

## Experiment Design

### Metric Choice

Invariant Metrics: Number of cookies, Number of clicks

Evaluation Metrics: Iteration 1: Retention, Net Conversion

Iteration 2: Net Conversion, Gross Conversion (Retention was dropped to reduce the duration)

I chose number of cookies and number of clicks as invariant metrics as these would not change in our experiment. In fact we need these metrics to be comparable in a control group and experimental group to observe the effect of our experiment.

*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.*

Number of cookies: This is an independent invariant metrics as it is not affected by our experiment which only happens after clicking on “Start free trial button”

Number of user-ids: Not a good invariant metric because the number of users who enroll in the free trial is dependent on the experiment. It is not an ideal evaluation metric because it not normalized.

Number of clicks: Again, this is an independent invariant metrics as it is not affected by our experiment which only happens after clicking on “Start free trial button”

Click-through-probability: This is a good invariant metric because the clicks happen before the user sees the experiment, and are thus independent from it. However, the number of cookies and number of clicks are already sufficient for our experiment. Hence, for simplification, I decided to proceed without including it in current evaluation.

Gross conversion: Good evaluation metrics as it helps us measure if our experiment would help reduce the cost of enrollment for students that will eventually drop out. Since it is dependent on our experiment, it is not a good invariant metrics

Retention: Good evaluation metrics as it helps us measure if our experiment would positively affect the financial outcome. Since it is dependent on our experiment, it is not a good invariant metrics

Net conversion: Good evaluation metrics as it helps us measure if our experiment would positively affect the financial outcome. Since it is dependent on our experiment, it is not a good invariant metrics

The experiment is to clearly convey to the audience about the required invested time to complete the course. This would prepare the students beforehand and hence avoid leaving the course due to frustration. Therefore, we are essentially trying to measure if our experiment would increase retention and reduce the cost of enrollment for students that eventually drop out. Metrics Gross Conversion, Retention and Net Conversion measures just that!

Hence, I started with the three metrics. But later, it was found that in order to reduce the duration of the experiment to a practical level, retention should be dropped as an evaluation metrics. Therefore, our new evaluation metrics are: Gross conversion and Net conversion. The first metric measures whether we lower our costs by introducing the screener. The second metric measures if the change affects our revenues.

In order to recommend whether to launch the experiment or not, Gross conversion should have a practically significant decrease, and make sure that there is no drop in the net conversion.

### **Measuring Standard Deviation**

Analytic estimate of standard deviation:

Gross Conversion: 0.0202

Retention: 0.0549

Net Conversion: 0.0156

Is it okay to go ahead with analytic estimate of standard deviation?:

Gross conversion and net conversion both have the number of cookies as their denominator, which is also our unit of diversion. Therefore, it is okay to proceed with just the analytical estimate of standard deviation.

## Sizing

### Number of Samples vs. Power

I would not use Bonferroni correction because my evaluation metrics are covariants.  
The number of pageviews you will need to power you experiment appropriately is  
4741212

Steps: Using  $\alpha=0.05$ ,  $\beta=2$  and  $d_{min}$  (given for each evaluation metrics), calculate the number of pageviews for each metric individually. The largest value is the number of pageviews required. However, since we need it for both control group and experimental group we multiply this number by 2.

### Duration vs. Exposure

I diverted the whole traffic, fraction 1 and after second iteration I got duration=18 days

Using the total page views as 4741212, the total duration was 119 days which is very long for an experiment! This made me revisit my earlier decisions. I realized that the main reason for a large duration was including retention. Therefore I dropped that as my evaluation metrics and selected gross conversion. The largest total page = 685326. This divided by total page views per day gave 18 days long duration.

*Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?*

The experiment is not risky as it requires a simple change to the existing system. It would not affect the enrolled students in any way and will make the new students better aware of what the course demands. The duration of the experiment is not very long-only 18 days and it does not involve any sensitive data that may hurt any sentiments. Therefore, overall I think there aren't any major risks and hence it is safe to divert the whole traffic.

## Experiment Analysis

### Sanity Checks

95% confidence interval for:

Number of cookies: Lower bound: 0.4988 Upper bound: 0.5011 Observed: 0.5006

Number of clicks: Lower bound: 0.4959 Upper bound: 0.5041 Observed: 0.5005

All the metrics pass sanity check!

### Result Analysis

## Effect Size Tests

Gross Conversion: Lower bound: -0.0291, Upper bound: -0.012

Statistically and practically significant!

Net Conversion: Lower bound: -0.0116, Upper bound: 0.0018

Statistically and practically not significant!

## Sign Tests

Gross Conversion: sign test p value= 0.0026, statistically significant

Net Conversion: sign test p value: 0.6776, not statistically significant

## Summary

In our case we need all our evaluation metrics (that is both gross conversion and net conversion) to reject the null. It might be useful to apply the Bonferroni correction if we were to launch the experiment when any metric would reject the null. If we were looking at only one metric in order to launch the experiment Bonferroni correction could have been used to reduce the risk that one metric is deemed significant by mistake. That is not true in our case.

There is no discrepancy between effect size hypothesis and sign tests. Both the effective size hypothesis tests and sign tests state that the change will practically significantly reduce the gross conversion, but the results for the net conversion rate is not in our favor.

## Recommendation

Our Gross conversion is negative and practically significant. This is a good outcome because we lower our costs by discouraging trial signups that are unlikely to convert. However, net conversion is statistically and practically insignificant and the confidence interval includes negative numbers. This implies that there is a risk that net conversion went down by a number that would incur losses to the business. Therefore, there is a risk that our experiment may lead to a decrease in net conversion which does not sit well with our condition for launching the experiment.

Hence, at this point, I wouldn't recommend launching the experiment. We need to dig deeper!

## Follow-Up Experiment

In a follow up experiment, I would like to make a change in the message that appears when students click the "start free trial" button. The message would say, "Remember, if

you finish the first part of the course within 2 weeks you earn 20% off on the tuition at the completion of the course”.

The hypothesis is that this message would entice people more and increase retention.

The unit of diversion is a cookie, although if the student enrolls in the free trial, they are tracked by user-id from that point forward. The same user-id cannot enroll in the free trial twice. For users that do not enroll, their user-id is not tracked in the experiment, even if they were signed in when they visited the course overview page.

Since I am concerned with the retention, my invariant metrics and evaluation metrics would be the same as the current experiment. However, in addition to the statistical significance, I would also consider the business impact of the change. The retention (change) should be more significant in this experiment to make sure we are making financial gains through this change.