

A/B TESTING

Submitted By : Rahul Saxena

Experiment Overview:

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

- **Number of cookies:** That is, number of unique cookies to view the course overview page. ($d_{\min}=3000$)
- **Number of user-ids:** That is, number of users who enroll in the free trial. ($d_{\min}=50$)
- **Number of clicks:** That is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is triggered). ($d_{\min}=240$)
- **Click-through-probability:** That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page. ($d_{\min}=0.01$)
- **Gross conversion:** That is, number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button. ($d_{\min}=0.01$)
- **Retention:** That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout. ($d_{\min}=0.01$)
- **Net conversion:** That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button. ($d_{\min}=0.0075$)

INVARIANT MERTICS

An invariant metric should not change across experimental and control groups. After conducting the experiment, they are used to perform sanity check on results. Since screener pops up after clicking start free trial button. So **#page views /cookies, #clicks & click through rate** will remain unchanged. *Gross Conversion, Retention, Net Conversion* metrics get affected after screener so these should not be considered as Invariant metrics.

Therefore, invariant metrics chosen for the experiment are:

- Number of cookies
- Number of Clicks
- Click through Probability

EVALUATION METRICS

Evaluation metrics are expected to change. Metrics which may help in evaluating the hypothesis should be considered as evaluation metrics. They should occur after the screener.

- Number of user-ids : It will not be considered as an evaluation metric as gross conversion is a fraction of user-id and using gross conversion it's a better way of tracking the effect of screener.
- Gross Conversion : It will be considered as an evaluation metric. It is the ratio of no. of user-ids & no. of unique clicks on "Start free trial".
- Retention : It will be considered as an evaluation metric. Since it is the ratio of no. of people who made payments & no. of unique user-ids.
- Net Conversion: It will be considered as an evaluation metric. Since it is the ratio of no. of people who made payments & no. of clicks on "Start free trial" button.

Overall goal is to minimize student frustration that left free trial because they didn't have enough time. For considering launch of this Test :

1. Net Conversion should not decrease as students to continue past free trial would increase or remain same.
2. Retention is students who stay after 14 days of trial. It should not be decreased.
3. Gross Conversion should decrease as screener will filter out the students.

MEASURING STANDARD DEVIATION

Using the rough estimates of the baseline values for the metrics, we have calculated Standard Deviation of evaluation metrics analytically.

Gross Conversion	0.0202
Retention	0.0549
Net Conversion	0.0156

For Gross Conversion & Net Conversion, the unit of diversion is equal to unit of analysis. So, the analytical estimate of S.D. tends to be empirical estimate of S.D. But in case of Retention Unit of diversion is not same as unit of analysis. After calculating cookies, duration & exposure, if we decide keep retention, we have to calculate empirical variability of retention.

SIZING

NUMBER OF SAMPLES VS. POWER

We are not using Bonferroni correction. Bonferroni correction is too conservative for these metrics. I want all my metrics to be significant.

$$\alpha = 0.05 \text{ \& } \beta = 0.2$$

$$\text{Ratio of page views to clicks} = \frac{3200}{40000} = 0.08$$

$$\text{Ratio of page views to enrollments} = \frac{660}{40000} = 0.0165$$

$$\text{No. of page views for Gross Conversion} = \frac{25835 * 2}{0.08} = 645875$$

$$\text{No. of page views for Retention} = \frac{39115 * 2}{0.0165} = 4741212$$

$$\text{No. of page views for Net Conversion} = \frac{27413 * 2}{0.08} = 685325$$

We will have to satisfy all our conditions so, will select maximum no. of pageviews.

$$\text{No of Page Views} = \mathbf{4741212}$$

DURATION vs EXPOSURE

Duration can be calculated based upon the exposure which is inversely proportional to the Risk involved. Here, the screener is reminder about time commitment, it constitutes minimal risk. None of the participants will suffer from physical harm nor is the data too sensitive, so even if we will give more than 50% exposure it will be okay. I am considering 75% exposure.

$$\text{No. of days for 4741212 page views with 75\% exposure} = \frac{4741212}{0.75 * 40000} \sim 158 \text{ days}$$

It's a long time period and we should reduce the time duration. We can exclude Retention as an evaluation metric and will consider Net conversion, with revised no. of page views as 685325.

$$\text{No. of days for 685325 page views with 75\% exposure} = \frac{685325}{0.75 * 40000} \sim 23 \text{ days}$$

Excluding retention as a metric still allows us to test our hypothesis with net conversion. The two metrics are correlated. Retention measures the difference in the rate at which people drops from enroll to completion of trial. Net conversion uses the no. of users to complete the trial and retains even after the trial. So there is no such issue even after excluding Retention as a metric.

EXPERIMENT ANALYSIS: SANITY CHECKS

Metric	Lower Bound	Upper Bound	Observed Value	Result
No. of Cookies	0.4988203921	0.5011796079	0.5006396669	Pass
No. Clicks	0.4958844957	0.5041155043	0.5004673474	Pass
Click through probability	0.08121035975	0.0830412674	0.08212581357	Pass

RESULT ANALYSIS

EFFECT SIZE TESTS

95% Confidence interval around the difference between the experiment and control group for evaluation metrics.

Metric	d_{min}	LB	UB	Statistical Sign	Practical Sign
Gross Conv	0.01	-0.029123358	-0.011986390	True	True
Net Conv	0.01	-0.0116046244	0.001857178971	False	False

Gross Conversion is both Statistical Significant and Practical Significant

Net Conversion is neither Statistical Significant nor Practical Significant

SIGN TESTS

Metric	p-value	Statistically Significant
Gross Conversion	0.0026	Yes
Net Conversion	0.6776	No

SUMMARY

Bonferroni correction is not used here because our launch decision is based upon two metrics Gross Conversion & Net Conversion. In order to launch, we need both the metrics to match our expectation. We risk not launching as if at least one metric (out of 2) fail to reject null, when null is not the true effect.

If we were to launch the experiment when any metric would match our expectations, then we would have to use Bonferroni correction. Bonferroni is used to reduce risk but can only be used conditionally.

The sign test mirrors that of size test, that gross conversion is significant but net conversion is not. Both the tests are giving same results on being metrics to significant, so no further study is required.

RECOMMENDATION

For making a recommendation on whether to launch screener or not, we need to check evaluation metrics.

Gross Conversion is both statistically & practically significant; it means we are successful in decreasing the no. of enrollments.

Net Conversion is neither statistically nor practically significant. The CI of Netconversion includes negative of practical significance boundary. So, the no. of people who will stay past trial could reduce. This is risky & not a good change according to this metric.

Considering both the points, we should not launch the experiment.

FOLLOW-UP EXPERIMENT

We can perform another experiment by changing the number of hours from screener to prerequisite knowledge required to pursue that course. This screener may help students to get an idea about what knowledge they should have before joining Udacity course. If the student does not have that knowledge, he/she may click on suggested courses that are listed in prerequisites and can join those particular courses.

Null Hypothesis: No significant difference between control and experiment groups.

Unit of diversion = Cookie

For testing this hypothesis, we have to measure #cookies, #clicks, #enrollments, #Payments and from these we will calculate Gross Conversion & Net Conversion.

If Gross Conversion & Net Conversion will result to be statistically & practically significant then we will be able to launch our test.

REFERENCES:

1. Udacity Discussion forum
2. <http://www.exp-platform.com/Documents/controlledExperimentDMKD.pdf>
3. <http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/type-i-and-type-ii-error/>