcomes regarding the use of gamification?

3. Negative results: Does the paper report significant negative outcomes regarding the use of gamification?

4. Unsupported hypotheses: Does the paper report hypotheses that could not be confirmed?

There are many competing definitions of gamification. Definitions that are too broad are of little use for research, as they make the defined term itself meaningless. We therefore chose two narrower definitions for the term to measure each reviewed paper against. Deterding et al. (2011) refer to gamification as "the use of game design elements in non-game contexts", while Huotari and Hamari (2012) define gamification from a service marketing perspective as "a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation". While one focuses on affordances used in games, the other concentrates on the game-like experiences evoked by those affordances. Both describe the use of individual affordances, not full-fledged games. Studies discussing complete games were therefore marked as non-pertinent to the question at hand. Studies that did not test hypotheses regarding the effects of gamification or used insufficient empirical means to do so were likewise excluded. Statistically significant results that support gamification having a positive effect from the perspective of the person employing gamification were marked as positive results. If the goal of gamifying a system was to increase participation, for example, results that showed significant increases in participation for the experimental group were marked as positive. Likewise, results that supported the opposite of the intent behind the specific instance of gamification were marked as negative. If the paper included (usually positive) hypotheses about the effects of gamification that were not supported by data, those were marked in the fourth category.

3   Results

Figure 1 shows the distribution of our ratings of the results of the 15 papers marked by Hamari et al. as positive or partially positive. 7 had to be marked as non-pertinent. A prominent example is the work by Eickhoff et al. (2012), employing a full-fledged annotation game in their treatment group. While their results are promising, their treatment is better described as a game with a purpose (Ahn 2006) rather than gamification. A different, but equally non-pertinent example is (Hamari & Koivisto 2013). Their study deals with the motivations for the use, not the effects of gamification. In a final example, Li et al. (2012)

used experimental setups for their groups that diverged so much from each other that it is difficult to pinpoint their results to specific sources.

Of the remaining 8 papers, 3 reported significant negative effects in addition to positive ones. Domínguez et al. (2013) report that the experimental group of students did significantly better on certain tasks but also significantly worse on a final examination and on partici-pation. They conclude that "quantitative analysis suggests that cognitive impact of gami-fication over students is not very significant" (Domínguez et al. 2013). Downes-Le Guin et al. (2012) report that their gamified experimental setup was significantly more interesting than other versions, but they also report a significant decrease in speed and in task completion rates. The study by Hamari (2013) did not confirm any major hypothesis. The only significant correlations found were between user activity and the amount of times the users viewed their own badges or those of others. It seems easy to attribute these correlations to the fact that more active users are more likely to find and use a badge system in the first place.
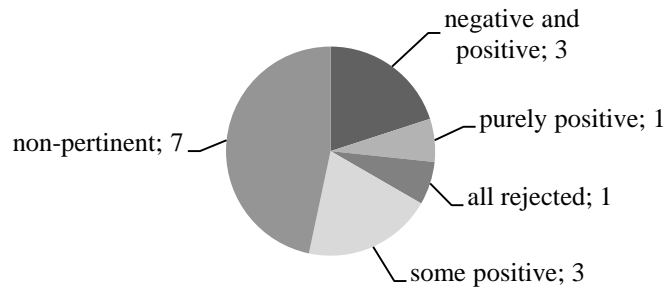

Figure 1 Revised rating of paper results.

The second study that showed purely positive results dealt with the removal of gamification from a previously gamified system (Thom et al. 2012). The authors show a significant decrease in activity following the removal. It is unclear whether this decrease means that gamification had a positive effect or whether it was the removal of expected rewards that resulted in reduced motivation. Such an effect (a reduction of intrinsic motivation through extrinsic rewards) has been extensively studied in psychology (e.g. Deci et al. 2001).

Three papers reported a mix of positive results and unconfirmed hypotheses. Denny (2013) saw an increase in student activity, but no significant differences in question authoring, response quality, and perceived learning value when gamifying an online learning tool. Farzan et al. (2008) saw an increase in the amount of content shared, but not in the amount of users sharing content in the long run. Finally, Hakulinen et al. (2013) measured an improved time management in their experimental group, but could not confirm significant differences in individual submission scores, final scores, or carefulness in completing tasks. Overall, evidence of gamification's success is hardly conclusive. 3 out of 15 papers marked as positive in Hamari et al.'s review can be considered net-positive without caveats. All 3 could only confirm part of their hypotheses.

4   Discussion

The results described above are hardly encouraging for researchers and practitioners dealing with gamification. While it is too early to say that gamification simply does not work, it is also difficult to draw the opposite conclusion. Those researchers that reported net-positive results could only verify a subset of their hypotheses. Increases in user activity seem to be achievable, but little else has been shown. Following Gartner's Hype Cycle, these results could indicate that we are about to enter the trough of disillusionment. The high expectations for gamification are not confirmed in experiments and it might only be a matter of time until visibility drops. The results of studies on gamification are not without hope, however.

One observation in particular seems worthy of discussion: Hakulinen et al. report a significant change of behavior in a small group of their users, while most of them behaved similarly independent of their experimental conditions (Hakulinen et al. 2013). A few users were greatly motivated by the introduction of badges in an e-learning environment, while most did not seem to be affected. Denny (2013) and Farzan et al. (2008) suggest such differences as well. While it is obvious that different users would react differently, the fact that there seems to be only a small group of users affected by a specific affordance seems important to both practitioners of gamification and to the interpretation of results of studies. If one is measuring participation, for example, a strong effect on a small group of partici-pants might be lost in statistical variation. Within-subjects experimental setups might be useful in identifying such different types of users. A very rough calculation shows that any individual motivational affordance is likely to affect only a small subset of users. A study among US residents of the ages from 18-44 found that 55% of employed respondents "would be interested in working for a company that offers games as a way to increase productivity" (Saatchi & Saatchi S 2011). A possible conclusion from this data is that about half of a system's users can be expected to react positively to game related affordances at all.

Affecting half your users might still be a worthwhile goal and would likely give significant results. Unfortunately, it is probable that user bases are partitioned even further. Literature in game studies often refers to an essay on types of players in multi-user dungeons (Bartle 1996). Bartle identifies four types of users based on their motivations to play the game – achievers, socializers, explorers, and killers. These types have been widely adopted beyond their original scope, even though the validity of Bartle's original statements is questionable there. Yee (2007) identified three major components of motivations for play in massively multiplayer online role-playing games – achievement, social, and immersion. These are further subdivided into 10 separate components of motivation. Clearly, players of these types of games do not all play for the same reasons. If one gamifies by taking individual game design elements from such games and then implementing them in a non-game context, one will not reach all members of the target audience that are motivated by games, but only those that are motivated by the affordances chosen. Arbitrarily assuming an even split between differently motivated users, any single affordance would only reach about 15 percent of users. Percentages get even worse when implementing affordances that only make use of one of the sub-components mentioned by Yee. Gamification aimed purely at advancement, for example, would then affect an even smaller subset of users. If one considers different types

of games as well it is easy to imagine that only single-digit percentage of users will be positively affected by any individual gamification effort. It is important to stress again that the above calculation is very rough and based on many arbitrary estimates. Furthermore, Yee (2007) notes that users are not limited to one motivation component. Players may be interested in both achievement and immersion, for example. Nevertheless, one can draw a few conclusions here:

- A narrow gamification strategy should not be expected to achieve large effect sizes across random populations. The implications for research are varied. For one, it would be very interesting to correlate users' game-playing habits and motivations (e.g. using Yee's inventory of items) to their response to gamification. For another, quantitative research designs should be adjusted to the expected effect size. Furthermore, qualitative research might be very useful in understanding the reasons for being affected by certain motivational affordances and not by others.

- The current focus in gamification on badges, levels, leaderboards, and points is too narrow to be useful in the long run. All of these speak to the achiever type and ignore other components of motivation.

- It is important to understand the target audience of a gamified system if one wants to gamify successfully. Saatchi & Saatchi S (2011) report major differences in interest in gamification, for example. While only 55% of their employed respondents indicated interest, 68% of males 18-34 responded that way as did 81% of respondents that spent more than $20 on app purchases in the 12 months preceding the study.

5   Conclusion

Scientific studies on gamification report mixed results that paint a rather bleak picture. Most implementations of gamification studied show that their effect on motivation or participation is far lower than the hype would make one believe. The emergence of these studies indicates that gamification might be entering the trough of disillusionment predicted by the Gartner Hype Cycle. We have presented some possible reasons for these rather negative results. The main reason we see is that user motivation can have many different components and that even motivations to play games differ widely. As such, only parts of a target audience will ever be motivated by gamification and these are likely to be smaller, the narrower the implemented motivational affordances are. We suggest that further research in the area keeps such a split in the target audience in mind and either directly studies how different groups of users react differently to gamification or adjusts research designs appropriate to the expected (low) effect size. Furthermore we suggest that collaboration between researchers, system operators, and practitioners is needed if research should reflect conditions as they are in practice instead of sterile, small-scale implementations. Finally, the low number of pertinent results indicates the need for further rigorous empirical studies if we are to truly understand the effects of gamification.

References

Ahn, L. von. (2006). Games with a purpose. *Computer*, *39*(6), 96–98.

Bartle, R. (1996). Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. Retrieved July 12, 2012, from http://www.mud.co.uk/richard/hcds.htm

Deci, E. L. Koestner, R. & Ryan, R. M. (2001). Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research*, *71*(1), 1–27.

Denny, P. (2013). The Effect of Virtual Achievements on Student Engagement. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, pp. 763–772.

Deterding, S. Dixon, D. Khaled, R. & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification." In *MindTrek '11 Proceedings*. New York, NY, USA: ACM, pp. 9–15.

Domínguez, A. Saenz-de-Navarrete, J. de-Marcos, L. Fernández-Sanz, L. Pagés, C. & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, *63*, 380–392.

Downes-Le Guin, T. Baker, R. Mechling, J. & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research*, *54*(5), 1–21.

Eickhoff, C. Harris, C. G. de Vries, A. P. & Srinivasan, P. (2012). Quality through flow and immersion: gamifying crowdsourced relevance assessments. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, pp. 871–880.

Farzan, R. DiMicco, J. M. Millen, D. R. Brownholtz, B. Geyer, W. & Dugan, C. (2008). When the experiment is over: Deploying an incentive system to all the users. In *Symposium on Persuasive Technology, In conjunction with the AISB 2008 Convention*. Aberdeen, Scotland.

Gartner. (2013). Hype Cycle for Emerging Technologies, 2013. Retrieved May 30, 2014, from http://www.gartner.com/document/2571624

Hakulinen, L. Auvinen, T. & Korhonen, A. (2013). Empirical Study on the Effect of Achievement Badges in TRAKLA2 Online Learning Environment. In *Learning and Teaching in Computing and Engineering (LaTiCE), 2013*, pp. 47–54.

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.

Hamari, J. & Koivisto, J. (2013). Social motivations to use gamification: An empirical study of gamifying exercise. In *Proc. ECIS 2013*. Utrecht, Netherlands.

Hamari, J. Koivisto, J. & Sarsa, H. (2014). Does Gamification Work?—A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences.*.

Huotari, K. & Hamari, J. (2012). Defining gamification: a service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference*. New York, NY, USA: ACM, pp. 17–22.

Li, W. Grossman, T. & Fitzmaurice, G. (2012). GamiCAD: a gamified tutorial system for first time autocad users. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM, pp. 103–112.

Saatchi & Saatchi S. (2011). Engagement Unleashed: Gamification for Business, Brands, and Loyalty. Retrieved May 30, 2014, from http://de.slideshare.net/Saatchi_S/gamification-study

Thom, J. Millen, D. & DiMicco, J. (2012). Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 1067–1070.

Yee, N. (2007). Motivations for Play in Online Games. *CyberPsychology & Behavior*, 772–775.

Contact information

Jan Broer (jbroer@ifib.de)

ifib: Institut für Informationsmanagement Bremen GmbH
and
AG Informationsmanagement, Universität Bremen

Am Fallturm 1
28359 Bremen
Germany