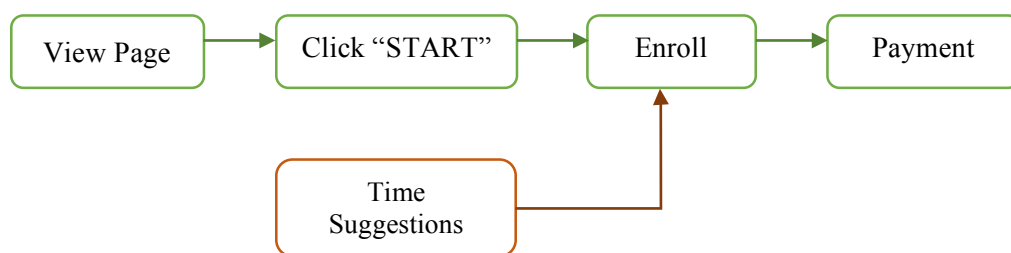


## Summary

A/B testing is a methodology used online when you want to test a feature or new product's performance. Udacity plans to improve the entire enrolled student experience and coaches' capacity to support students who are likely to complete the course by adding a time-spent-on-course feature (Free Trial Screener). This project aims at analyzing two versions of Udacity's website (experience group and control group), and determine whether the new feature should be launched in order to reduce the number of students who left the free trial because of the time issue. This project includes metrics choice, measuring variability, sizing, sanity checks, analysis, recommendation, and the follow-up experiment. Data used in this project was given by Udacity. The progress for the two sets looks like this:



## Experiment Design

**Null Hypothesis:** This approach might not significantly decrease the number of students who left the free trial because they don't have enough time.

**Alternative Hypothesis:** This approach will obviously reduce the free trial cancellation rate.

## Hypothesis (From the instruction)

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)

## Metric Choice

### Invariant Metrics:

**Number of Cookies:** The number of unique cookies to view the course overview page.

**Number of Clicks:** The number of unique cookies to click the "Start Free Trial" button (which before the free trial screener is triggered).

**Click-Through-Probability:** The number of unique cookies to click the "Start the trial" button divided by the number of unique cookies to view the course overview page.

These three metrics won't be affected by the screener approach because click the "Start" button has happened before the screener notice so that the users won't be noticed differently yet. Also, cookies were assigned as the division of measure which means it

should be randomly and equally split into both experience group and control group. Therefore, these are good invariant metrics but not evaluation metrics.

### Evaluation Metrics:

**Gross Conversion:** The number of user-ids to complete checkout and enroll in the free trial divided by the number of unique cookies to click the “Start” button.

**Net Conversion:** The number of user-ids to remain enrolled past the 14 days boundary divided by the number of unique cookies to click the “Start” button.

**Retention:** The number of user-ids to remain enrolled past the 14-day boundary divided by the number of user-ids to complete checkout. The retention is expected higher in experiment than the control group.

Our goal is to reduce the number of users who couldn't continue the course because of the time issue and improve the learning experience at the same time. Gross conversion is a good metric to track the number of users enrolled in the free trial and Net conversion metric could help us to check the number of users who still remain in the course. We can use the combination of Gross conversion and Net conversion metrics as evaluation metrics. The Gross conversion is expected to decrease and Net conversion is expected to increase or keep the same in the experiment group.

### Unused Metrics:

**Number of User-id:** The number of users who enroll in the free trial. This assumes that the fewer users enroll, the more users would complete the course because their education experiment improved and thus users are not equally distributed between the control and experiment groups. So, this could be a good metric but not an invariant metric. Since Gross conversion is more robust than number of user-id, we choose Gross conversion.

### Measuring Standard Deviation

For binomial distribution, the analytical standard deviation is:  $\sqrt{\frac{p(1-p)}{N}}$

Given 5000 cookies to view the course overview page per day:

**Gross Conversion:**

$$P_{gc} = \frac{660}{3200} = 0.20625$$

$$SD_{gc} = \sqrt{0.20625 * (1 - 0.20625) / (5000 * 0.08)} = 0.0202306$$

**Retention:**

$$P_r = 0.53$$

$$SD_r = \sqrt{0.53 * (1 - 0.53) / (5000 * (\frac{660}{40000}))} = 0.054949$$

**Net Conversion:**

$$P_{nc} = 0.1093125$$

$$SD_{nc} = \sqrt{0.1093125 * (1 - 0.1093125) / (5000 * 0.08)} = 0.01560154$$

## Sizing

### Number of Samples vs. Power

I didn't use Bonferroni correction, because these metrics are more related and more likely to move at the same time which means it's too conservative.

Given alpha = 0.05, beta = 0.02. To calculate how many page views we need, we can use this [link](#).

Metrics	$\hat{p}$	$d_{min}$	Sample Size	Pageviews
Gross conversion	0.20625	0.01	25835	645875
Retention	0.53	0.01	39115	4741212
Net conversion	.1093125	0.0075	27413	685325

Pageviews required is the maximum of pageviews of the selected evaluation metrics. Therefore, the pageviews we need is 4741212.

### Duration vs. Exposure

If we divert 100% users, we need 119 days for testing retention metric but 17 days for gross conversion and 18 days for net conversion. 119 Days is a really long time to run. Therefore, we only need 18 days to run the test with 100% diversion and 34 days for 50% diversion if we drop the retention metric.

In general, this screener may cause a decrease of the users who plan to enroll. Considering the fact that we do not want to expose 100% of the traffic to the experiment, I would take 75% diversion for 23 days.

## Experiment Analysis

### Sanity Checks

For the three invariant metrics that we choose at the beginning, given the 95% confidence interval for the value we expect to observe, we expect them equal diversion into the experiment and control group. According to the result, they all pass the sanity checks ([find here](#)).

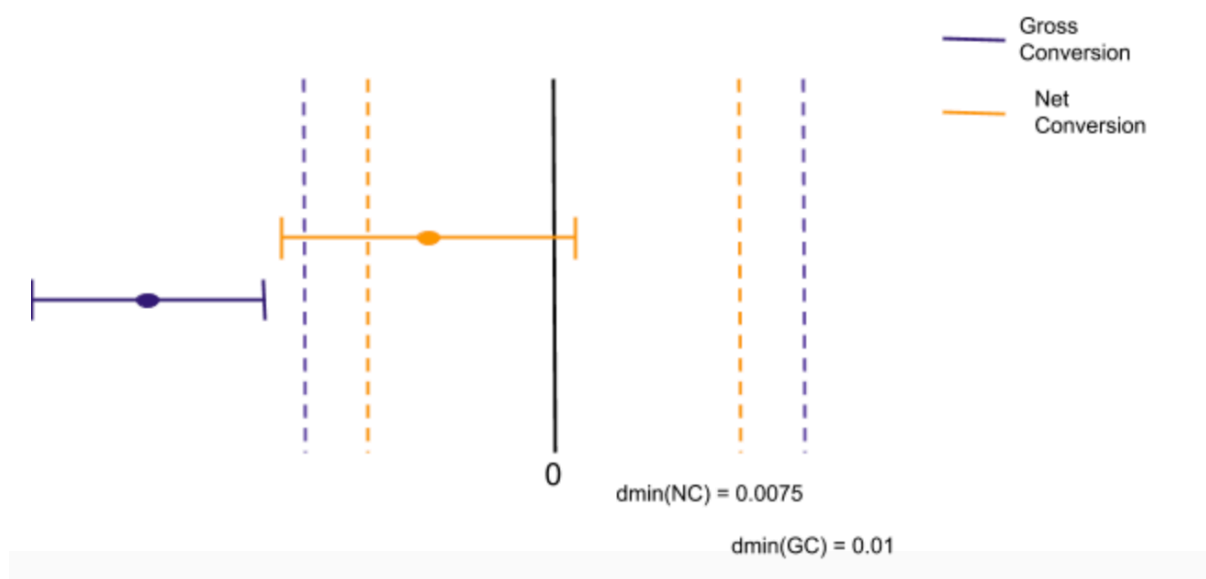
Invariant Metrics	Lower Bound	Upper Bound	Observed	Pass
# of cookies	0.498820	0.502280	0.500640	Yes
# of clicks	0.495885	0.504115	0.500467	Yes
Click-Through-Probability	-0.001296	0.001296	0.000057	Yes

## 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.

Evaluation metric	Lower Bound	Upper Bound	$d_0$	Statistical significance	Practical significance
Gross Conversion	-0.029123	-0.011987	-0.020555	Yes	Yes
Net Conversion	-0.011605	0.001857	-0.004874	No	No



According to our interval [result](#), Gross conversion is confidently significant change both statistically and practically. However, Net conversion is neither statistically nor statistically and practically significant.

### Sign Tests

Evaluation metric	# of days w/ positive change/total days	P value	significance
Gross Conversion	4/23	0.0026	Yes
Net Conversion	10/23	0.67764	No

We got the same result as Effect size test had above.

## Summary

In the experiment, Udacity tested a change by dividing potential students into two group- experiment and control group. Students in the experiment group are asked how much time they had available to devote course if they want to start a free trial and students in the

control group are continually enroll without any suggestion as usual. We use three invariant metrics for sanity check and two evaluation metrics for size and sign test check. Our expectation is to reduce the number of students who do not have enough time spending on Udacity study in order to increase the other student's education experience which means we want a decrease in gross conversion and an increase in net conversion. We didn't use Bonferroni correction because it is too conservative to have a reasonable alpha. In this case, the risk type I error (false positive) and type II errors (false negative) increase as the number of metrics increases, and change the final decision.

Net conversion both fail the significant of effect size test and sign test whereas Gross conversion is significantly impacted by the new screener.

We didn't consider retention as our evaluation metrics because we may not get enough sample page views in a limited time.

## Recommendation

We expect to decrease the number of users who enroll in the free trial without significant reduction of the number of users who stay enrolled after 14-day. According to our data, there is a statistically and practically significant decrease in Gross conversion which matches our expectation. However, Net conversion indicates that the remaining users are no significant differences or even less than the control group. Considering this, my advice is not to launch this experiment.

## Follow-Up Experiment

In my opinion, I don't think Net conversion is a perfect metric to measure the overall student experience because

1. The fewer students enroll the free trial means less potential students may stay in course after 14-day.
2. This experience is to improve the overall students' experience, not just the new-enroll students'. Therefore, we may add some metrics that can test some metrics that include the whole group's experience.

There are two **evaluation metrics** I may consider using:

Course Completion Conversion: That is, number of user-ids to complete the course divided by number of user-ids to enroll in the course.

Course Drop Conversion: That is, number of user-ids who left the course divided by number of user-ids to enroll in the course.

A variety of methods that could improve the overall student learning experience by reducing students who don't have enough time. An ideal approach is that we can test a change where after student finish the 14-day free trial and summarize how much time they spent on devoting to the course per day on average and notice them this course usually require how many hours per day. If they spend less than the minimum suggestion, they will

be asked if they still want to stay, otherwise, they will be encouraged to continue completing the course.

The **hypothesis** is that student all have the opportunity to try the free trial course to learn if they are interested in this course and want to spend enough time to finish the course by reducing the number of students who are not have or willing to have more time on this course. If they already spend reasonable time on the course, Udacity will boost their enthusiasm and provide improved coach experience.

**Unit of Diversion:** user-id

**Invariant Metrics:** user-id

**Evaluation Metrics:**

Retention Conversion: the number of user-ids to remain enrolled past the 14-day boundary divided by the number of user-ids to complete the checkout and enroll the free trial.

Second Payment Conversion: the number of user-ids to complete the second checkout divided by the number of user-ids to the number of user-ids to remain enrolled past the 14-day boundary before the time-spent-summary screener is triggered.