# Effect of rate of appearance of in-game items on user behavior in educational gaming

A short summary of a research project conducted on the PaGamO platform

## Question

How is user behavior affected by changes in the rate at which items (stimuli) in an educational game appear increase or decrease?

## Background information

On the PaGamO platform, users attempt to conquer as much territory as possible and build a large empire by answering questions. As they conquer land, users are regularly presented with in-game items, including potions, terraforming tools, and tools used to attack opponent lands. Previous research conducted by the company demonstrated that presenting users with more tools for attacking opponent lands leads to increased usage and improved retention rates.

## Research methodology (provided by BoniO as-is)

Test subject: 200 active users, divided into four equal groups. All these groups generally have the same retention rate, answered question, and attacking rate.

Research questions as follow

1. Would the retention rate (answered question, and attacking rate) increase if the offensive items dropping rate increased?
2. Would the retention rate (answered question, and training rate) increase if the terrain items dropping rate increased?
3. Would the retention rate (answered question, attacking rate, and training rate) increase if both offensive and terrain items' dropping rate increased?

| | No dropping | High offensive item dropping |
|---|---|---|
| No dropping | 50 | 50 |
| High terrain item dropping | 50 | 50 |

We would like to know:

(1) How would you conduct this experiment, including how much you want to increase the items' dropping rate?

(2) What are the numbers you want to collect from the experiment?

(2) Based on your findings, how do you suggest to proceed to next experiment or to develop new feature/ algorithm in our platform?

# Potential Changes

Whether or not this specific experiment (this specific research question) is adopted is actually still up for debate. This specific topic is limited to the item-dropping rate, the effects of which will very likely vary from one game to another. The ability to draw generalized conclusions that will have any consequence for the educational gaming industry is, thus, quite limited. A possible way to fix this is to add more variables to the research. The company's representative did suggest using location (part of Taiwan), gender, and any demographic information as variables. However, this will introduce more complexity while not achieving very much. Therefore, another way to do so is to amend the research question to be more generalized to help us better draw conclusions that are applicable across the industry. An example would be going from "item-dropping rate" to perhaps "stimuli" or something just as generalized. This can avoid the comparability issue. In this case, there would be more emphasis on categorizing and coming up with a way to make sure that the stimuli in PaGamO is analogous to those in other games, or at least on making it possible to compare the amount of stimuli across games.

There are many other potential changes to this methodology proposed by the company. The company, for example, recently developed an item response theory (IRT) model to model the ability of each student in a particular course or subject area and use mathematical inference to adjust their estimates based on the students' answers to practice questions. This improved model for student learning can be used to help evaluate the efficacy of the program, instead of using the retention rate. Specifically, we can focus on the changes in the IRT model's estimate of the user's understanding of the course content of a particular course. Not only does this showcase the application of this research in gaming as a whole -- it focuses on the effects of these changes on the *educational* effects of the game. This can possibly

provide more value to this research as a significant amount of research has already been done on the gaming industry as a whole -- but educational gaming is a different challenge. Ultimately, using the IRT model as a benchmark for learning was the use in developing the system to begin with and therefore it is only natural to use it for this experiment as well.

Furthermore, 200 active users may be too few for this research project. It would certainly contribute to data accuracy if more active users were pooled into the experiment. Specifically, an issue that this can create is that the research project necessarily *has* to be limited in scope because this limited number of users can only be tested in a very limited number of tests. Therefore the research project will have problems being generalized as aforementioned.

Another potential issue is the interaction between the several variables tested. We posited that the various variables tested (two, that is) will not affect each other and instead the relationship between the two will be multiplicative. This, however, cannot be tested while there are only two variables. In the long term, more trials will need to be done to draw conclusions on the relationship between the two variables.