Design an A/B Test by Xinlin Feng

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 cookies load before the user get the web page, so it is a invariant metric.
Number of clicks: The clicks Happen before the user get the web page, so it is a invariant metric.
Click-through-probability: The same as Number of clicks.

Gross conversion: It is a good evaluation metric since the result is influenced by the suggestion, say if a user have a less than 5 hours/week study, then he or she might to select not to enroll, this will change the number of user ids. it's directly dependent on the effect of the experiment and also it shows the probability to succeed.
Retention: It is a good evaluation metric since the result is influenced by the suggestion, say if a user have a less than 5 hours/week study, then he or she might to select not to pay after free trial. it's directly dependent on the effect of the experiment and also it shows the probability to succeed. However, I won't launch a experiment on Retention because it take a very long period to observe.
Net conversion: It is a good evaluation metric since the result is influenced by the suggestion, say if a user have a less than 5 hours/week study, then he or she might to select not to pay after free trial. it's directly dependent on the effect of the experiment and also it shows the probability to succeed.

Number of User-ids: Neither of above, because the number of users enrolled in the free trial is dependent on the experiment, I expect to see different value in the control and experiment group, therefore it cannot be a good invariant metric. Meanwhile, User-ids are tracked only after enrolling in the free trial, The number of enrolled users can fluctuate a lot with respect to number of "start free trial" clicks on a particular day, which can influence our examination on the effect of "start free trial" page rendering. A more appropriate way would be number of user-ids divided by number of "start free trial" clicks.

State what results you will look for in your evaluation metrics in order to launch the experiment?
I will looking for a statistic significant decrease in gross conversion while the net conversion will not decrease statistically significant.

Measuring Standard Deviation

Metric	value
Unique cookies to view page per day	40000
Unique cookies to click "Start free trial" per day	3200

Enrollments per day	660
Click-through-probability on "Start free trial"	0.08
Probability of enrolling, given click	0.20625
Probability of payment, given enroll	0.53
Probability of payment, given click	0.1093125

The equation for calculation is

$$\frac{\sqrt{p(1-P)}}{\#Pageview \times rate}$$

Then at 5000 pageviews:

Evaluation Metric	SD
Gross Conversion	0.0202
Retention	0.0549
Net Conversion	0.0156

The analytical estimate of standard deviation tends to be near the empirically determined standard deviation for those cases in which the unit of diversion is equal to the unit of analysis. This is the case for Gross Conversion and Net Conversion, but not Retention. If we do ultimately decide to use Retention, then we should calculate the empirical variability.

Sizing

Number of Samples vs. Power

I did not use Bonferroni Correction during my analysis phase. The metrics in the test has high correlation and the Bonferroni correction will be too conservative to it.
I Used http://www.evanmiller.org/ab-testing/sample-size.html to calculate the pageview.

Gross Conversion:

- Baseline Conversion: 20.625%
- Minimum Detectable Effect: 1%
- alpha: 5%
- beta: 20%
- 1 - beta: 80%
- sample size = 25835 enrollments/group
- Number of groups = 2
- total sample size = 51670
- clicks/pageview: 3200/40000 = .08
- pageviews = 51670/.08 = 645875

Retention

- Baseline Conversion: 53%
- Minimum Detectable Effect: 1%
- alpha: 5%

- beta: 20%
- 1 - beta: 80%
- sample size = 39155 enrollments/group
- Number of groups = 2
- total sample size = 78230
- enrollments/pageview: 660/40000 = .0165
- pageviews = 78230/.0165 = 4741212

Net Conversion

- Baseline Conversion: 10.9313%
- Minimum Detectable Effect: .75%
- alpha: 5%
- beta: 20%
- 1 - beta: 80%
- sample size = 27413 enrollments/group
- Number of groups = 2
- total sample size = 54826
- clicks/pageview: 3200/40000 = .08
- pageviews = 54826/.08 = 685325

Pageviews Required: 4741212

Duration vs. Exposure

With 40000 cookies per day, and 100% traffic, it will take 119 days. If we eliminate retention, this reduces the number of required pageviews to 685325, and an result in a 18 day experiment with 100%

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

A 119 day experiment with 100% diversion of traffic presents both a business risk (potential for: frustrated students, lower conversion and retention, and inefficient use of coaching resources) and an opportunity risk (performing other experiments).

However, In general, this is not a risky 18-day experiment, given that it does not affect existing paying customers, and it will not take any sensitive information from individuals. There are some substantial impact on new enrollment, Like they might felt been blamed by web site for not paying enough hours, and thus give up.

Experiment Analysis

Sanity Checks

Number of cookies:

Total Control group pageview: 345543

Total Experiment group pageview: 344660

Total pageview: 690203

Probability of cookie in control or experiment group: 0.5

Standard error: SE = $\sqrt{0.5 \times 0.5 \times (\frac{1}{345543} + \frac{1}{344660})} = 0.0006018$

Margin of error (m) = SE * 1.96 = 0.0011796
Confidence Interval = [0.5-m,0.5+m] = [0.4988,0.5012]
Observed value = 344660/690203 = 0.5006
Result: Pass Sanity Check

Number of clicks:
Total Control group clicks: 28378
Total Experiment group clicks: 28325
Total pageview: 56703
Probability of cookie in control or experiment group: 0.5
Standard error: SE = $\sqrt{0.5 \times 0.5 \times (\frac{1}{28378} + \frac{1}{28325})} = 0.0021$
Margin of error (m) = SE * 1.96 = 0.0041
Confidence Interval = [0.5-m,0.5+m] = [0.4959,0.5041]
Observed value = 28378/56703 = 0.50046

Result: Pass Sanity Check

Click-through-probability:
Total Control click-through-probability: 0.0821
Total Experiment group click-through-probability: 0.0822

Standard error: SE = $\sqrt{0.0821 \times (1 - 0.0821) \times \frac{1}{345543}} = 0.000467$
Margin of error (m) = SE * 1.96 = 0.0009
Confidence Interval = [0.0822-m, 0.0822+m] = [0.0812,0.0830]

Observed value = 0.0822

Result: Pass Sanity Check

Result Analysis
Effect Size Tests
Gross Conversion:

	Control	Experiment
Clicks	17293	17260
Enrollment	3785	3423
Gross Conversion	0.2189	0.1983

SE = $\sqrt{(1 - 0.2086) \times 0.2086 \times (\frac{1}{17293} + \frac{1}{17260})} = 0.0044$
m = SE * 1.96 = 0.0086
Pooled Probability = 0.2086

D hat = 0.1983 - 0.2189 = -0.0206
Confidence Interval = [-0.0292,-0.0120]

It is statistically significant for not containing 0; It is practically significant for not containing dmin = 0.01

Net Conversion:

	Control	Experiment
Clicks	17293	17260
Enrollment	2033	1945
Net Conversion	0.1176	0.1127

SE = $\sqrt{(1-0.1151)\times 0.1151\times(\frac{1}{17293}+\frac{1}{17260})} = 0.0034$
m = SE * 1.96 = 0.0067
Pooled Probability = 0.1151
D hat = 0.1127 - 0.1176 = -0.0049
Confidence Interval = [-0.0116, 0.018]

It is not statistically significant for containing 0; It is not practically significant for containing dmin = 0.0075

Sign Tests

Metric	P-value	Statistically significant
Gross Conversion	0.0026	Y
Net Conversion	0.6776	N

Summary
I did not use Bonferroni correction because I want both gross conversion significantly decrease and net conversion does not significantly decreases requirement fulfilled. For both effect size hypothesis tests and the sign tests, gross conversion in experiment group is significantly less than gross conversion in the control group, net conversion in both groups has not significantly differed.

Recommendation
My final decision is launching the experiment is not recommended.
Gross conversion in new change would significantly reduce enrollment that does not pay. Net conversion in new change is being statistically and practically insignificant and the confidence interval includes negative numbers so, I'm not sure that this experiment would not decrease net

conversion or not. Therefore, there is a risk that the introduction of the trial screener may lead to a decrease in revenue.

Follow-Up Experiment

As a student, I know to complete any online program/courses we need much help and many materials. Some students might just get frustrated just because there is nobody to exchange idea or consult with, and thus quit the course after free trial. My idea is we can suggest student form some study team while they enrolled in a program, thus it can motivate students' enthusiasm about study and increase the retention.

Setup: Upon enrollment students will either be randomly assigned to a control group in which they are not into a team, or an experiment group in which they are.

Null Hypothesis: Participation in a team will not increase the number of students enrolled beyond the 14 day free trial period by a significant amount. $H_0 : \mu(\text{increase}) = \mu(\text{origin})$.

Unit of Diversion: The unit of diversion will be user-id as the change takes place after a student creates an account and enrolls in a course.

Invariant Metrics: The invariant metric will be user-id, since an equal distribution between experiment and control would be expected as a property of the setup.

Evaluation Metrics: The evaluation metric will be Retention. A statistically and practically significant increase in Retention would indicate that the change is successful.

If a statistically and practically significant positive change in Retention is observed, assuming an acceptable impact on overall Udacity resources the experiment will be launched.

References:

http://www.evanmiller.org/ab-testing/sample-size.html
https://en.wikipedia.org/wiki/Bonferroni_correction