# Experiment Doc Template - Patreon Growth Team

# Experiment Name

## (*Asana Link*)

***One-line summary*:  Our goal is to [*improve this KPI*] by [*doing this experiment*].**

# I.  Hypothesis

|  |
| --- |
| *This is a belief about the aspect of our user experience that we want to improve in this experiment.  It is probably something that we believe is holding back success for patrons or creators.*    *Include any 1) supporting qualitative data; 2) supporting quantitative data*    *[e.g. “We believe that creators experience writer’s block when filling out their pages which stops many from launching.”]* |
| **Hypothesis:** |

# II.  Predicted Impact

|  |
| --- |
| *“We predict that if we [idea], then [metric] will [increase by x%] because [justification].”* |
| **Overall Prediction:** |

*We will estimate this experiment’s impact on our metrics in this section. This will help to prioritize this experiment.*

*Assume the variant is already at 100% when predicting impact, and ask yourself what our KPIs would look like if that were true.*

*In most experiments, these will be funnels, with the most recent/final metric in the bottom row, and leading indicators/intermediate metrics above.  Create a separate table for each funnel.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Baseline Value** | **Predicted Value** | **Predicted Increase %** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |
| --- |
| ***Bottom-line impact:***  *[e.g. If this experiment is successful, we predict that we will generate 16 new financially active creators/month].* |
| **Bottom-line impact:** |

*How confident are we in this? Why? (Think of past experience, best practices)*

# III.  Spec/Wireframes

**Sufficient Test**

|  |
| --- |
| *What is the minimal experiment we can do that will validate or invalidate iterating on this experiment?  We want to do the least possible amount of work that is likely to provide a meaningful result.  Then we can iterate and improve on the feature further.  Think small.  Anything else put in the “if this works” section below.* |
|  |

# IV.  Experiment design

**For a detailed explanation of designing and analyzing experiments, read this document:** [**Analytical framework for AB tests at Patreon**](https://paper.dropbox.com/doc/Analytical-framework-for-ab-tests-at-Patreon-YBuvYei9d9BwAHlMEosNW)**.**

**1. Observation Units, Treatment and Response**

|  |
| --- |
| ***Observation unit***  *This answers the question “Who participates in the experiment? Who is randomly assigned?”*    *We identify it with three components:*     1. ***Plain unit:*** *creators, site visitors, patrons, etc.* 2. ***Identifier:*** *`user\_id`, `device\_id`, `mandrill\_message\_id`, etc.* 3. ***Eligibility:*** *what did they have to do to be considered for the experiment.*     *Also make a note of who will specifically be excluded from this experiment, and why.*    ***[Examples:*** *creators identified by user\_id who have created a campaign since xx/xx/xxxx, visitors identified by device\_id that made at least one logged-out request to patreon.com during February.]* |
| **Observation unit**     1. Plain unit:        1. Identifier:        1. Eligibility: |

|  |
| --- |
| ***Treatment groups:***    *A treatment group is the group of observation units that get assigned to a particular treatment.*    *For each treatment (including control), list 1) name, 2) how the observation units’ experience will differ from other treatment groups*    ***[Example****: CPV2: Every time the observation unit loads a creator page, we render CPV2, CPV3: Every time the observation unit loads a creator page we render CPV3]* |
| **Treatments (aka variants):** |
| ***Event(s) on which treatment group is assigned*** |

|  |
| --- |
| ***Response unit:***  *This is a measurement of the behavior of the observation units after they experience one of the treatments.  We expect this measure to be different across treatment groups.*    *The response unit must be at the observation unit level.*    *You can additionally mention what aggregation of the metric will you compute for each treatment group. This is often the mean, but can also be the median, max, etc.*    *We’ll flesh out the funnels, including leading indicators to the response metric, in the Predicted Impact section below.*    *[one response unit primary]* |
| **Response unit:** |

**2.  Timeline**

|  |
| --- |
| **First day of experiment:** |
| **First day when experiment will be ready for analysis:** |
| **Initial roll-out % and subsequent increases\*:** |
| **[date when it becomes significant]** |
| **Expected size of population to be exposed to experiment:** |

*Factors that determine roll-out %:*

* *Risk: If we mess something up, how badly could this turn out?*
* *Data collection: are we getting to significance as fast as we can?*
* *Overlapping experiments - reduce percentage to reduce effect of impact saturation*

# VI.  Analytics Implementation

**For a detailed explanation of implementing experiments on the engineering side, read this document:** [***Setting Up Experiments***](https://paper.dropbox.com/doc/Setting-up-Experiments-fRzjMr0q7Q5JvqRhncccZ#:h2=Setting-up-Experiments)**.**

*What events and properties should we be tracking?  These should directly correspond to the analysis we mapped out - the actions and segments we are interested in - rather than logging everything.*

***Note****:  Whenever possible, the treatment group should be recorded in the relevant event as a property.  E.g. if we’re showing something differently on the creator page, we should show the assigned treatment on `Creator Page : Landed`.*

**1.  Events and properties**

|  |  |
| --- | --- |
| **Flag name:** |  |
| **UTM properties *(if any)*:** |  |

[add or modify - data analyst responsible]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Event name *(including domain)*** | **Properties *(if any)*** | QA: Appears in amplitude | QA: All properties are correctly passed | QA: Appears in raw tables | QA: All properties are correctly passed | QA: Passes naming standards |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**2.  QA**

*Special QA considerations such as user states, devices, additional flows that should not be impacted.  Might not apply to every experiment.*

* *[e.g. Test on Mobile, test for logged-out users, test for existing patrons, etc.]*

# VII.  If This Works We Should

|  |
| --- |
| *“If this works, we should…” These ideas go beyond the sufficient test.* |
|  |

# VIII.  Results

**Pre-Analysis Notes**

*[Scratch pad for thoughts on analysis--ideas to try, stumbling blocks, etc.*

*Could this experiment have adverse effects on a KPI that we were not going to measure directly?  If so, include that metric above to be analyzed before and after.*

*Consider qualitative ideas like surveys, or scroll maps, etc. - what might help shed light on any results?]*

* *Segment by…*
* *Segment by…*
* *Segment by…*

**Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Baseline** | **Observed value** | **Predicted Increase %** | **Observed Increase %, with 95% confidence interval** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

*These KPIs/funnels should be in the same format as the Predicted Results above.  Include a table for each funnel in the experiment.*

# IX.  Learnings & Next Steps

*What did we learn?  Why did we observe what we did?*

*How does this affect how we think about our users? Does it give us new hypotheses to test? Does it give us ideas for new experiments? Should we double down here?*