

Experiment Design

Metric Choice

Invariant metrics: Number of cookies, Number of clicks, Click-through-probability

Evaluation metrics: Gross conversion, Retention, Net conversion

Number of cookies: the unique cookies to view the course overview page should not be affected by the feature change, as the action happens before the free trial screener trigger. If the experiment and control has high variability, it means something might go wrong—the system or the data capture method.

Number of user-ids: signing up happens after the experiment, so it could not be used as an invariant metric. It could not serve as evaluation metric because it has not been normalized. The difference of number of user-ids could be affected by the number of cookies in two different groups, which means it is not a very representative metric of the experiment.

Number of clicks: this should also be invariant, with the same reason as number of cookies.

Click-through-probability: given the result is actually computed using the first two invariant metrics, it should also be invariant.

Gross conversion: users' enrollments are expected to be affected due to the screener. The metric is expected to witness decline, as some users might consider they do not have enough time and choose not to enroll. However, the decline should not be too much otherwise it would affect revenue and harm the business.

Retention: the metric kind of tell directly if the hypothesis is true—whether the pop-up question will reduce rate of enrolled students who quit the free trial. Though the quit might due to other reasons than without enough time. **But this metric is a little complicated here. Will elaborate further in the later part of the report.**

Net conversion: we hope the experiment group will have an increase in this metric. It is also acceptable that the net conversion remain the same as control group, however, it should not be lower than the control group.

Measuring Standard Deviation

Gross conversion: 0.0202

Retention: 0.0549

Net conversion: 0.0156

Gross conversion: would be comparable to the empirical variability, as the unit of analysis is the same as unit of diversion, which is a cookie

Retention: would be worth doing an empirical estimate if there is time, as the unit of analysis is user-id rather than cookie.

Net conversion: would be comparable to the empirical variability, as the unit of analysis is the same as unit of diversion, which is a cookie

Sizing

Number of Samples vs. Power

Will not use Bonferroni correction, as the evaluation metrics are highly related. Using the correction would be over conservative.

Number of pageviews: 685325.

NOTE:

Given the original evaluation metrics I chose, the largest number of pageviews required should be 4741212 or 119 days, which is based on retention. However,

1) the sample size would be too big and require too much time to run the experiment,

2) the standard deviation of retention might be underestimate, as the unit of analysis is not the same as that of the unit of diversion,

I decided to choose the second largest number of pageviews for the sample size, which would be 685325, based on Net conversion metric. Therefore retention will not be able to serve as an evaluation metric in the rest of the experiment.

Duration vs. Exposure

Fraction of traffic exposed: 1,

Length of experiment: 18 days

Risks:

Adding a pop-up question is not a big change and not very risky for Udacity, I decided to the largest fraction possible to make the experiment run faster.

Number of cookies, number of clicks, click-through-probability use cookies, so the data is not very sensitive, as it's not traceable.

But Gross conversion, Retention and Net conversion need information of user-id, which can trace back to the individual users. The dataset therefore is more sensitive and should be more careful to deal with.

Experiment Analysis

Sanity Checks

	Lower Bound	Upper Bound	Observed	Passes
Number of cookies	0.4988	0.5012	0.5006	Yes
Number of clicks	0.4959	0.5041	0.5005	Yes
Click--through-probability	-0.0013	0.0013	0.0001	Yes

Result Analysis

Effect Size Tests

	Lower Bound	Upper Bound	Statistical significance	Practical significance
Gross conversion	-0.0291	-0.0120	Yes	Yes
Net conversion	-0.0116	0.0019		

Sign Tests

	p-value	Statistical significance
Gross conversion	0.0026	Yes
Net conversion	0.6776	

Summary

I did not use Bonferroni correction through the experiment given the evaluation metrics I chose are highly correlated. Using Bonferroni correction here would be over conservative.

The only discrepancy I observed is the statistical significance of retention, which probably because the sample size is not big enough for this metric. (That's why this evaluation metric mainly serve as a reference in the analysis.)

Recommendation

Do not launch the free trial screener.

Gross conversion has both statistical and practical significance. However both the lower and higher bound is below 0, which means this the number of users enrolled will definitely decline after the screener;

Net conversion did not show statistical or practical significance in this experiment. But given the confidence interval includes negative areas

(the lower bound is lower than 0), it means the net conversion could decline after the experiment. This is not what we expected to see.

Follow-Up Experiment

Providing more support options through the course will help to reduce the number of frustrated students who cancel early in the course.

Experiment:

In the experiment, if the student spent more than 10 minutes on a quiz (or average time of the quiz given past record), ask if the students want to ask for help with three given options and links.

The choices will be : 1) links to udacity forum of that section, 2) links to google/ stackoverflow, prompting suggested keywords; 3) links to reserve coaching session; 4) or just cancel and prefer to think by the user him/herself.

Hypothesis:

The hypothesis is that with the option to look for extra support will help decline users reduce the number of frustrated users who left the free trail because they face challenges.

Unit of diversion:

user-ids

Invariant Metric:

Number of user-ids that enter the quiz.

Probability of users who spent more than 10 minutes on the quiz:
$$\frac{\text{number of user-ids that spent more than 10 minutes on the quiz}}{\text{number of user-ids that enter the quiz}}$$

Evaluation Metric:

Retention: number of user-ids to remain enrolled past the 14day boundary (and thus make at least one payment) divided by number of user-ids to enroll the free trail.

Note:

Received advices and instructions from Union A programs tutor.