In Pursuit of Evidence: a Design and Empirical Study of a Gamified Online Marketplace

Bauke Buikstra¹,², Rens Kortmann¹, Eric Klaassen², Laurens Rook¹, Alexander Verbraeck¹

¹ Delft University of Technology, Faculty of Technology, Policy & Management, Jaffalaan 5, 2628BX Delft, The Netherlands
² BLOOM, Keizersgracht 119, 1015CJ Amsterdam, The Netherlands
{l.j.kortmann, l.rook, a.verbraeck}@tudelft.nl
{bauke, eric}@bloomamsterdam.com

ABSTRACT
Gamification seems a promising method to engage people and stimulate their activity, both offline and online, but has a dearth of empirical evidence in scientific literature. This study investigated whether gamification can have a positive effect on the amount of user-generated content on online marketplaces. Two similar gamification treatments were designed for the mobile website of OLX India, a marketplace for used goods. The treatments were quantitatively evaluated in a double-blind controlled experiment on 51,103 OLX users, who were randomly selected and evenly assigned to a control group and two treatment groups. The users who actively engaged with either of the treatments created more than 6 times as much content as users who did not engage with the treatments, resulting in an overall content increase of over 18% for the treatment groups compared to the control group. Gamification on online marketplaces is effective for users who actively interact with a treatment, not for all users per se.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Evaluation/Methodology, User-Centered Design; H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous; J.4 [Social and Behavioral Sciences]: Psychology

General Terms
Management, Design, Experimentation

Keywords
Gamification, gameful design, controlled randomised experiment, online marketplace, user-generated content, network effects

1. INTRODUCTION
Gamification has gained attention as a method to motivate and engage people, mostly online platforms. Gamification uses the strategies of persuasive technology and motivating elements from games to affect behaviour during a task, thus serving a facilitating role in a service or product [9, 24]. The main potential of gamification is likely the increased user engagement because of the added elements of play, fun and competition, creating a ‘gameful experience’. Depending on the system where it is implemented, increased engagement can result in an increased or transformed motivation to participate, use, learn, have social interaction or perform tasks [8, 9, 22, 24, 38, 46].

1.1 Theoretical Perspective on Gamification
The gamification field contains strongly divided opinions, definitions and movements [10, 22]. Some even see gamification as a re-invented marketing tool to trick and exploit customers in order to increase profits [14]. Within scientific literature there are generally two theoretical perspectives on gamification: an experiential perspective led by Hamari and others, in which gamification is defined by the experience a user has when interacting with a system [24, 28] and a systemic perspective (originally) led by Deterding and others, where gamification is defined as “the use of game design elements in non-game contexts” ([11], p. 10). Deterding recently published work in which he proposes to combine both perspectives into a more socio-technical view; recognising the game design elements, but also the user and context perspective: “gamification is a holistic socio-technical systems design practice […] one that understands humans interacting with technology as assemblages, activity systems, or ecologies of heterogeneous and intertwined actors” ([10], pp. 312-313). The goal of gamification is then to afford systemic and emergent motivational experiences in socio-technical systems. This view acknowledges most of the criticism gamification has received [6, 18, 20]. Deterding’s perspective on gamification [10] is adopted throughout this research.

1.2 Scarcity of Gamification Evidence
A scarce supply of scientific literature and empirical research results and its young age indicate that gamification is currently still in its infancy: the design principles are scattered and diverse; there are multiple definitions; there is a rampant growth of self-proclaimed gamification experts; and gamification has been deemed a marketing buzzword. Gamification literature demands for more empirical research and case studies; especially research that is executed with a valid design and evaluation methodology [10, 17, 22, 37, 39, 42, 46]. A recent literature review of available empirical gamification studies concluded that only 5 studies were methodologically well executed [24]. Most of these studies
concluded that gamification only works in part of the hypothesised relationships between game elements and outcomes. The context being gamified and the qualities of the users were identified as the main underlying confounding factors. No structured game or gamification design methods were used, or at least not set out in the papers and reports. Game elements were selected and implemented on a logical (or seemingly random) basis, following the nature of the product or service to enhance [8, 15, 24–26, 29].

1.3 Research Goal and Approach

Our goal is to find evidence for the described potential of gamification. Our approach is a design and empirical study of a gamified system. Empirical studies that follow a structured, user-centered gamification design method are highly demanded in gamification literature. In order to follow this approach, the application domain of online marketplaces was chosen, since the general challenges of these platforms align with the possible added value of gamification. An online marketplace was used as case for this research. The remainder of this paper is organised as follows: background on the application domain and case; methods used for the design and effect study; empirical results; discussion and conclusions.

2. METHODS

2.1 Application Domain

Several industries have been challenged by disruptive new business models, where supply and demand are connected directly through an online platform. Airbnb, Uber and Craigslist are commonly known examples. These are online marketplaces: an online platform that purely facilitates communication of two or more user sides (buyers and sellers), through online content created by these users, which is known as user-generated content (UGC). Their main value for society is that the search and/or transaction costs for users are considerably reduced [23, 43]. Characteristic for online marketplaces are network effects: “the utility that a given user derives from the good depends on the number of other users who are in the same ‘network’ as he or she” ([30], p. 424). There are different types of networks effects (see figure 1), but online marketplaces heavily rely on cross-side network effects for growth, where sellers (therefore content) and buyers (therefore transactions) are interdependent [1, 13].

Figure 1. Network effects in online marketplaces; adopted from [13]

For online marketplace users, supplying content is generally a bigger effort than browsing content [2, 3]. Due to cross-side network effects, this content bottleneck can hinder overall platform growth. Thus, stimulating the amount of content provided by sellers is vital for online marketplaces [2, 13, 23]. The potential of gamification to aid in online marketplace challenges can easily be discerned. UGC is driven by user activity, where the decision to create content partly depends on the level of engagement of a user with the online marketplace [7]. Therefore, these marketplaces form a suitable case to apply and test gamification. It is worth investigating, as it is an interesting combination of a relevant societal subject and a promising but disputed method.

As a design case for our study, an online marketplace was found which suits our research approach. It is the mobile website of OLX India, which is an online marketplace for used goods. It is trying to grow its user base, but the existing users do not post a sufficient number of ‘listings’ (advertisements to sell used goods, created by users, thus UGC), which prevent both the demand and supply user sides from growing. Thus, the challenges of OLX very closely resemble the general online marketplace challenges.

2.2 Gamification Design Method

Looking for a gamification design method, a plethora of design principles, foundations, frameworks and perspectives derived from game design theories was found. No tested, proven and established single method is available [9, 24, 46]. However, one newly developed method was discovered. Based on the perspective described in the previous paragraph and taking into account the pitfalls and misuse of gamification, Deterding [12] created a design method called ‘Lens of Intrinsic Skill Atoms’. It is based on the interaction between user and system and incorporates motivational theory. It uses design lenses to tackle challenges and iterative prototyping to verify the outcome of preliminary designs until the desired outcome is reached. The design method was theoretically reviewed and slightly adapted for our purposes, resulting in the steps as shown in figure 2 on the next page. The adapted method has been applied to the OLX case.

2.3 Design Process & Resulting Treatment

The goal for OLX is to increase the number of new listings posted by users. One of the main challenges for OLX users is that they do not know what to sell. They do not think they have items of value to others in their home and do not recognise the types of items eligible for sale, thus do not post a listing. This was derived from OLX internal reports, interviews/discussions with OLX employees and a workshop with OLX users. The design process focused on this challenge. Initial ideas were developed and transformed into prototypes. These were evaluated and refined through two workshops, in which respectively five OLX users - Indian students at Delft University of Technology - and one OLX expert were involved. All the design lenses from the design method [12] were used as input during ideation and in both the workshops. The ideas and concepts that were used in the final design come from the following design lenses [12]: on-boarding, scaffolded complexity, interim goals, limited choice, appeal to motivations and template. After three prototyping iterations, a final design was selected: ‘the Selling Assistant’ (SA), featured in figure 3 on the next page. The underlying concept is that appealing explicitly to the identified user challenge with a call to action will engage users. The posting process is broken up into smaller steps. Users are asked to select an item to sell from a small category list, rather than to ‘sell something’. For each possible choice, the demand for a new listing in that category is made explicit, appealing to user motivations and making them aware of interested buyers. Two versions were created – ‘Suggest’ and ‘Know’ – which differ only in the wordings of their home page button. These two versions were created at request of OLX, to examine and correct for a possible effect of button wording on clicking behaviour.
Figure 2. Design method adapted from [12], as applied in this research

Figure 3. Selling Assistant workflow and experiment groups
2.4 Randomised Controlled Experiment

To evaluate the effect of the Selling Assistant an online, double-blind, randomised controlled experiment was used. Such an evaluation has a high general validity (if preconditions such as sample size are met) and is relatively easy to apply in our case [8, 16, 25, 33]. No personal user data were collected. The experiment was live for exactly 7 days, from Monday February 23, 2015, 19:30 (IST) until Monday March 2, 2015, 19:30 (IST). During the experiment, a percentage of users was randomly included into the experiment and uniformly divided over three groups: Suggest, Know (the treatment groups) and Original (the control group); see figure 3. In total, 51,104 users were included in the experiment: 16,744 in Original, 17,156 in Suggest and 17,204 in Know. The following variables were collected for each user:

- **Treatment ID**: Original, Suggest or Know;
- **Clicked**: whether or not the user clicked on the SA button on the home page (e.g. interacted with the treatment);
- **Listings**: number of new listings posted;
- **Page views**: number of pages viewed;
- **City**: from which city the OLX website was accessed;
- **Browser**: internet browser used to access the OLX website;
- **Source**: how a user reached the OLX website (referral link, search engine, direct visit or marketing campaign);
- **City tier**: based on city, common Indian city classification based on population size, infrastructure state, etc.

*Clicked and treatment ID* are the independent variables and *listings* is the target dependent variable. *Page views, city, browser and source* are extraneous variables used to examine interaction effects with independent variables, suggested by OLX based on their internal research. *City tier* was derived as extraneous variable from *city*, because the users in the design workshop indicated that they expected a difference in the effect of the SA within small and big cities; since the concentration and supply diversity of offline used goods shops in big cities is much higher than in small towns. Previous scientific studies also suggest incorporation of such demographic predictor variables [34].

The experiment was implemented with third-party web service provider Optimizely [44, 45]. For the data preparation, validation and analysis, the software packages R Studio (version 0.98), IBM SPSS statistics 22, including the PROCESS macro by Hayes [27], and Microsoft Excel 2011 were used. The target variable *listings* resulted in count data, which is non-normal but resembles a Poisson distribution with over-dispersion and excess zero’s. Therefore, a negative binomial regression model and the non-parametric Mann-Whitney U test were used as statistical methods [5, 21, 47]. For proportional data (proportion of users that posts a listings), a Z-test for proportions was used.

3. RESULTS

Table 1 features an overview of the tested alternative hypotheses (corresponding null hypotheses assume no change). Table 2 on the next page contains the main results of the experiment.

Users who interacted with the SA (by clicking on the home page button) posted more than 6 times as many listings on average during the experiment than users who did not interact with the SA (alternative hypothesis 1 accepted). Clicked was a highly significant predictor in negative binomial regression models for Suggest users (p < 0.01) and Know users (p < 0.01) – with *listings* as dependent variable and controlling for the extraneous variables measured. 2.1% and 2.4% of the users in the two treatment groups clicked on the Selling Assistant button. This result indicates that just a small fraction of the users interacted with our gamification treatment (N=354 in Suggest and N=411 in Know). The extra productivity of those users still contributed only slightly to the average number of listings per user, since the vast majority (>99%) of users did not post a listing at all. The users in the Suggest and Know groups posted respectively 18.2% and 18.9% more listings in total than users in the Original group. In a negative binomial regression model there was no significant effect of treatment ID on *listings*, so the influence of the SA on the OLX users depends on their interaction with this gamification treatment.

When comparing the Original variant to the Suggest and Know variants, a Mann-Whitney U test showed no significant difference in the number of listings per lister (alternative hypothesis 2a rejected). Nor was there a significant difference in the proportion of users that posted 1 or more listings, in a Z-test for proportions (alternative hypothesis 2b rejected). If hypothesis 2a or 2b was supported, this could have indicated more on the actual effect of the SA, as this would show if users were stimulated to post more listings or more users were stimulated to post a listing. A correlation between interacting with the treatments and posting content was confirmed, but – based on the experiment results – no causality can be confirmed. Did users post more because of the treatments, or is the correlation due to the fact that only the generally active users interacted with the treatments?

Examining the extraneous variables throughout the regression models showed that *page views, city, city tier and browser* were significant predictors of the general variance in *listings*. This confirms that demographic factors (*city and city tier*) have an influence on the content creation activity of users on an online marketplace [34]. However, significant mediation or moderation effects of these demographic variables on the relationship between *clicked* and *listings* could not be found in the results of our experiment. Therefore, effects of demographic factors on the way users are influenced by gamification – as suggested in literature [34, 35] – could not be validated.

The users in the Suggest and Know treatment groups showed no significant differences in posting behaviour, for neither of the hypotheses. This was expected, based on the small difference between their designs. It shows consistency between the two variants of the experiment.

### Table 1. Overview of tested alternative hypotheses

<table>
<thead>
<tr>
<th>#</th>
<th>Hypothesis</th>
<th>Accepted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Users who have interacted with the gamification treatment (the Selling Assistant) will post more listings on average than users who have not interacted with the gamification treatment.</td>
<td>Yes</td>
</tr>
<tr>
<td>2a</td>
<td>The average number of listings per posting user will be higher for the gamification treatment groups than for the control group.</td>
<td>No</td>
</tr>
<tr>
<td>2b</td>
<td>The proportion of users that posts a listing will be higher for the gamification treatment groups than for the control group.</td>
<td>No</td>
</tr>
</tbody>
</table>
4. DISCUSSION

4.1 Theoretical Implications

Our study and its results bear two important theoretical implications for gamification research. First of all, using gamification on online marketplaces to stimulate creation of user-generated content seems to be effective, as long as users actively interact with the gamification treatment. Only the users who fully experienced the SA by clicking on it showed a significant behaviour change. The fact that the mere presence of an extra button does not directly stimulate users to post more content was expected. However, we did not expect the share of users who clicked the button (and experience the full treatment) to be as low as 2%, limiting the overall treatment effect on the whole user base. Hamari implemented badges as a game element – using a more traditional gamification perspective – on online marketplace Sharetribe. His field experiment yielded results similar to our findings: users who actively engaged with the badges were significantly more likely to perform actions and create content on the marketplace. Also, Hamari could not prove an overall increase in user activity because of the general badge presence alone [25]. Based on these results, we would like to pose the question: does facilitative interaction undermine the effectiveness of a gamification treatment? In other words: when trying to engage an entire user base of an online marketplace, should gamification be incorporated into the core working of the marketplace, rather than being an add-on that can be voluntarily invoked?

Secondly, the adopted gamification perspective [10] and associated design method [12] are novel, within the already novel and manifold gamification field. The emergence of multiple perspectives and methods within the gamification field could be a positive development, in the sense that gamification is evolving from a single method into a more mature research field. However, it also creates confusion, because the term gamification refers to multiple phenomena, making it harder to identify and compare relevant scientific literature. This confusion fuels the existing scarcity of comparable and methodologically valid empirical gamification studies, as identified in [10, 19, 24, 34, 39, 46]. The design method by Deterding [12] partly copes with this by synthesising an existing variety of methods and design practices into one, with solutions to most of the known gamification criticism incorporated. This creates a potential single starting point for structured empirical gamification research, which can be compared and evaluated on the same grounds. Such a single starting point is facilitated in this study by providing thorough documentation of the design method as applied [4], which allows relatively easy reproduction within another case.

Furthermore, we want to theoretically reflect on two points. First, Hamari suggests that because of the utilitarian nature of a system, a hedonistic or affective treatment – such as badges – is ignored by the majority of the users [25]. In this study, both the gamification treatment and the platform itself are of a utilitarian nature. The Selling Assistant is a practical user feature, rather than a fun experience. Therefore, our results do not support Hamari’s hypothesis. We do think it is interesting to further examine the influence of a utilitarian or hedonic system and gamification treatment on the effectiveness of such a treatment. Second, the importance of user context in gamification design and effectiveness has been acknowledged literature [10, 24, 39]. One could argue that culture is an important element of this context, on multiple levels of abstraction (country, region, tribe, family, team, etc.) and the influence of culture on player behaviour in serious games has been described in an earlier study, also in India [36]. We expect that the regional cultural effects in our experiment are partly averaged out by the large number of users, but it is advisable to take into account the cultural background of users before the results are generalised to other groups.

4.2 Managerial Implications

Apart from theoretical implications, three recommendations for online marketplace managers can be derived from our study. First of all, gamification seems a good method to tackle specific issues for specific user groups that do not create content, due to a shared challenge they face. Such challenges can be identified and resolved by gamification treatments, which have been iteratively designed and tested. However, in general, practical hurdles for users to post content, next to motivational hurdles, could make it harder for gamification to be effective on an online marketplace than on a general online platform (such as a social network), where these practical hurdles do not exist. In the case in this research, if a user is theoretically convinced by the Selling Assistant that it is beneficial to post a listing in a certain category, he/she might still be practically limited by not having an item in such a category that he/she can sell.

Secondly, recommending the Selling Assistant to online marketplaces in general would be unwise, because it goes against the view of gamification as a socio-technical design process, which is highly influenced by specific contextual factors of a marketplace and its users. But on the whole, making the offer and demand on an online marketplace more explicit to its users does seem a valuable way of solving supply and demand imbalances. It corresponds with findings of what make online marketplaces successful: transparency and full information for users, no middleman, fast and easy transactions [23, 43].

Thirdly, the gamification design method [12] provides handholds outside of the normal recipe that is used when creating new features on an online marketplace. It forces the designer to categorise and connect both the users’ thought processes and actions, which is important when improving an online marketplace. Moreover, implementation of the design method into the current way of working of most online marketplaces will most likely be a seamless process, as it aligns with widely adopted methods of online companies, including the ‘lean start-up’ methodology [40] and concepts of continuous improvement and frequent user contact [31].

<table>
<thead>
<tr>
<th>Experiment group</th>
<th>Users</th>
<th>Listings</th>
<th>Listings per user</th>
<th>Listings through SA</th>
<th>Users who interacted with SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>16,744</td>
<td>143</td>
<td>0.0085</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>Suggest</td>
<td>17,156</td>
<td>169</td>
<td>0.0099</td>
<td>27</td>
<td>354</td>
</tr>
<tr>
<td>Know</td>
<td>17,203</td>
<td>170</td>
<td>0.0099</td>
<td>26</td>
<td>411</td>
</tr>
</tbody>
</table>

Table 2. Main experiment results
4.3 Limitations
What type of user was stimulated by the gamification treatment in this study cannot be discerned from the experiment data. Are they users who were actually struggling with the problem of not knowing what to sell, or are they users who are generally active, who like to experiment with new features (and therefore clicked on the Selling Assistant button) and who were planning to post content anyway? Incorporating qualitative information on users’ motivations and historical behaviour on OLX would have created a more complete picture of gamification effects. Using this approach, more observed correlations could have been confirmed as causal relations.

It’s relatively easy to agree with Deterding’s social-technical perspective on gamification [10], since he summarises the relevant gamification developments of the past years, reflects on the definitions posed (including his own) and incorporates and acknowledges much of the criticism gamification has received. His perspective and design method have proven to be effective in this first empirical application, but general statements on its effectiveness cannot be made at this stage. Moreover, as we saw in the empirical study by Hamari [25], implementing one simple game element – badges – on an online marketplace can yield similar results as this study produced. Also, a recent empirical study by Lieberoth showed that participants who interacted with a gamified system (game look and feel with simple game elements) showed the same increase in overall motivation (measured in terms of ‘value’, ‘enjoyment’, ‘autonomy’, ‘relatedness’ and ‘competence’) as participants who interacted with a system that had all the interface elements of a gamified system, but no actual game mechanics or dynamics [35]. In other words: applying a thorough and structured user-centered design method (as we did in this study) is not necessarily more effective than using game elements or game interfaces only.

5. CONCLUSIONS & RECOMMENDATIONS
This study set out to investigate the effect of gamification on the amount of content generated by users on an online marketplace. OLX was used as an example online marketplace, which demanded for more user-generated content. A novel but theoretically sound gamification perspective was adopted and the associated design method was used to create a gamification treatment, through several iterative prototyping and ideation workshops with OLX users and experts. The effect of gamification was notable in a randomised controlled experiment: users who interacted with our treatment posted a significantly larger amount of content (p < 0.01) than users who did not interact. However, only a small fraction of users (2 – 2.5%) chose to interact with the treatment, limiting the total increase in listings between the original and treatment groups. This conclusion closely aligns with a similar empirical gamification study conducted before [25], which supports the generalisability of the findings. An influence of demographic factors on the effect of gamification could not be found. Lastly, the adopted design method provides a possible starting point for structured empirical gamification research. Many questions regarding the types of users that were affected remain unanswered, which confirms the need for more empirical research.

For future research, we want to make three recommendations based on our results. First of all: it is important to connect the behavioural effects of the treatment to changes in motivational elements of competence, relatedness and autonomy of users [10, 24, 35, 41, 42]. However, one needs to be aware that a trade-off exists between, on the one hand, working with a small user base, which is aware of being in an experiment (not double-blind, so more bias) and which allows for psychological measurements and, on the other hand, working with a large user base in a double-blind experiment, like we did, which allows for measuring only behavioural data. Principally, we recommend using a mixed methods experimental approach that incorporates behavioural experiment data, historical user data and psychometrics. Our second recommendation is to use two control groups, of which one receives a dummy treatment. Also, the experiment timeline should be at least two weeks, to correct for ‘novelty’ or ‘Hawthorne’ effects (users displaying an inherent positive attitude towards new features) [24, 32, 34, 35]. Thirdly, an important point to investigate in more depth is the difference between applying gamification from the systemic [11], experimental [28] or socio-technical design perspective [10]; preferably within the same case. This will allow the gamification (design) research to become more evidence based and adopt one single design method as starting point for empirical research.

For a more elaborate description of this study, including the design process and experiment results, please refer to the full MSc thesis [4].

6. ACKNOWLEDGEMENTS
We would like to thank the people from OLX involved in this research, who very generously made their platform available for our research.

7. REFERENCES


