* + 1. Virtual Atom Smasher and GeoTag-X: analytics use cases

**Analytics data definition**

The first step in setting up learning analytics in a CCS project is the definition of the analytics data, i.e. what we are going to measure and how we are going to measure it. As analytics are the low-level data that allow for the exploration of the learning and the engagement of the participants, the definition of the analytics must stem directly from the learning objectives that are initially set for each project. During several meetings and discussions with the project leaders and developers, the team agreed on a set of expected learning outcomes (ELO) (see Table 2, page 36), on the corresponding indicators to assess each ELO, and finally on the low-level analytics data (also called analytics events) necessary to compute the indicators.

For example, in both projects, the ratio of bad clicks around the project interface is an indicator of how well the interface is mastered. Each time a participant clicks in a wrong place, an event is generated and later we can compute several indicators, depending on the level of detail that we want to analyze. For instance, a global ratio of bad clicks can be computed and analyzed. Observing the evolution of this ratio through time might reveal if the users are improving and learning how to use the interface appropriately. We can also specify in the analytics where the bad clicks occur, and then we can infer on the usability of different parts of the interface.

A more complex ELO might involve many different analytics events and even to track the chronological sequence of events. For instance, we consider that a Virtual Atom Smasher participant acquired the adequate playing strategy when (s)he manages to execute a set of actions in a specific order. A participant is playing properly when (s)he changes the parameter value, clicks the button estimate, and finally clicks validate. In this case, the needed analytics data involve user clicks on the buttons “*change parameter”*, “*estimate”*, and “*validate”*. Among the many potential learning indicators that reflect the playing strategy, we mention the number of parameter changes applied before a validation, and the quality of the estimation represented by a value called “goodness-of-fit”.

Analytics data collection

Once the analytics data are defined, they are implemented through the application using the CCLTracker library. CCLTracker provides a high-level API to ease the implementation of complex monitoring tasks, such as, time watching a video, or time spent doing a task. CCLTracker is connected to Google Analytics (GA) through Google Tag Manager, thus reducing the dependency with GA and easing the migration to other technologies. GA provides a GUI to easily visualize and manipulate analytics data. Basically, GA allows to create segments (e.g. users who complete one task only, users who apply the expected strategy, etc.), crossing data (e.g. gender of users who are contributing more than 50 tasks), or filtering (e.g. top 10 referrals bringing more than 100 users to the application). GA is limited regarding the aggregation of data. Thus, more complex aggregation and manipulations need to be implemented in specialized statistical packages.

Analytics processing

An R package called “rga” is used to export analytics data from GA to R. The data used to assess the learning and engagement, are typically collected according to the following format:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *User Id* | *Combined event* | *Complementary info* | *Date* | *Hour* | *Minute* |

* **User Id** is an anonymous identifier that the developers define and make sure it is shared among all the data sources of the activity of the participants, i.e. analytics, online surveys, in-app questionnaires, etc.
* **Combined event** is an event tag combined with some extra information. The extra information can, for instance, inform on the part of the UI where it occurs, or on the answer of the participant in a related activity.
* **Complementary event info** is a field that does not always contain relevant information. It might be used for some types of events to give a timestamp, a percentage of completion of an activity, or the duration of time spent on the activity.
* **Date, hour** and **minute** are the timing information provided by Google for each occurring event. As Google does not provide timing at the second level, in some cases, a more precise timestamp is needed in order to chronologically order the events, and in this case it can be transmitted in the complementary info field.

Typically, on a normal active day, participant activities on Virtual Atom Smasher and GeoTag-X can generate a few thousands events. These events are collected through the API and then processed and aggregated to fit the needs of the statistical analysis.

In order to combine the analytics data with the survey data that we have at hand, we generate a dataset of the format:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *User Id* | *Indicator1* | *Indicator2* | *…* | *IndicatorN* |

The indicators are aggregated values computed to reflect aspects of the learning or the engagement of an individual participant identified by the User Id. The set of indicators that we used in a typology analysis are depicted in the following table:

|  |  |  |
| --- | --- | --- |
|  | **Engagement Indicators** | **Learning Indicators** |
| **GeoTag-X** | * total duration in days * nb. of active days in the whole duration * mean nb. of events per day * Total nb. of events in the total duration * peak nb. of events per day * skewness of nb of events | * mean nb of bad clicks per day (independently of the location of the bad click) * ratio of bad clicks (BCR) * total nb. of started tutorials * total nb. of completed tutorials * total nb. of skipped tutorials * mean time before skipping tutorials * mean time to complete tutorials * total nb of distinct completed tasks (info provided from server-side) * total nb of started tasks * total nb of completed tasks (not necessarily distinct) * mean-min-max-standard deviation of task duration |
| **Virtual Atom Smasher** | * total duration in days * nb. of active days in the whole duration * mean nb. of events per day * Total nb. of events in the total duration * peak nb. of events per day * skewness of nb of events | * mean nb of bad clicks per day (independantly of the location of the bad click) * ratio of bad clicks (BCR) * mean nb. of interface change per day * ratio of interface change (CR) * mean nb. of estimations per day of activity * min fit value achieved in the whole period of activity * mean fit value achieved per day of activity * nb of changes on parameters * nb of validations of the changes * mean nb of changes before a validation * mean scale value for the changes * nb of taken questionnaires (–very small – omitted) * mean score (– omitted) * nb of started tutorials * nb of wrong answers in the intro course * nb of correct answers in the intro course |

Table 5 Indicators used in a cluster analysis of participants in Virtual Atom Smasher and GeoTag-X.

Notice that to represent the activity intensity we use the number of events instead of the time devoted on each active day. Although GA provides a measure of session durations, but this measure is not very accurate (ref. <https://support.google.com/analytics/answer/2731565?hl=en>), hence we preferred to count the number of events generated by the activity of the participants. This measure of number of events is only meaningful to compare participants in the same project and on the same period of time, as the events are defined by the developer and could change through different versions. This is for instance the reason that prevented from collecting analytics of GeoTag-X starting from May when the last version was published, as the full implementation of analytics ended on the 5th of August.

This represents one of the important lessons learned from working with learning analytics: Analytics implementation should be considered starting from the design phase subsequently to the definition of the learning objectives and should be implemented and tested as intrinsic part of the project so that they could be deployed as soon as the project is operational.

In contrast with Zooniverse projects that ran for more than 18 months with tens of thousands of participants, our pilot projects ran for shorter periods of time and have gone through different versions with the needed analytics implemented gradually. All these factors limited the interesting period of analysis to be about three months for GeoTag-X and to less than 6 months for Virtual Atom Smasher both with less than 300 hundreds identified participants. Nevertheless, we applied the profiling approach we proposed summarized in Table 5 to see if it is possible to discover short-term engagement-learning profiles.

### Analysis of the learning analytics of VAS

We conducted a cluster analysis on our VAS analytics dataset with the objective of identifying participant types with respect to their engagement and learning indicators (presented in Table 5, page 49).

A hierarchical clustering is implemented and discussed here. Additional experiments with k-means clustering algorithms were also conducted providing similar typology of the participation that the hierarchical clustering offers.

We conducted the hierarchical cluster analysis of VAS users through a period of nearly 6 months, from the 8th of May till the 20th of October 2015. During that period we gathered learning analytics from 511 users.

#### VAS typology

A hierarchical cluster analysis allowed identifying four distinct groups of users as shown in Table 1 below. The different groups of VAS users are named as follows:

***One-day passing*** is the user group who did not interact a lot with the game as they generated few events, and most of them did not apply any parameter validation. This group contains 291 users (more than 50%). Every user connected to the game during one day, on average.

***Ephemeral*** represents the group of users who generally returned to the game shortly after the first introduction but never came back (few of them returned in a period of three months). This group contains 151 users (33%). What distinguishes this *ephemeral* group is that they had slightly more activity than the *one-day passing* group and far less than the other two groups.

***Struggling*** is a group that represents about ten percent of the participants. They seemed to struggle with the game. Hence we call them the “*struggling*” group. These participants played but they apparently did not manage to achieve very good scores on the game. Precisely, they made many parameter changes and more validations on average than the other groups, but they also have higher goodness-of-fit values (ref. D4.7 3.4.5 for an explanation of the goodness-of-fit parameter in VAS), which indicates that they did not know how to estimate well. Also participants in this group have higher ratios of interface change than the first two groups, this indicates that they spent their time moving around the game and probably trying to figure out how to play. It is also noticeable that abide those who only visited the game for one day, most of these *struggling* people returned nearly daily to the game on short periods (less than a week) and then abandoned, and this explains the high activity ratio of this group. During their activity, the *struggling* group generates more events than the remaining *committed* group as we can see from the indicators (Table 1): mean number of events per day, mean number of estimations, mean number of validations and mean number of changes on parameters.

***Committed*** group includes a minority of the participants (12 users) who returned to the game on periods that exceeds 50 days, but not necessarily regularly since the group achieves a low activity ratio (only one user achieved an activity ratio of 70% while 8 other participants achieved less than 10% ratio). We describe this group as *committed* because they interacted with the game for relatively long periods. It appears that this small group understood the mechanics of the game as we will discuss shortly.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant type (cluster groups) | One-day passing | Ephemeral | Struggling | Committed |
| Size | 291 | 151 | 57 | 12 |
| nb. of active days in the total duration | 1.03 | 1.44 | 2.05 | 12.50 |
| total duration | 1.04 | 4.13 | 4.89 | 101.92 |
| activity ratio (nbActivDays/duration) | 1.00 | 0.87 | 0.92 | 0.15 |
| mean nb. of events per day | 27.26 | 83.19 | 399.14 | 135.20 |
| total nb. of events in the total duration | 29.19 | 127.30 | 752.74 | 1734.83 |
| peak nb. of events per active day | 28.37 | 105.38 | 542.84 | 569.92 |
| skewness of nb of events | 0.00 | 0.10 | 0.08 | 1.30 |
| mean nb of bad clicks per day (independently of the location of the bad click) | 4.80 | 24.05 | 72.19 | 15.13 |
| ratio of bad clicks (BCR) | 0.15 | 0.31 | 0.19 | 0.12 |
| mean nb. of interface change per day | 14.13 | 30.17 | 235.59 | 96.27 |
| ratio of interface change (CR) | 0.48 | 0.30 | 0.58 | 0.68 |
| mean nb. of estimations per day of activity | 1.09 | 1.34 | 33.50 | 9.65 |
| min fit value achieved in the whole period of activity | 2.95 | 3.03 | 1.89 | 1.83 |
| mean fit value achieved per day of activity | 3.12 | 3.38 | 7.07 | 2.98 |
| nb of started tutorials | 0.13 | 1.85 | 3.68 | 5.33 |
| nb of wrong answers in the intro course | 0.08 | 12.50 | 22.26 | 0.00 |
| nb. of correct answers in the intro course | 0.00 | 1.04 | 1.35 | 0.00 |
| mean nb. of validations per day of activity | 0.32 | 0.90 | 5.72 | 2.71 |
| mean nb of changes on parameters per day of activity | 2.04 | 5.37 | 81.50 | 23.62 |
| mean scale value for the changes | 0.20 | 0.15 | 0.15 | 0.13 |
| nb of changes before a validation | 4.66 | 5.66 | 18.44 | 10.93 |

Table 1 The mean values of the indicators of VAS activity according to the different participant groups

Table 1 and the bar charts in Figure 2 show that the groups have distinct engagement profiles (the first 7 indicators in Table 1), but the clear separation between the groups relies also strongly on the indicators that reflect the learning outcomes.

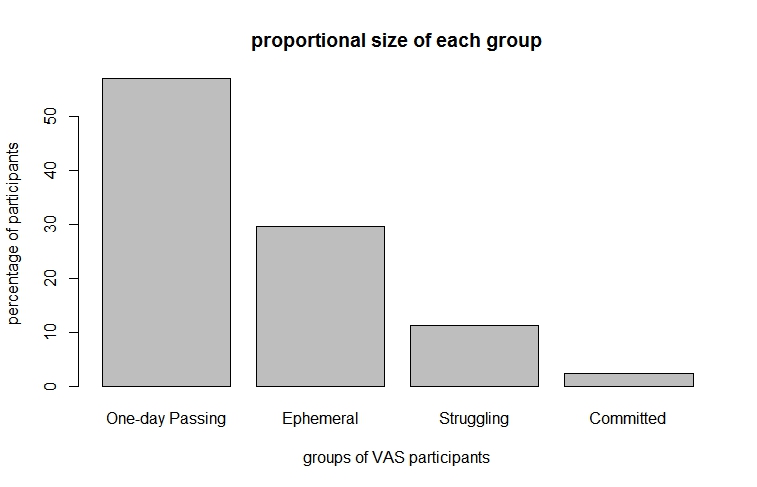
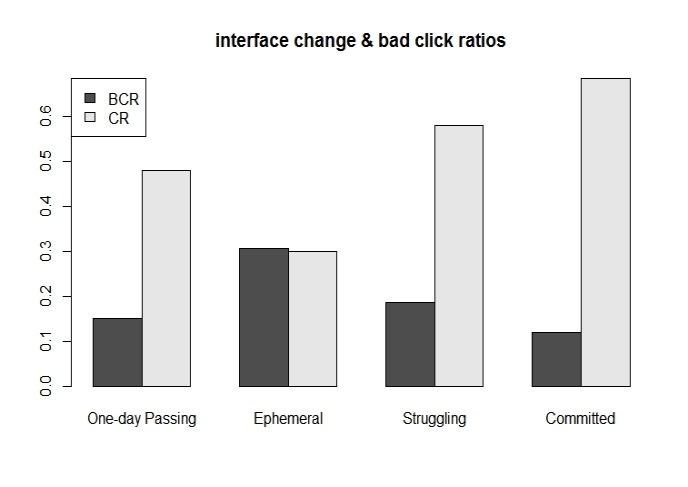
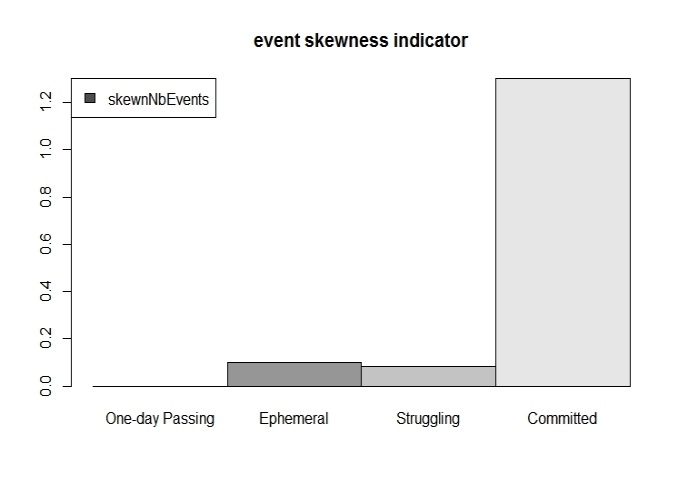
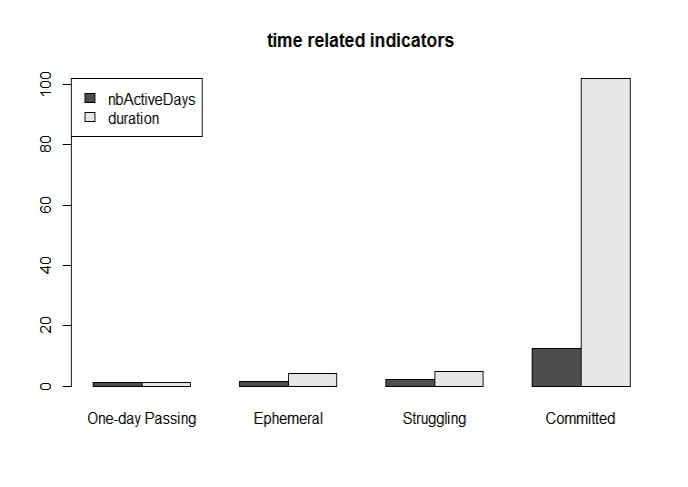
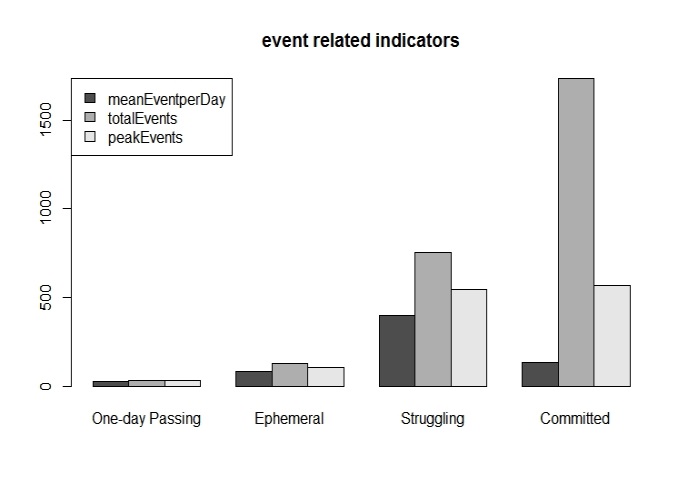
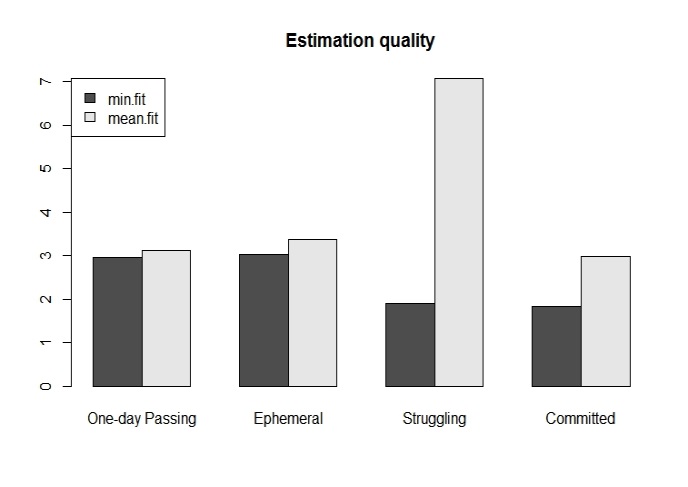
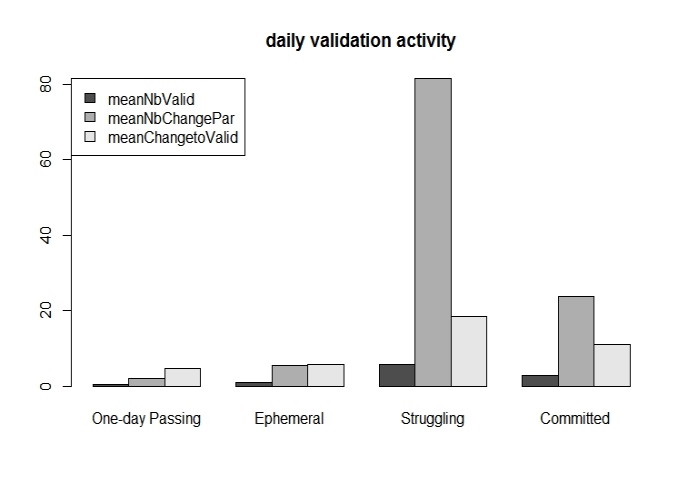


Figure 1 Distribution of VAS participant types in terms of number of participants

Figure 1 shows that around 60% of the users participated just for one day. Around 38% of the users participated for at least 2 days and finally 2.3% of them stayed connected to the game for longer periods.



(a)

(f)

(b)

(e)

(d)

(c)

Figure 2 Some indicators of VAS activity according to the different participant groups.

Figure 2 shows the values of some selected indicators of the group centers. Figure 2 (a) contrasts the longer period during which the *committed* group returned to the game with the short periods after which the other groups entirely disconnect from the game. Figure 2 (b) shows that the struggling group has more intense activity than the *committed* group. Their peak numbers of events are comparable, and the *struggling* group even generated more events on average than the committed group. It appears that these participants had interest in the game but lost it so they abandoned after few days, Further analysis of their other data could explain the reasons of this abandon.

The heavier activity of the *struggling* group is also visible in Figure 2 (c) that show their higher number of parameter changes and validations. Figure 2 (d) reveals that they made estimations of worse quality as their goodness-of-fit is greater than 4 on average, while the other groups had an average goodness-of-fit close to 3 which is considered as a good value in contrast with values greater than 4 which are considered of average quality (cf. D4.7 3.4.5)

Figure 2 (b) and (c) show that the majority of the users, i.e. the *one-day passin*g and the *ephemeral* groups did not generate a high number of events and they applied a very limited number of validations.

Figure 2 (d) represents the skewness of the number of daily events, this skewness is mostly meaningful for the *committed* group who kept the connection with the game for long periods. This skewness is positive which reflects that the committed participants were more active in the first days, and their activity reduced with time.

Finally, Figure 2 (f) shows that the *ephemeral* group has a higher ratio of bad clicks and high ratio of interface change than the other groups. The *one-day passing* group has the highest ratio of interface change. These two indicators indicate that the users explored the game but not necessarily knew how to interact with it.

Globally, this first-hand typological analysis reveals two important findings. First, the majority of the users (more than 85%) stops after at most two visits. Second, 10% of the participants who are in the struggling group had an intense activity compared to the other groups but they disconnected from the game shortly (in less than a week). This finding direct us to further explore the behavior of this group and understand the reasons of their abandon.

The typological analysis also reveals that there is a minority, which is the *committed* group, who understood the mechanics of the game as they have the smallest bad click ratio (BCR) and achieved good goodness-of-fit values, and had a good playing strategy since they did not apply too many parameter changes. On the other hand, the count of daily events of this group is positively skewed which indicates that mastering the game is not enough to fully engage the participants, as they lose interest with time. This indicates that actions should potentially be taken to keep these participants engaged.

Table 2 shows the correlations between the minimum goodness-of-fit and the engagement indicators, the negative values, especially the significant negative correlations with the number of active days and the total number of events, suggest that stronger engagement leads the participants to better estimate the parameters of the game with time.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | nb. of active days in the total duration | total duration | Activity ratio | mean nb. of events per day | total nb. of events in the total duration |
| min fit value achieved | -0.60646 | -0.05115 | -0.42259 | -0.30411 | -0.71289 |
| *significance* | *0.03656* | *0.8746* | *0.1711* | *0.3365* | *0.009258* |

Table 2 correlations between the engagement indicators and the minimum goodness of fit achieved by the participants in the committed group

This correlation between the minimum goodness-of-fit and the number of active days and the total number of events are also significantly negative when they are computed for all the participants (*resp*. -0.3249736 and -0.5054271). Hence we can see here an explanation of one of our research question (*RQ1.3)* which stipulates that different patterns of engagement could lead to different learning outcomes (in occurrence here, learning how to estimate parameters in VAS). The *committed* group has an average minimum goodness-of-fit of 1.83 which is lower than the global average which is 2.51. and it seems from the correlation analysis that smaller values can be achieved with higher time engagement with the game as it is the case of the *committed* participants.

This clustering analysis opens interesting questions: “Who are the *committed* users?”, and “How did they learn to estimate parameters?”.

An answer to these questions can come from the segmentation tool available on Google Analytics web interface that allows applying segmentations. Though we track two segments of users: Those who are following the expected flow of actions and those who are not (i.e. they were not playing as it was expected). Namely, the right flow of actions is: (1) change parameter values, (2) click estimate button, (3) repeat 1 and 2, and finally (4) validate.

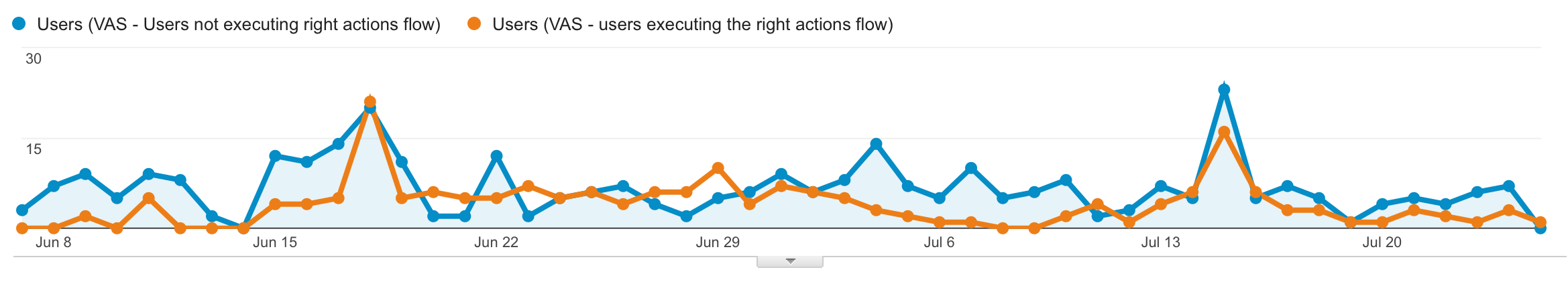


Figure 3 A timeline representing the evolution of the number of VAS participants who are playing according to the xepected flow of actions (orange line) and those who are not (blue line)

(The period depicted here is included in the analytics collection period)

Figure 3 shows the number of users per day who were following the right flow of actions (orange line) versus the number of users per day who did not follow the right actions flow (blue line). We observe that the 17th of Jun and 15th of July users played properly the game. On those days, the VAS developer leader was giving seminars at CERN introducing the game and helping new users to play it. Consequently, users were playing properly for few days after the seminar. Later, new users arrive and they had problems to find out the right way of playing. Based on this information revealed by the analytics, the VAS user interface was re-designed in order to ease the participation.

When crossing the segment information provided by GA with our clustering groups, we can clearly see in Table 3 that the users who didn’t manage at all to flow the right flow of actions fall in the *one-day passing* and the *ephemeral* groups and that the majority of the *struggling* and all the *committed* users manage to apply the right flow of actions (at least once)

|  |  |  |
| --- | --- | --- |
| **group** | **Nb. of unsuccessful participants** | **Nb. of successful participants** |
| one-day passing | 174 | 117 |
| ephemeral | 105 | 46 |
| struggling | 6 | 51 |
| committed | 0 | 12 |

Table 3 Distribution according to the different groups, of the number of participants who managed or not to apply the right flow of actions in VAS.

#### Discussion

The set of the 511 identified users includes users who confirmed their signing up via email and the others who did not confirm and hence whose activity was not logged on the server. This second group counts 244 out of the 511 VAS participants that we tracked. Most of these participants never returned to the game, and they naturally fall in the group of *one-day passing* participant. As they are not entirely identified, it is not possible to track their behavior regarding the in-game questionnaires or the other surveys that are collected separately. Regarding the typology analysis presented here, these unconfirmed users are treated similarly to the other confirmed users.

Theoretically, a correlations analysis can reveal relationships between the different indicators. But since most of the participants except the *committed* group have limited activities, some correlations are not available. Also the size of the *committed* group is too small for the correlations to be significant.

Despite the interesting typology that the analytics allows to detect in the behaviour of VAS participants, correlations among the different engagement and learning indicators are not strong enough to deepen the analysis. More participation for longer periods would in the future, allow to test our research hypothesizes regarding the relationship between engagement and learning.

We were confronted to the same limitations when joining analytics and survey data. There are not enough participants who answered both the pre and post tests and who have meaningful tracked activity during the period of analytics collection, statistical analysis is hence limited.

#### Conclusions on the findings from learning analytics in VAS

The CCLTracker framework has been used in VAS to monitor learning analytics indicators. The major outcomes of using learning analytics and analysing learning indicators are:

1. We observed a strong correlation between high engagement and good performance on the game.
2. The learning analytics data shows a high level of activity in the game, especially from the *struggling* group. Eventhough parameter tuning was seemingly difficult, the game contains a number of learning materials allowing users to spend time on the game even if they are not adjusting parameters.
3. Analytics as well as interactions with new users on dissemination activities revealed that VAS user interface needed adaptation to allow users to play properly and progress. This information was used by the developers to adapt the game dynamic and update accordingly the user interface.