

Final Project

Experiment Design

Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here. For each metric, explain both why you did or did not use it as an invariant metric and why you did or did not use it as an evaluation metric. Also, state what results you will look for in your evaluation metrics in order to launch the experiment.

Invariant metrics:

of cookies:

I used the number of cookies as an invariant metric, since adding the time commitment popup screen would not affect this. I did not use it as an evaluation metric since knowing the # of unique cookies to view the overview by itself wouldn't provide any useful measurement for evaluating the experiment results.

of clicks:

I used the number of clicks as an invariant metric, since adding the time commitment popup screen would not affect this. I did not use it as an evaluation metric since knowing the # of unique cookies to click the "Start Free Trial" button by itself wouldn't provide any useful measurement for evaluating the experiment results.

Click-through-probability (CTP):

I used the CTP as an invariant metric, since adding the time commitment popup screen would not affect this. I did not use it as an evaluation metric since knowing the CTP by itself wouldn't provide any useful measurement for evaluating the experiment results.

Evaluation metrics:

Gross conversion:

I did not use gross conversion rate as an invariant metric since the time commitment pop-up screen could affect this. I used gross conversion rate as an evaluation metric to assess if the time commitment popup screen impacted the number of people who enroll in the free trial. To launch the experiment, I want to see the gross conversion rate reduced. The reduced gross conversion rate would confirm that the popup is working to turn away people who do not have enough time to commit to the course. This aligns with the experiment goal of "...reducing the number of frustrated students who left the free trial because they didn't have enough time...".

Net conversion:

I did not use net conversion rate as an invariant metric since the time commitment pop-up screen could affect this. I used net conversion rate as an evaluation metric to evaluate whether the time

commitment popup screen had a significant impact to the number of students to continue past the free trial and eventually complete the course. To launch the experiment, I want to confirm that there is not a significant reduction in the net conversion rate. This aligns with the experiment goal of “... without significantly reducing the number of students to continue past the free trial and eventually complete the course.”

In summary, both the gross conversion and net conversion metrics need to behave as expected to launch the experiment.

Not selected:

of user IDs:

I did not use the number of user IDs as an invariant metric since adding the time commitment pop-up screen could affect this metric. I also did not use the number of user IDs as an evaluation metric. Even though the number of user IDs is relevant in this experiment, using the raw number of user IDs as a measurement in this experiment could be misleading if there were more pageviews in the experiment group vs. the control group (or vice versa). Instead, I used Gross Conversion (probability of enrolling, given the student clicked on the “Start Free Trial” button) which will not be impacted by number of pageviews in each group.

Retention:

I did not use retention rate as an invariant metric since the time commitment pop-up screen could affect this. I originally selected retention rate as an evaluation metric, since retention (reducing the # of frustrated students who left the free trial because they didn’t have enough time) was part of what we’re trying to change by adding the time commitment popup and setting clearer expectations up front. I removed this as an evaluation metric once I realized that I needed 4,737,818 page views. It would take 236 days to collect this many page views, which I thought was too long of a time-period.

Measuring Standard Deviation

List the standard deviation of each of your evaluation metrics. For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the empirical variability, or whether you expect them to be different (in which case it might be worth doing an empirical estimate if there is time). Briefly give your reasoning in each case.

Gross Conversion

Analytic estimate of standard deviation: 0.0202

Do you think the analytic estimate would be comparable to the empirical variability? Yes

Reasoning:

The unit of diversion is a **cookie** for this experiment which is explained in the “Experiment Overview” section of the project instructions. The unit of analysis for gross conversion is also a cookie as you can

see from the calculation below. Therefore, the analytical variability and empirical variability should be consistent, hence no need to calculate an empirical variability.

Gross Conversion = enrollments / unique **cookies** to click “Start Free Trial” button

Net Conversion

Analytic estimate of standard deviation: 0.0156

Do you think the analytic estimate would be comparable to the empirical variability? Yes

Reasoning:

The unit of diversion is a **cookie** for this experiment which is explained in the “Experiment Overview” section of the project instructions. The unit of analysis for net conversion is also a cookie as you can see from the calculation below. Therefore, the analytical variability and empirical variability should be consistent, hence no need to calculate an empirical variability.

Net Conversion = payments / unique **cookies** to click “Start Free Trial” button

Sizing

Number of Samples vs. Power

Indicate whether you will use the Bonferroni correction during your analysis phase, and give the number of pageviews you will need to power your experiment appropriately.

Bonferroni Correction

Bonferroni correction will not be used during the analysis phase.

Pageviews required: 683,626

Duration vs. Exposure

Indicate what fraction of traffic you would divert to this experiment and, given this, how many days you would need to run the experiment. Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?

Fraction of Traffic Exposed: 1.0 (100%)

Length of Experiment (Duration): 18 days

Reasoning:

I chose to divert 100% of the traffic to the experiment because I think the experiment poses minimal risk. By diverting 100% of the traffic, we need to run the experiment for 18 days to get enough pageviews which seems like a reasonable length.

How risky?

I think this experiment poses minimal risk to the Udacity students.

1. The risk of a student encountering any physical, psychological, emotional, social, or economic harm from this experiment is minimal.
2. The only new data being collected from the student is the number of hours per week they can set aside to complete the course, which the student should clearly understand from the question asked – “How many hours per week are you committed to learning?”. I would not consider this data to be sensitive. I also wouldn’t expect this data to be stored by Udacity, so there wouldn’t be any risk of the data being made public or tied back to the student.
3. The experiment could potentially benefit the student by setting clear expectations about recommended time commitment up front and thus preventing the frustration in the future when they find out they don’t have enough time to complete the course.
4. The student also has a clear alternative – to access the course materials for free instead of enrolling in the free trial.

Experiment Analysis

Sanity Checks

For each of your invariant metrics, give the 95% confidence interval for the value you expect to observe, the actual observed value, and whether the metric passes your sanity check. For any sanity check that did not pass, explain your best guess as to what went wrong based on the day-by-day data. Do not proceed to the rest of the analysis unless all sanity checks pass.

	Lower Bound	Upper Bound	Observed	Passes
# of cookies	0.4988	0.5012	0.5006	Yes
# of user IDs	0.4959	0.5041	0.5005	Yes
# of clicks	-0.0013	0.0013	-0.0001	Yes

Result Analysis

Effect Size Tests

For each of your evaluation metrics, give a 95% confidence interval around the difference between the experiment and control groups. Indicate whether each metric is statistically and practically significant.

	Lower Bound	Upper Bound	Statistical Significance?	Practical Significance?
Gross Conversion	-0.0291	-0.0120	Yes	Yes
Net Conversion	-0.0116	0.0019	No	No

Sign Tests

For each of your evaluation metrics, do a sign test using the day-by-day data, and report the p-value of the sign test and whether the result is statistically significant.

	p-value	Statistical Significance?
Gross Conversion	0.0026	Yes
Net Conversion	0.6776	No

Summary

State whether you used the Bonferroni correction, and explain why or why not. If there are any discrepancies between the effect size hypothesis tests and the sign tests, describe the discrepancy and why you think it arose.

Bonferroni Correction

Bonferroni correction was not used during the analysis phase, since the expectations for both the gross conversion (we want to see a decrease) **AND** net conversion (we do not want to see a decrease) needed to be met to launch the experiment. Since, **ALL** the conditions needed to be met for us to reject the Null Hypothesis and launch the experiment, we did not need to apply Bonferroni Correction to protect against the additional risk of one of the metrics incorrectly rejecting the Null Hypothesis by chance.

If the scenario was different and we only needed one of the two conditions to be met for us to reject the Null Hypothesis and launch the experiment, then we would need to apply Bonferroni Correction to protect against the additional risk of one of the metrics incorrectly rejecting the Null Hypothesis by chance.

Any discrepancies between the effect size hypothesis tests and the sign tests? None

Recommendation

Make a recommendation and briefly describe your reasoning.

There were two criteria that needed to be met to launch the experiment:

1. Significant reduction in the gross conversion rate.
2. Not a significant reduction in the net conversion rate.

The first condition for Gross Conversion was met. The confidence interval ranged from -0.0291 to -0.0120, indicating that the Gross Conversion went down by at least the practical significance boundary (-0.01). So, we got the result we were looking for in decreased enrollment.

The second condition for Net Conversion was not met. The confidence interval ranges from -0.0016 to 0.0019, indicating that there was no statistically significant change, but the confidence interval does include the negative of the practical significance boundary (-0.0075). So, we cannot guarantee there was not be a significant reduction in Net Conversion (the number of students to continue past the free trial and eventually complete the course).

Given that only the first condition was met and there's a chance that the second condition was not met, I would recommend completing further testing before launching the experiment.

Follow-Up Experiment

If you wanted to reduce the number of frustrated students who cancel early in the course, what experiment would you try? Give a high-level description of the follow up experiment you would run, what your hypothesis would be, what metrics you would want to measure, what your unit of diversion would be, and your reasoning for these choices.

Experiment Overview – Free Trial Skills Confirmation:

The previous experiment was created with the assumption that students were leaving the free trial frustrated because they did not have enough time to commit to complete the course, but what if that is just part of the problem. For example, what if many of the students were eager to complete the course and could commit to 5 or more hours each week, but they lacked the pre-requisite knowledge/skills required to successfully complete the course in a timely manner.

In this follow-up experiment, I will test a change where if the student clicked “Start Free Trial” and they indicate that they have 5 or more hours to commit to the project, they are asked another question about the skills they currently possess. The popup screen will list out several of the most important skills that are usually required for the student to be successful. The student will need to check a box next to each skill they currently possess. If the student does not check all the boxes (indicating they possess all the recommended skills), a message will appear indicating that the Udacity course usually requires the student to possess all the skills, and suggest the student review a list of courses, web sites, articles, etc. prior to enrolling in the course. At this point the student would have the option to continue enrolling in the free trial, or instead can choose to review the other recommended material first.

Hypothesis:

Having the student confirm they possess the required skills before enrolling in the free trial will set clearer expectations for the students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have the pre-requisite skills required to complete the course—without

significantly reducing the number of students to continue past the free trial period and eventually complete the course.

If this hypothesis holds true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

Metrics to measure:

Invariant metrics

I would use the following invariant metrics, since adding the pre-requisite skills question would not affect these.

- # of cookies
- # of clicks
- Click-through-probability (CTP)

Evaluation metrics

Both evaluation metrics listed below would need to behave as expected to launch the experiment.

- Gross conversion
- Net conversion

I want to see the gross conversion rate reduced (both statistically and practically significant). The reduced gross conversion rate would confirm that the popup is working to turn away people who do not have the pre-requisite skills recommended to complete the course successfully. This aligns with the experiment goal of "...reducing the number of frustrated students who left the free trial because they didn't have the pre-requisite skills required to complete the course...".

I want to confirm that there was not a significant change to the net conversion (neither statistically or practically significant). This aligns with the experiment goal of "... without significantly reducing the number of students to continue past the free trial and eventually complete the course."

Note: I did not include Retention as an evaluation metric, since it was determined in the last experiment that it would take too long to gather the amount of pageviews required for the experiment.

Unit of Diversion:

Cookies will be used to divide the students into the experiment and control groups. There could be instances where the same student sees both the experiment and control instances, but it would be rare since the student would need to access the site from a different browser or clear the cache between visits. Also, once a user has enrolled in the free trial, I would expect them to immediately log in upon accessing the Udacity site, so they would not have the option to click the "Start Free Trial" button.

I did not want to use page views since it would be much more likely the same student sees both the experiment and control instances. All the student would need to do is reload the page and then they could be routed to the other group.

Note: For tracking purposes, once the student enrolls in the free trial, they will be tracked by user ID from that point forward. If the user does not enroll in the free trial, their user ID will not be tracked.