

Design an A/B Test

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.

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.

1 Experiment Design

1.1 Metric Choice

1.1.1 Invariant Metrics:

- Number of cookies (number of unique users to visit the course overview page): Unit of Diversion is cookies, it's evenly distributed in control and experiment group. It is therefore appropriate as an invariant metric.
- Number of clicks (number of unique cookies to click the start free trial button): The click happens before the users see the experiment, it's independent, and also appropriate as an invariant metric.
- Click-through-probability (unique cookies to click the "start free trial" button per unique cookies to view the course overview page): This is also an invariant metric since the clicks happen before the experiment.

1.1.2 Evaluation Metrics:

- Gross conversion (the number of user-ids to complete checkout and enroll in the free trial per unique cookie to click the "start free trial" button): A good evaluation metric since it depends directly on the effect of the experiment.

- Retention (the number of user-ids to remain enrolled past the 14-day trial period, making at least one payment, per number of user-ids to complete checkout): It's an evaluation metric because users are completely aware of the time commitment, they made a deliberate decision to enroll in the course or continue learning for free. We can expect more students staying beyond the free trial in the experiment group

- Net conversion (the number of user-ids to remain enrolled past the 14-day trial, making at least one payment, per the number of unique cookies to click the "start free trial" button): The net conversion is the final result connected to deliberate user decisions along with previous two evaluation metrics. We expect that having a "5 or more hours per week" button will increase the net conversion.

To launch the experiment one of the following must be true:

- in the experiment group more students stay beyond the free trial (means increased retention),
- in the free trial less students enroll in the course (means decreased Gross Conversion), but at the same time no less students stay beyond the free trial (means not decreased Net Conversion).

1.1.3 Unused metrics

- Number of user-ids (the number of users to enroll in the free trial): User-ids are tracked only after enrolling in the free trial, therefore it's not a bad invariant metric. It could be used as an evaluation metric, however it's not ideal, because it's not normalized.

1.2 Measuring Standard Deviation

Metric	Standard Deviation
Gross Conversion	.0202
Retention	.0549
Net Conversion	.0156

The analytical estimate of standard deviation tends to be close to the empirically determined standard deviation when the unit of diversion is equal to the unit of analysis.

Both Gross Conversion and Net Conversion use number of cookies as the denominator, which is also unit of diversion. Therefore, the analytical and the empirical estimates would be comparable.

Retention uses "Number of users enrolled the courseware" as the denominator. The unit of analysis and the unit of diversion are not the same. Therefore, the analytical and the empirical estimates would be different.

1.2 Sizing

1.2.1 Number of Samples vs. Power

We will not deploy the Bonferroni correction, because the metrics are highly correlated and the Bonferroni correction will be too conservative to it.

	Gross Conversion	Retention	Net Conversion
Baseline Conversion	20.625%	53%	10.9313%
Minimum Detectable Effect	1%	1%	0.75%
alpha	5%	5%	5%
beta	20%	20%	20%
1 - beta	80%	80%	80%
Sample size	25835 enrollments/group	39155 enrollments/group	27413 enrollments/group
Number of groups	2	2	2
Total sample size	51670 enrollments	78230 enrollments	54826 enrollments
Calculation	3200 clicks /40000 pageviews = 0.08	660 enrollments /40000 pageviews = 0.0165	3200 clicks /40000 pageviews = 0.08
Pageviews	645875	4741212	685325

I think we should drop the Retention, because it will take a considerably long time. This reduces the number of required pageviews.

Total pageviews required: 685325

1.2.2 Duration vs. Exposure

The whole experiment is not considered as highly risky because:

- additional functions are completely not harmful to users,
 - although payment details are being collected, there is no change in this process (at least it's not more risky).
- An 18-day experiment with 100% diversion rate is quite reasonable,

2 Experiment Analysis

2.1 Sanity Checks

We expect equal diversion into the experiment and control group. The confidence interval is 95%.

Metric	Expected Value	Observed Value	CI Lower Bound	CI Upper Bound	Pass
Number of Cookies	0.5000	0.5006	0.4988	0.5012	Yes
Number of clicks on "start free trial"	0.5000	0.5005	0.4959	0.5042	Yes
Click-through-probability	0.0821	0.0822	0.0812	0.0830	Yes

2.2 Result Analysis

2.2.1 Effect Size Tests

	dmin	Observed Difference	CI Lower	CI Upper	Statistically significant	Practically significant
Gross Conversion	0.01	-0.0205	-0.0291	-0.0120	Yes	Yes
Net Conversion	0.0075	-0.0048	-0.0116	0.0019	No	No

2.2.2 Sign Tests

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

2.2.3 Summary

Requirement for launching the experiment:

- the null hypothesis must be rejected for all evaluation metrics,
- the difference must exceed the practical significance threshold.

The Bonferroni correction is a method for controlling for type I errors when using multiple metrics. Because the acceptance criteria in our case requires statically significant differences for all evaluation metrics, the use of the Bonferroni correction is not appropriate. In simple words, the correction is applied if we use logical “or” on all metrics, but we use “and” for all our metrics.

2.3 Recommendation

Gross Conversion is significantly lower; Net Conversion is not significantly decreased. This means a decrease in enrollment with no significant decrease in students staying for 14 days and paying.

However, the practical lower boundary (-0.0075) lies between the observed difference (-0.0049) and its lower CI (-0.0116). Therefore, I highly recommend to postpone launching the experiment and conduct additional research.

3 Follow-Up Experiment

I remember my first project on Udacity and my frustration about at the very beginning of the path. The course materials are well-designed, structured and explained. My first difficulty was however making a first step in the project.

It was always hard to begin. The process of working/reviewing was not so clear, it made me think that I should perfectly polish my project before sending it. I also thought that I probably have only one chance to send it, if I send – I receive a bad mark which will stay with me till my graduation. I also didn’t understand that if I need some help I can make an appointment with a coach.

Therefore, I propose a new “Introduction into your first project” course that would explain how to successfully finish your first project, what to do if you have difficulties, if you need help. It will go through the whole process of working on a project and explain it in details.

Setup:

After enrollment students will be randomly assigned to a control group in which they have a new additional mini-course.

Null Hypothesis:

Having the mini-course will not significantly increase Retention.

Unit of Diversion:

User-id. Because the split takes place after creating an account and enrolling in a course.

Invariant Metrics:

User-id. We can expect an equal distribution between experiment and control as User-id appears before the experiment (immediately after enrolling) and is stored in a database.

Evaluation Metrics:

Retention. A statistically and practically significant increase in Retention would mean that the mini-course is successful.

References

<https://classroom.udacity.com/nanodegrees/nd002/syllabus>

<http://www.evanmiller.org/ab-testing/sample-size.html>

<http://graphpad.com/quickcalcs/binomial1.cfm>