Experiment Overview: Free Trial Screener

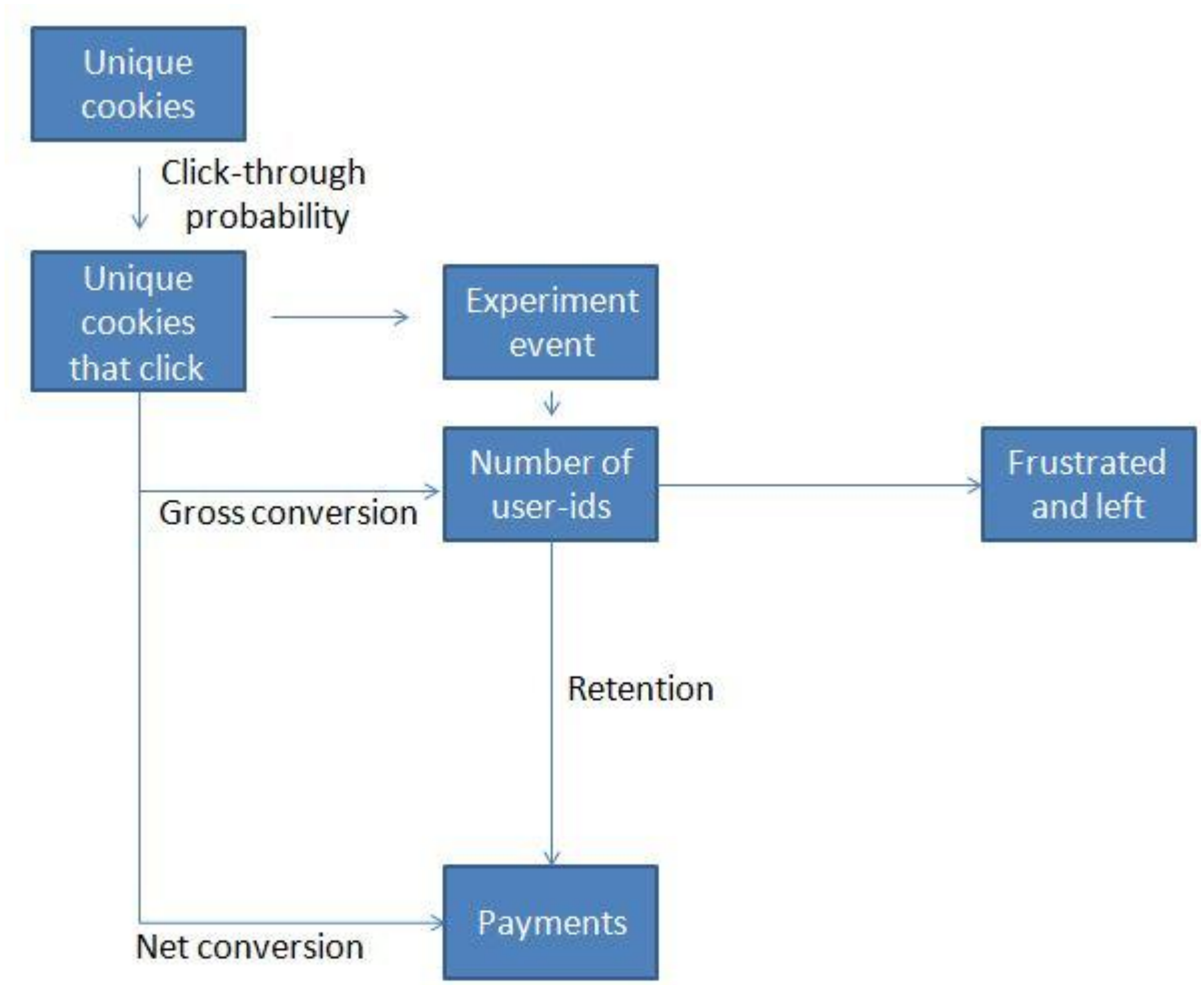
At the time of this experiment, Udacity courses currently have two options on the home page: "start free trial", and "access course materials". If the student clicks "start free trial", they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid 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.

Experiment Design



Invariant metrics: These metrics are used to measure the unchanged part of the experiment, so that later on, we could conduct sanity check. In this case, any metrics that are recorded or calculated by data recorded before the experimental event happened could be considered invariant, which is:

- Number of cookies.
- Number of clicks.
- Click-through-probability.

Evaluation metrics: These metrics are used to measure the parts that are supposed to change in the experiment. I left out the number of user-ids, because there's another metric, gross conversion, which incorporated the information that the number of user-ids covered, and it also play nicer at scale (run from 0 to 1):

- Gross conversion: Measure how the experiment event affect the number of enrolls.
- Retention: Measure the number of students to continue past the free trials (based on enrolls)

- Net conversion: Measure the number of students to continue past the free trials (based on cookies)

	Cookies	Clicks	Enrolls	Payments
Current	200	100	50	25
Expectation	200 (No change)	100 (No change)	40 (Significantly Decrease)	25 (No significant reduction)

The idea is that, the message will reduce the number of people who do not have enough commitment to the course, showing a significant decrease in the number of enrolls. These enrolls, however, are expected to be of better quality, increasing the number of students to continue past the free trial.

- Gross conversion (Enrolls / Cookies): Significantly lower.
- Retention (Payments / Enrolls): Significant increase
- Net conversion (Payments / Cookies): No significant reduction.

Measuring Standard Deviation

In this experiment, I will assume normal distribution and use an analytic calculation of metric variability, based on:

- A large number of datapoints: Over 28,000 clicks in each of the control and experimental groups.
- Well defined distributions: each of the data points measured generates an outcome of either a success (enrollment or payment, depending on metric) or failure

Standard Deviation:

$$\sigma = \sqrt{\frac{\frac{X}{N} * (1 - \frac{X}{N})}{N}}$$

- Gross Conversion: 0.0202
- Retention: 0.0549
- Net Conversion: 0.0156

Gross conversion and net conversion have the same unit of analysis and unit of diversion, cookies, therefore the analytical estimation of the standard deviation is likely to match with the empirical estimation.
Meanwhile, retention has user-ids as the unit of analysis and cookies as the unit of diversion, so retention's analytical estimation of the standard deviation shouldn't match with the empirical estimation.

Sizing

Number of Samples vs. Power

I will not use Bonferroni correction, because this experiment have a low number of evaluation metrics (gross conversion and net conversion).
At $\alpha = 0.05$, $\beta = 0.20$:

Metrics	Standard Deviation	Practical Boundary	Sample Size	Pageviews
Gross Conversion	0.0202	0.01	25830	645750
Net Conversion	0.0156	0.0075	27411	685275
Retention	0.0549	0.01	39078	4737818

The largest sample size is the retention rate at 4,737,818 pageviews. It's going to take too much time and resources to collect the needed pageviews. So I'm going to exclude retention out of my experiment, I still have net conversion, which cover payments, so there shouldn't be any loss of information.

Duration vs. Exposure

Faction of Traffics	Faction of Traffics exposed to the experiment	Traffics (Total)	Days
15%	7.5%	6000	115
25%	12.5%	10000	69
50%	25%	20000	35
100%	50%	40000	18

I'm inclined to put 100% of traffics to the test, because:

- The experiment is not risky:
    - At most, only 50% of the traffics will be exposed to the changes.
    - No one could get hurt from seeing a pop-up message asking how many hours they are going to commit to education.
    - The data is not sensitive, and are under the form of aggregation, which is untraceable.

- Shorter duration allows for more experiments, which bring tremendous profit to the organization.

Experiment Analysis

Sanity Checks

	Number of cookies	Number of clicks on "start free trial"	Click-through-probability on "Start free trial"
Control	345543	28378	0.082126
Experiment	344660	28325	0.082182
$\hat{p}$	0.5	0.5	
Standard Error	0.0006	0.0021	0.000467
Margin of Error	0.00118	0.004116	0.000915
Lower	0.4988	0.4959	0.08121
Upper	0.5012	0.5041	0.083041
Observed	0.49936	0.5005	0.082182
	Pass	Pass	Pass

Result Analysis

Effect Size Tests

	Gross Conversion	Net Conversion
$\hat{p}$	0.208607	0.115127
$\hat{d}$	-0.02055	-0.00487
Standard Error	0.004372	0.003434
Margin of Error	0.008568	0.006731
Lower	-0.02912	-0.0116
Upper	-0.01199	0.001857
d_{min}	(-) 0.01	(-) 0.075
	Statistical Significance	Not Statistical Significance
	Practical Significance	Not Practical Significance

Sign Tests

	Gross Conversion	Net Conversion
Success	4	10
Failure	19	13
p-value	0.0026	0.6776
	Statistical Significance	Not Statistical Significance

Summary

- The launch decision is based on the matching of expectation of all evaluation metrics. Meanwhile, Bonferroni correction is to be used to prevent type I error in the case that any matching metric is enough to launch the experiment; therefore it is not appropriate to use it in this case.
- There wasn't any discrepancy between the effect size hypothesis tests and the sign tests.

Recommendation

$$Gross\ conversion = \ Net\ conversion + \frac{Frustrated\ students}{Clicks}$$

There was a significant reduction in the gross conversion, leading to a significantly lower number of frustrated students who left.

Net conversion represents the number of students to continue past the free trial. In this case, there was no significant reduction to it, but the confidence interval does include the negative of the practical significance boundary. That is, it's possible that this number went down by an amount that would matter to the business.

Therefore, I would recommend to not launch the changes.

Follow-Up Experiment: How to Reduce Early Cancellations

Hypothesis: Students who received a warm welcome will have a higher motivation, increasing the rate of completion.
In this experiment, after completing the enrollment process, students will be added to a real time chat group, where they could see and interact with their peers.
The hypothesis is that, this social interaction will help motivate the students, thus reducing the number of early cancellations – without significantly reducing the number of cancellations.

Now the students have already enrolled, so it will be easier to track them using user id. Therefore the unit of diversion is the user id.

Evaluation metrics:

- Average hours since enrollment.
- Retention.

I use the hours since enrollments to measure the early cancellations. A successful experiment will have a significantly larger average hours since enrollments. It could positively affect the retention rate, or it could not have any affect at all, but the bottom line is that, there should not be any negative change to the retention rate.