

## Experiment Design

### Metric Choice

#### Invariant Metrics

- Number of Cookies – This is the number of unique users who visited the course overview page. It is distributed equally in control and experiment group. This number should not vary when we change display of “start free trial” page
- Number of clicks – This is the number of users clicked the free trial button. This metric is also the same as number of cookies in the way that it is the same for control and experiment group because the course overview page is still the same.
- Click-through-probability – This is unique cookies to click the "start free trial" button per unique cookies to view the course overview page. Because the users did not see the page we tested before they choose to click. So, this metric is a good choice for invariant metric.

#### Evaluation Metrics

- Gross conversion – This is number of users who enrolled in the free trial/ Number of users who clicked the Start Free Trial button. This number depended on how we show “start free trial page”. This is the number that we got after users see the message “5 or more hours per week” which had the direct impact for users to make a decision to enroll or not. In this experiment, we expected that this value should decrease substantially.
- Retention - This is the number of user-ids to remain enrolled past the 14 days trial period /number of users who enrolled in the free trial. The effect of users after seeing the message “5 or more hours per week” will have an effect on this value. The users who saw the message and still enroll were the ones who willing to spend the time in the courses and lead to still remain enrolled after trial period. The value should be higher in experimental group.
- Net conversion – This is number of user-ids remained enrolled for 14 days trial and at least make their first payment/ number of users clicked the Start Free Trial button. This value is the product of 2 previous value (Gross conversion and Retention), and we expect that this value should not drop. It should remain the same or increase.

#### Not fall into any metrics

- Number of user-ids – it should not be invariant metric and evaluation metric because this enrollment may depend on how we show the “start free trial page”, so it is not the invariant metric. It is also not quite useful to use this value as an evaluation metric because it is quite standalone. It will be more useful if we compare it with some value such as Number of users who clicked the Start Free Trial button.

## Measuring Standard Deviation

Using the baseline value in below table

Baseline	
Unique cookies to view page per day:	40000
Unique cookies to click "Start free trial" per	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

We can calculate the Standard Deviation like below table

	p	N	SD ( $\sqrt{p * (1-p) / N}$ )
Gross Conversion	0.20625	400	0.02023
Retention	0.53	82.5	0.05495
Net Conversion	0.1093125	400	0.01560

Gross Conversion and Net Conversion

- They use the number of cookies as denominator which is the unit of diversion which means the unit of diversion is the same as the unit of analysis. So, the analytical estimate can be comparable to the empirical analysis.

Retention

- This uses "Numbers of users who enrolls in the free trial" as a denominator which is not the unit of diversion. So, the analytical and the empirical estimate are not the same.

## Sizing

### Number of Samples vs. Power

I chose to not using Bonferroni correction because the metrics in the test has high correlation and Bonferroni seems to be too conservative to it. Then, I calculated pageviews needed using  $\alpha = 0.05$  and  $\beta = 0.2$

	Gross Conversion	Retention	Net Convnersion
Baseline Conversion	0.21	0.53	0.11
Minimum Detectable Effect	0.01	0.01	0.01
alpha	0.05	0.05	0.05
beta	0.20	0.20	0.20
1 - beta	0.80	0.80	0.80
Sample Size	25,835.00	39,155.00	27,413.00
Number of Groups	2.00	2.00	2.00
Total Sample size	51,670.00	78,310.00	54,826.00
Clicks/ Pageviews	0.08	0.02	0.08
Pageviews	645,875.00	4,746,060.61	685,325.00
Pageview Required	4,746,060.61		

Analysis: Retention need 4,746,060 pageviews which means 117 days to finish the experiment. This seems very unrealistic. So, I decided to drop this metric out and use only Gross Conversion and Net Conversion from now on.

### Duration vs. Exposure

Realistic Pageviews	685,325.00	685,325.00
Redirected Percent	0.5	1
Number of days used	34.27	17.13

Based on the table above, if we redirect 100% of traffic for experiment, it will take only 18 days. And we can also choose to use only 50% of the traffic in this experiment, which means 25% will go to control group and 25% will go to experimental group and we will take 35 days to complete the test. In some A/B testing, this amount of value may be a little bit high because we will test with 25% of all users. There may be a potential economic lose in the case that if the experimental users saw the message "5 or more hours per weeks" and feel that they cannot commit this amount of time. They may choose not to enroll and the result of experiment will lead to lower conversion, lower retention and we will also end up with inefficient use of coaches. However, extending the duration of this experiment would just mean we lose the same customer in a longer amount of time. And if we consider other aspects of the risks, this experiment did not affect the chance that anybody will get hurt and we did not deal with any sensitive data. Therefore, it is not risky and we can divert all of 100% traffic into this test to reduce the time, lower the expense and increase a chance to some other tests.

. Because if we reduce the risk by half, we will need 2 months to complete this test and it may be too long, which is another risk.

## Experiment Analysis

### Sanity Checks

	Number of cookies	Number of clicks on "start free trial"
STANDARD DEVIATION	0.00060	0.00210
MARGIN OF ERROR	0.00118	0.00412
LOWER BOUND	0.49882	0.49588
UPPER BOUND	0.50118	0.50412
OBSERVED	0.50064	0.50047
Sanity Check	PASS	PASS

For number of cookies and number of clicks on "start free trial", the observed value is within upper bound and lower bound. So, these two pass the sanity check.

Click-through-probability on "start free trial"	
Control Value	0.08213
Standard Deviation	0.00047
Margin of Error	0.00092
Lower Bound	0.08121
Upper Bound	0.08304
Experimental Value	0.08218
Sanity Check	PASS

The same go for Click-through-probability on "start free trial", the observed value is within lower bound and upper bound, so this metric also pass the sanity check.

## Result Analysis

### Effect Size Tests

The below table show the calculation

<b>Gross conversion</b>		
N_cnt = clicks_controlled		17293
X_cnt = enroll_controlled		3785
N_exp = clicks_experiment		17260
X_exp = enroll_experiment		3423
p_pooled = (X_cnt + X_exp) / (N_cnt + N_exp)		0.2086
se_pooled = sqrt(p_pooled * (1-p_pooled) * (1./N_cnt + 1./N_exp))		0.00437
d = X_exp / N_exp - X_cnt / N_cnt		-0.02055
z-score = 1.96 So, se_pool * zScore =		0.00857
lower = d - se_pooled =		-0.02912
upper = d + se_pooled		-0.01199
<b>Net conversion</b>		
N_cnt = clicks_controlled		17293
X_cnt = enroll_controlled		2033
N_exp = clicks_experiment		17260
X_exp = enroll_experiment		1945
p_pooled = (X_cnt + X_exp) / (N_cnt + N_exp)		0.1151
se_pooled = sqrt(p_pooled * (1-p_pooled) * (1./N_cnt + 1./N_exp))		0.00343
d = X_exp / N_exp - X_cnt / N_cnt		-0.00487
z-score = 1.96 So, se_pool * zScore =		0.00673
lower = d - se_pooled =		-0.01160
upper = d + se_pooled		0.00186

#### Gross Conversion

- The confidence interval is [-0.02912, -0.01199] which does not contain 0, the metric is statistical significant.
- The confidence interval also did not contain d\_min value, so it is also practically significant.

#### Net Conversion

- The confidence interval is [-0.01160, 0.00186] which contains 0, the metric is not statistical significant.
- The confidence interval also contains d\_min value, so it is also not practically significant.

## Sign Tests

	Number of success	Number of Trial	Probability	P-Value	Statistically Significant @ alpha .05?
Gross Conversion	4	23	0.5	0.0026	yes
Net Conversion	10	23	0.5	0.6776	no

Gross Conversion – Statistically significant

Net Conversion – Not statistically significant

## Summary

I chose to not using Bonferroni correction because the metrics in the test has high correlation and in order to launch, we need these 2 metrics to match our expectation (decrease in gross conversion and no decrease in net conversion). We are in the situation that all metrics need to be all matching not any metrics need to be matching. If it is the latter case, it may be appropriate to use Bonferroni. But in this case, which is the prior one, we should not use Bonferroni.

The effect size test and sign tests are showing the same result for gross conversion which is practically significant. But they are also both result in not practically significant in net conversion rate.

## Recommendation

The purpose of this experiment is the see if we filter the student by showing by time commitment, it should result in increasing the number of paid users. However, the result from this A/B test using Effect Size Tests and Sign Tests show that the net conversion is not significantly different. And the confidence interval of the net conversion does include the negative of the practical significance boundary. So, I recommend to not apply this feature and the team may try to come up with another idea and retest again.

## Follow-Up Experiment

Using my own experience in taking the online course, for me, when I choose to take any online courses, I am quite sure that I can finish the course without any problem because I have a broad knowledge of programming languages and any programming related technologies. But when I talked with my friends or colleagues, they feel that they may not understand the content after enrollment or cannot finish the course. Based on this experience, I would like to test that if we have “pre-test” for the course, we can filter out the unsure students and can make the students who pass the pre-test have more confident in the course. The test is quite simple, just pop up dialog “You need to take 10 minutes free pre-test to ensure your background knowledge

is suitable for this course.” after users apply for the free trial. And we will try to see the result from this experiment. The steps of this experiment will be like below:

**Null hypothesis** – passing the pre-test will not increase the number of students who enroll after 14 days free trial period.

**Unit of Diversion** – They should be number of cookies, number of clicks and click-through-probability. We will apply this change to only after users create new account and start a free trial.

**Invariant Metric** – – They should be number of cookies, number of clicks and click-through-probability. Because we need to equally distribute the users to experimental group and control group.

**Evaluation Metric** – It should be Retention and Net Conversion. A statistically and practically significant increase in Retention and Net Conversion will show that this change is successful.

If we can get a statistically and practically significant positive change in Retention and Net Conversion, we will release this new feature to every student.