Experiment Design

Experiment Overview: Free Trial Screener

At the time of this experiment, Udacity courses currently have two options on the home page: "start free trial", and "access course materials". If the student clicks "start free trial", they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid version of the course. After 14 days, they will automatically be charged unless they cancel first. If the student clicks "access course materials", they will be able to view the videos and take the quizzes for free, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.

In the experiment, Udacity tested a change where if the student clicked "start free trial", they were asked how much time they had available to devote to the course. If the student indicated 5 or more hours per week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours per week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free. At this point, the student would have the option to continue enrolling in the free trial, or access the course materials for free instead.[This screenshot](#) shows what the experiment looks like.

The hypothesis was that this might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have enough time—without significantly reducing the number of students to continue past the free trial and eventually complete the course. If this hypothesis held true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

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.

Metric Choice

Invariant metrics:

 -Number of cookies: This metric is not affected by the experiment and should not differ between control and experiment group (this page view data is collected prior to a user seeing the free trial screener).

 -Number of clicks: This metric is not affected by the experiment and should not differ between control and experiment group (this click data is collected prior to a user seeing the free trial screener).

 -Click-through-probability: This metric is not affected by the experiment and should not differ between control and experiment group because this is based on the prior two metrics that do not change.

 -Why I didn't use the other metrics:

In the customer funnel, some of clicks will be converted to enrollment, of which some will then be converted to payments. This experiment is trying to lower enrollment without affecting payment numbers (leading to higher retention).

Number of user-ids: This number will change as the experiment will affect number of students who enroll in a course.

Gross conversion: This experiment is essentially trying to lower this number and therefore will not be a good invariant metric. Enrollment numbers are likely to change.

Retention: This experiment could potentially affect both payment and enrollment numbers, as the hypothesis expects the change to increase this number.

Net conversion: While we don't want this number to go down, this experiment could potentially affect payment numbers (by turning away people that we don't intend to turn away), leading to a decreased net conversion rate.

Evaluation metrics:
-Gross conversion: The hypothesis expects this number to go down as enrollment will go down.
-Net conversion: The hypothesis expects this number to stay the same. If this number goes down, it signals that there is a problem with our hypothesis.
-Why I didn't use the other metrics: The invariant metrics (Number of cookies, Number of clicks, Click-through-probability) will not work as evaluation metrics as they will not change as a result of the experiment. Retention would be a good evaluation metric, but I decided not to use it as it would take too long to run the experiment to get enough samples.

Measuring Standard Deviation
Given a sample size of 5,000 cookies viewing the course overview page
Evaluation metrics:
-Gross conversion: 0.0202
-Net conversion: 0.0156

I expect these analytic estimates to be comparable to the empirical variability because both of the metrics are based on unique cookies which is the same as the unit of diversion.

Sizing
Number of Samples vs. Power
I did not use Bonferroni correction for my analysis and given my selected evaluation metrics (Gross conversion, Net conversion), I will need 683,625 page views (for alpha = 0.05, beta 0.2).

Duration vs. Exposure
I decided to use all of traffic to divert to this experiment (1 on a scale of 0 to 1), and given this, it will take 18 days to run the experiment.

	Page views	Days
Gross conversion	645,300.0	17

Net conversion	683,625.0	18

The experiment could potentially turn away students who would have completed the free trial and finished the course successfully. However, because the number of pageviews required include both control and the experiment group, if I divert all the traffic to this experiment, only a half of the traffic will fall under the experiment group while the other half will continue to be part of the control group (no change). If I lower the traffic ratio, the experiment will take longer and may potentially suffer from other issues that aren't controlled for.

Experiment Analysis

Sanity Checks

95% Confidence Interval

Invariant metrics	CI-Lower	CL-Upper	Observed	Sanity Check
# of Cookies	0.4988	0.5012	0.5006	Pass
# of Clicks	0.4959	0.5041	0.5005	Pass
Click-through-probability	0.0812	0.0830	0.0822	Pass

Result Analysis

Effect Size Tests

95% Confidence Interval without Bonferroni Correction

	CI-Lower	CL-Upper	Statistical Significance	Practical Significance
Gross conversion	-0.0291	-0.0120	Yes	Yes
Net conversion	-0.0116	0.0019	No	No

*I only used days that included enrollment and payment data (23 days total).

Sign Tests

Sign test using day-by-day data without Bonferroni Correction

	Gross conversion	**Net conversion**
Total Days	23	23
Days Cont>Exp	19	13
Days Cont<Exp	4	10
Sign Test P-value (two-tailed)	0.0026	0.6776
Significant?	Yes	No

Summary

I did not use Bonferroni correction because for the experiment to be successful, both of the metrics need to be significant (each metric supports different hypothesis-1) reduce students from leaving free trial who don't have enough time 2) without reducing students who complete the free trial and the course). There is no discrepancy between the effect size and the sign test, as they both indicate that gross conversion is significant and net conversion is not significant.

Recommendation

I would recommend an additional experiment until we can have statistical and practical significance on both metrics for two hypothesis. Gross conversion shows a statistical and practical significance with the negative minimum change (1%). Net conversion does not seem to have changed with statistical and practical significance, but the confidence interval includes the negative practical significance boundary.

Follow-Up Experiment

To encourage students to complete the course, preferably as early as they can, I would run an experiment that would motivate their competitive side by showing them other students' average time to finish vs. their progress for each interval or milestone in the course. To run this experiment, I would have a button 'check my progress' for only the control group and when they click on it, it will show their performance vs. others' performance. The hope is that those who click on the button will have a better chance of completing the course. The unit of diversion would be user id's because each user will be enrolled and logged in to see this button. Evaluation metric would be probability of completion, given enrollment. if duration of course completion is too long, we can use the probability of completing certain interval or milestone, given enrollment instead, as an evaluation metric.