# Instructions

<https://docs.google.com/document/u/1/d/1aCquhIqsUApgsxQ8-SQBAigFDcfWVVohLEXcV6jWbdI/pub?embedded=True>

# Change Description

An experiment was conducted in which potential Udacity students were diverted by cookie into two groups, experiment and control. The experiment group was asked to input the amount of time they are willing to devote to study, after clicking a “start free trial button”, whereas the control group was not.

# Experiment Design

## Metric Choice

**Invariant metrics:**

Number of cookies: number of unique cookies to view the course overview page

Number of clicks: number of unique cookies to click the “Start free trial”

Click-through-probability: number of unique cookies to click the “Start free trial” button divided by number of unique cookies to view the course overview page

*Reason:* cookies are used as unit of diversion so it should be invariant metric. The change is not expected to have influence on click through probability.

**Evaluation metrics:**

Gross conversion: number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the “Start free trial” button. (dmin =0.01)

Retention: number of user-ids to remain enrolled past the 14-day boundary divided by number of user-ids to complete checkout (dmin=0.01) Too long to observe

Net conversion: number of user-ids to remain enrolled past the 14-day boundary divided by the number of unique cookies to click the “Start free trial” button. (dmin= 0.0075)

*Reason:* These two metrics show the satisfaction level of students enrolled in the free trial.

We expect the gross conversion to be lower in the experiment group since the number of enrollments should be lowered, and in contrast we expect not to see any significant difference between the values of net conversion rate of the experiment and control groups.

## Measuring Standard Deviation

Gross conversion: 0.0202

Net conversion: 0.0156

For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the 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.

## Sizing

### Number of Samples vs. Power

Number of pageviews: 685325

The sample size is very large, as we will see, when we consider the unique cookie number of overview page visitors. It is possible to reduce the alpha in order to reduce the false positive probability of the whole test anyway.

How to deal with large sample sizes, like retention metric in this case?

We can loosen the power of the test, and increase the alpha for the retention, since this metric is our bottleneck. Or increase the practical significance to 2%, instead of the default 1%.

Increasing the practical significance means that even though the change in the metric is statistically significant and practically significant based on the old criterion, now we deem it as insignificant. A good decision?

It may be worth trying, but [the provided data about the traffic](https://docs.google.com/spreadsheets/d/1Mu5u9GrybDdska-ljPXyBjTpdZIUev_6i7t4LRDfXM8/edit#gid=0) is insufficient for keeping the retention. So, I drop the retention, and continue with the net and gross conversion metrics.

### Duration vs. Exposure

Fraction of traffic exposed: 0.5

Length of experiment: 35

The experiment may cause some unexpected side-effects, so it is better not to expose all traffic to it.

# Experiment Analysis

## Sanity Checks

All pass

Analysis showed the expected equal distribution of cookies into the control and experimental groups at the 95% CI.

## Result Analysis

### Effect Size Tests

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Evaluation metrics | Lower bound | Upper bound | dmin | Statistical significance | Practical significance |
| Gross conversion | -0.0291 | -0.0120 | 0.01 | Yes | Yes |
| Net conversion | -0.0116 | 0.0019 | 0.0075 | No | No |

### A difference in gross conversion was found to be statistically signficant at the 95% CI, and the null hypothesis was rejected. Gross conversion also met the practical signficance threshold. Net Conversion was found to be neither statistically nor practically signficant at the 95% CI.

### Sign Tests

|  |  |  |
| --- | --- | --- |
| Evaluation metrics | p-value | Statistical significance |
| Gross conversion | 0.0026 | Yes |
| Net conversion | 0.6776 | No |

### Summary

I didn’t use Bonferroni correction, as the evaluation metrics are expected to move together. The more students who enrolled, the more students who paid. Bonferroni would be too conservative.

Sign tests are consistent with the effect size tests.

## Recommendation

There was a significant decrease on gross conversion and no significant change on net conversion.

To try to explain the results, I split people who click the “Start free trial” button into four categories:

Control group

|  |  |  |
| --- | --- | --- |
|  | Have time | No time |
| Have money | Click and enroll and pay | Click and enroll |
| No money | Click but not enroll | Click but not enroll |

Experiment group

|  |  |  |
| --- | --- | --- |
|  | Have time | No time |
| Have money | Click and enroll and pay | Not click |
| No money | Click but not enroll | Not click |

According to Simpson’s paradox theory, this situation is likely to happen when the ratios of each category are quite different. For example, let’s assume that the ratio of people having time vs. no time is 1:4. Plus, in the group having time, the ratio of people having money and no money is 1:4, while in the group having no time, the ratio of people having money and no money is 1:1.

Example for Simpson’s paradox

|  |  |  |
| --- | --- | --- |
|  | Have time | No time |
| Have money | 20 | 250 |
| No money | 80 | 250 |

Calculated gross conversion

|  |  |
| --- | --- |
| Control | 0.45 |
| Experiment | 0.2 |

I also think that retention rate is higher in experiment group, due to lower gross conversion and no change in net conversion.

I’ll recommend not launch this change after we have more evidence.

Other opinions:

I would run the experiment. If it is possible, I would run the experiment twice to be sure about the results. However, after running the experiment, we can monitor the results, and since this change is not risky, there is not that much harm if the effect of the experiment is not as significant as we expect.

My recommendation is not to launch. We can design a follow-up experiment at this point.

This experiment was designed to determine whether filtering students as a function of study time commitment would improve the overall student experience without significantly reducing the number of students who continue past the free trial. A statistically and practically significant decrease in Gross Conversion was observed but with not in Net Conversion. This translates to a decrease in enrollment not coupled to an increase in students staying for 14 days boundary to trigger payment.

# Follow-Up Experiment

I want to see if the students in experiment group who were diverted to “access course materials” are more likely to revisit Udacity due to higher level of satisfaction. But this may take a longer time to observe.

Other opinions:

The first opportunity for intervention was explored above wherein a poll regarding time commitment was used as to filter out students likely to become frustrated. This filter focused only on time commitment to the class and did not address other reasons why a student might become frustrated and cancel early. Even if the student was sincere in their response and diligent in their study, they may become frustrated if they don't have the suggested pre-requisite skills and experience. That is, their committed time may not be enough if they don't come in with the pre-requisite skill set. Adding a checklist of pre-requisite skills to the popup regarding time commitment may be informative.  A successful experiment would be one in which there is a significant decrease in Gross Conversion coupled to a significant increase in Net Conversion.