Shih, J. L. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Examining the Effects of Leaderboards in Gamified Learning Environment

Shurui BAI
Faculty of Education, The University of Hong Kong, Hong Kong SAR
shurui18@hku.hk

Abstract: Gamification is widely defined as the use of game design elements in non-game contexts. Recent years have witnessed a significant growth of gamification studies, and the most commonly used game elements among the studies are the leaderboards. Research findings on the effects of leaderboards so far have been mixed. On one hand, previous studies have shown that leaderboards can motivate participants to perform more tasks and improve their learning performance. On the other hand, leaderboards can also demotivate participants in completing their learning activities. This study aims to systematically investigate the effects of different variations of leaderboards using experimental designs in the higher education context. More specifically, this study will manipulate three main independent variables-anonymity of learners' identity, learners' ranking position, and learners' group size on a leaderboard. The dependent variables include participant academic performance, course engagement, participant intrinsic motivation and participant perception. Three hypotheses will be proposed and tested.

Keywords: Leaderboards, gamification, academic performance, motivation

1. Introduction and Previous Study

In recent years, gamification has become a buzzword across various academic disciplines. It is often defined as the application of digital game elements in non-gaming contexts to motivate user behavior (Deterding et al., 2011). It has been proposed that gamified practices will become a key element in motivating people to complete certain tasks such as improving their fitness (Nike+), answering other people's questions (Yahoo! Answers), or completing learning activities.

It is important to note that findings from previous comparison studies are limited because they usually included more than one game mechanics. This makes it impossible to parse out the effects and pinpoint a specific causal factor. For example, Hew et al. (2016) conducted two experiments to examine whether a digital points-badge-leaderboard system (Experiment groups) would promote superior quality of student work when compared to students without any digital game mechanics (Control groups). Results found that a gamified system positively motivated students to engage with more difficult tasks and produced better student artifacts than the control condition.

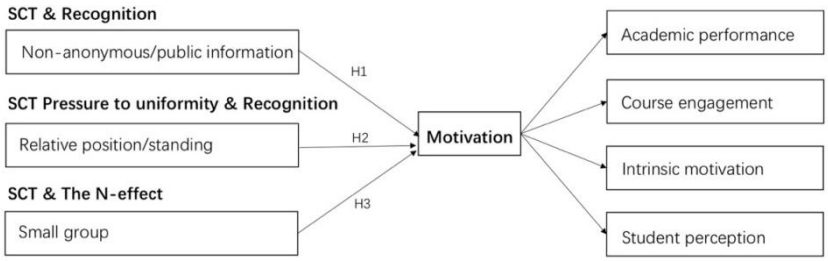
This study focuses only on one specific game element – the leaderboard, which is one of the most commonly used game elements in previous gamification studies (Sailer et al., 2014). Leaderboard is a high-score reputation table which rank users by their performance. Leaderboards satisfy a person's need for achievement by comparing his or her own abilities against others'. By displaying a list of players' ranks upward or downward, an order of competition against success criteria are available. However, in most cases, only learners on the top of leaderboards get satisfied and engaged, those at the bottom leave off comparison (Sailer et al., 2014).

In this study, we aim to investigate the effects of different variations of leaderboards on participant academic performance, course engagement, participant intrinsic motivation and participant perception. This study will manipulate three main independent variables – learners' identity, learners' ranking position, and learners' group size on a leaderboard. Three hypotheses will be proposed and tested.

2. Hypotheses

2.1 Hypothesis about Anonymous Information for Comparison

According to *Social Comparison Theory*, people have the tendency to use attained information to evaluate their opinions or abilities with others to obtain accurate appraisal when objective criteria are absent (Festinger, 1954). Public information about people's relative position helps increase their motivation effects (Webster et al., 2003). People will be motivated more if his rank is high, and is also known by others, but the motivation will be diminished if no one knows (Mcleod, 2011). Based on the social comparison theory, we propose the first hypothesis about the anonymity of personal identification on leaderboards setting: *H1: Non-anonymous personal identification on leaderboard setting enhances individual academic performance, engagement, intrinsic motivation; and students hold a positive perception towards it than the anonymous counterpart.*



Note: SCT Social comparison theory

Figure 1. Theories underpinning hypotheses of motivation

2.2 Hypothesis about Positioning Scheme

Leaderboards type can be divided based on positioning scheme: the relative positioning (also named no-disincentive leaderboards) and the absolute positioning (also named as infinite leaderboards). Relative leaderboards only present players above or below him/her. The absolute leaderboards display the literal ranks of every player. Players are always under threats of being beaten and falling off. The absolute leaderboards are commonplace in gamified learning (e.g., Özdener, 2018). According to the *pressure to uniformity* under *social comparison theory*, the effect is strong when people make comparison with those who are close to their abilities or opinions (Festinger, 1954). For example, learners prefer to compare their course assessment scores with competitors with similar levels. Therefore, relative leaderboards are assumed to perform better. The hypothesis is: *H2: Relative positioning scheme enhances individual academic performance, engagement, intrinsic motivation; and students hold a positive perception towards it in comparison to the absolute positioning scheme.*

2.3 Hypothesis about Group Size

The N-effect under social comparison theory manifests the negative relationship between the number of competitors (N) and a person's competitive motivation on individual tasks (Garcia & Tor, 2009). Social comparison becomes weak and competitive motivation diminishes under the environment of a large number of competitors. In one of Garcia's empirical study (2009), individuals completed a quiz significantly faster when they believed they competed with 10 rather than 100 other participants. Group sizes mediate the user's contribution to a community. Leung (2019) manipulated group sizes in three levels (10, 20 and 50) to compare user's response to different scores of their peer groups. In the educational field, group size is often confined to class size. Given the N-effect, we hypothesize that larger group size dampens individual motivation on student learning tasks. This leads to the third hypothesis: *H3: Small group size enhances individual academic performance, engagement, intrinsic motivation; and students hold a positive perception towards it.*

3. Methods

3.1 Participants

Participants will be recruited from a university in an Asian region. After consent forms are signed by individuals; demographic information will be collected for further data analysis. Gender is not presumed to be one influential factor, but the gender ratio will be reported for educational implication. The intervention duration for each sub-experiment is assumed as one semester (8-12 weeks).

3.2 Experiment Design

We will gamify our intervention by incorporating a plugin module called Level Up! to Moodle. This module enables learners to earn experience points in their courses. Level Up! can grant students a corresponding rank and badge automatically when they collect enough points. It is a user-friendly module which allows educators to manually display students' personal identity or not, so we can make comparison between anonymous and non-anonymous groups for the sake of the first hypothesis. As for the second hypothesis testing, we can activate the relative ranking or the absolute ranking by clicking a button "ranking" in setting. In the third experiment, we will recruit three classes (i.e., small, medium and large group sizes) to examine the effects of different group sizes.

3.2.1 Experiment One for H1

One class in a university will take part in this experiment within one semester to test the first hypothesis. We apply group sequential study, which means the treatment class will expose to multiple interventions in a set order (Thompson & Panacek, 2006). This design helps lead to a conclusion much earlier and allow a near-perfect match of participants. Internal validity of group sequential study does not depend on random assignment and has a higher statistical power (Charness et al., 2011), but the carryover effects may occur. We will completely counterbalance the effects by making the treatment order an independent variable. So different participants are exposed to different orders of treatments. There are two interventions in the treatment class: anonymous leaderboards (AL) and non-anonymous leaderboards (NAL). It is to say, participants in the treatment class will be divided into two groups by random draw, we apply AL-NAL order to half of the students and NAL-AL order to the rest of students at the same time.

3.2.2 Experiment Two for H2

A two-arm study is conducted: the relative positioning group and the absolute positioning scheme group. We administrate this experiment in a postgraduate course by randomly dividing a class of students into two groups of an equal number. The experiment will last for one semester and the pre-tests are conducted at the beginning of the intervention. We videotape every class activity and transfer the

in-class points into the learning management platform Moodle. The measurement of intrinsic motivation and engagement will be conducted after the intervention by two scales. Whether to use anonymous or non-anonymous leaderboards depends on the result of the first experiment.

3.2.3 Experiment Three for H3

Finally, we set up a favorable dashboard based on results from experiment one and two. Then we apply those leaderboards into three group sizes (i.e., small, medium and large). We have no pre-knowledge about the sample size, so we assign the large group size five times more than the small group size and twice more than the medium group size to maximize statistical power (Leung, 2019). For example, Leung (2019) manipulated group sizes in three levels (10, 20 and 50). Students initial levels, prior knowledge will be tested and measurement on intrinsic motivation and engagement will be conducted after the intervention.

3.3 Measurement of the Research Outcomes

As for individual achievement performance, we measure student academic performance partially through in-class and out-of-class activities. The preliminary points-adding criteria are designed based on teaching objectives. We also measure individual student academic performance via academic tests (e.g., final test scores, midterm test scores, and standardized test scores). Exams or assignments developed by instructors are common in the higher education context.

With regards to course engagement, two instruments were consulted: a) the Student Course Engagement Questionnaire (SCEQ) proposed by Handelsman, Briggs, Sullivan, and Towler (2005) and b) the self-developed points-adding criteria, by which we can measure student obvious participation engagement. Since emotional engagement is hard to measure by observation, and we know little about students who fail a certain assignment or test; therefore, SCEQ scale can complement our measurement by taking account of skills, participation/interaction, emotion, and performance engagement. We use Intrinsic Motivation Inventory (IMI) to measure the intrinsic motivation of individual students. After the intervention of three experiments, an interview will be held to understand student perception towards leaderboards in terms of anonymity, positioning scheme and group size.

References

Charness, G., Gneezy, U., & Kuhn, M. A. (2011). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization*. Doi:10.1016/j.jebo.2011.08.009.

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

Festinger, L. (1954). A Theory of Social Comparison Processes. *7*(2), 117-140. Doi:10.1177/001872675400700202.

Garcia, S. M., & Tor, A. (2009). The N-Effect: More Competitors, Less Competition. *Psychological Science, 20*(7), 871-877. Doi:10.1111/j.1467-9280.2009.02385.x.

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A Measure of College Student Course Engagement. *The Journal of Educational Research, 98*(3), 184-192. Doi:10.3200/JOER.98.3.184-192.

Leung, W. (2019). *How Do One's Peers on a Leaderboard Affect Oneself?* Paper presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk.

Mcleod, P. L. (2011). Effects of Anonymity and Social Comparison of Rewards on Computer-Mediated Group Brainstorming. *42*(4), 475-503. Doi:10.1177/1046496410397381.

Özdener, N. (2018). Gamification for enhancing Web 2.0 based educational activities: The case of pre-service grade school teachers using educational Wiki pages. *Telematics and Informatics, 35*(3), 564-578. Doi:10.1016/j.tele.2017.04.003.

Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2014). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s), 19*(1), 28-37.

Thompson, C. B., & Panacek, E. A. (2006). Research study designs: Experimental and quasi-experimental. *Air Medical Journal, 25*(6), 242-246. Doi:10.1016/j.amj.2006.09.001.

Webster, J. M., Duvall, J., Gaines, L. M., & Smith, R. H. (2003). The Roles of Praise and Social Comparison Information in the Experience of Pride. *The Journal of Social Psychology, 143*(2), 209-232. Doi:10.1080/00224540309598441.