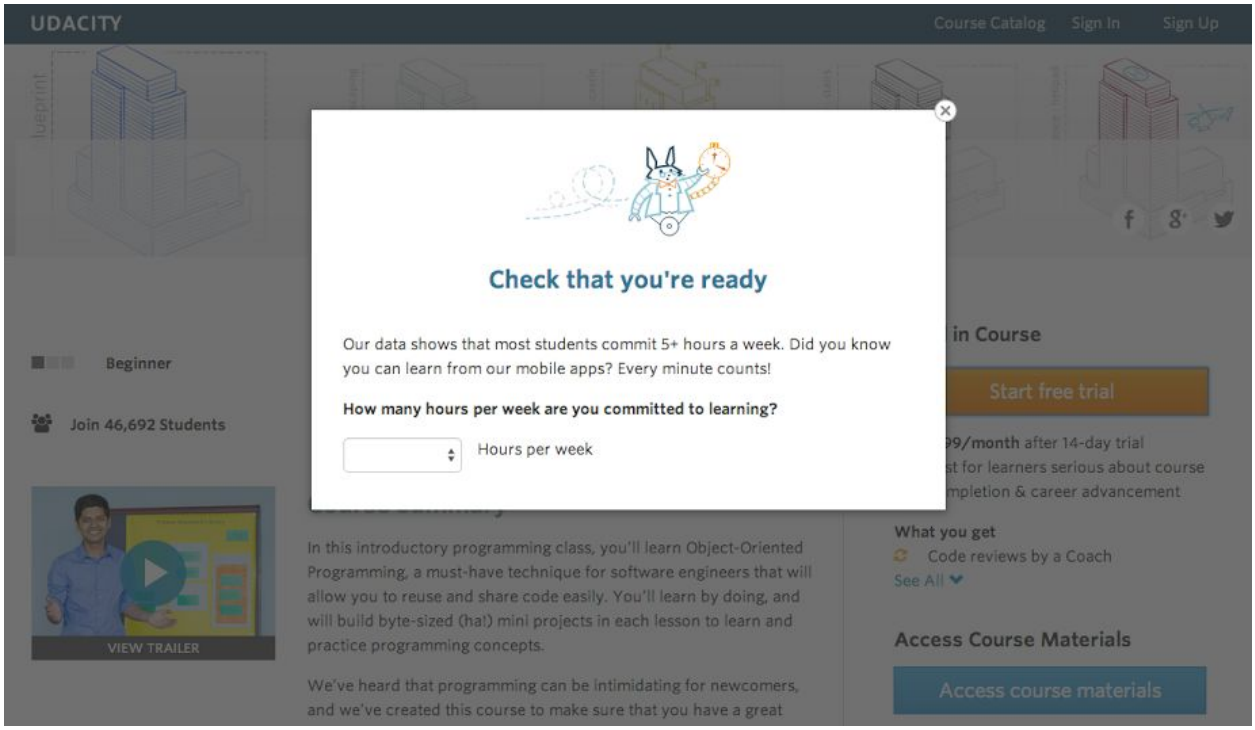
Katerina Baneva Project 7

A/B Test Design



A/B Experiment Overview: Free Trial Screener

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. 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.

Experiment Design

Metric Choice

Invariant metrics:

- **Number of cookies:** That is, number of unique cookies to view the course overview page. (d_{min}=3000). Is also the Unit of diversion. It is supposed to be equally distributed across the experiment and control groups. Also, the cookies see the page before the experiment. The metric is independent from the experiment.
- **Number of clicks:** That is, number of unique cookies to click the "Start free trial" button. (which happens before the free trial screener is trigger). (d_{min}=240). Therefore, the metric is independent from the experiment.There is an equal chance that the students from the control or experiment group will click.
- **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) The metric is independent from the experiment. It is a better metric compared to number of clicks since it normalizes to the size of the control and experiment groups.

Evaluation metrics:

- **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) Evaluating this metric can show us the effect of the experiment, meaning were we able to decrease the enrollments that would not convert into paying customers. Can not be used as invariant metric because the users who enroll in the free trial are dependent on the experiment.
- **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)The metric allows us to evaluate the financial outcome of the experiment and is directly connected with it. Can not be used as invariant metric because the users who enroll in the free trial are dependent on the experiment.

- **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) Can not be used as invariant metric because the users who enroll in the free trial are dependent on the experiment.The metric allows us to evaluate the financial outcome of the experiment and is directly connected with it.

Metrics that i did not include in my analysis:

- **Number of user_IDs**: That is, number of users who enroll in the free trial. (d_{min}=50) The number of students who enroll in the free trial is dependent on the experiment, so not appropriate for an invariant metric. It is not the best evaluation metric as it is not normalized.

Gross conversion should shows lower costs by introducing the screener and we should expect practically significant decrease.For Net Conversion it is it is sufficient not to decrease.

Measuring Standard Deviation

Evaluation Metrics	Standard Deviation
Gross conversion	0.0202
Retention	0.0549
Net conversion	0.0156

Unit of Analysis(denominator) for Gross conversion and Net conversion is equal to "unique cookies to click the "Start free trial" button", and for Retention the denominator (Unit of Analysis) is equal to "number of user-ids to complete checkout". Our Unit of Diversion is a cookie. If Unit of Diversion and Unit of Analysis are different the empirically computed variability will be much higher than the analytically computed variability.Therefore, for Retention it is better to move to empirically and analytically computed variability. For the other two evaluation metrics, the analytical estimate of the variance can be used.

Sizing

Number of Samples vs. Power

I decided that it is not necessary to use the Bonferroni correction because I need both of my evaluation metrics to be significant in order to make recommendation.

The evaluation metrics that I selected are Gross conversion and Net conversion. I will be needing 685 324 pageviews to power the experiment.

Duration vs. Exposure

I chose to divert 100 % of the traffic which means that 18 days will be needed to perform the experiment. I think the experiment does not hide any risk.

Experiment Analysis

Sanity Checks

	Number of cookies	Number of clicks	CTP
Lower bound	0.4988	0.4959	0.0812
Upper Bound	0.5012	0.5041	0.0830
Observed	0.5006	0.5005	0.0821
Passes	Yes	Yes	Yes

Result Analysis

Effect Size Tests

	Gross conversion	Net conversion
Ppool	0.20864	0.1151
Pcon	0.2188	0.1175
Pexp	0.1983	0.1126
dhat	-0.0205	-0.0048
SEpool	0.0043	0.0034
m	0.0085	0.0067
dhat - m	-0.0291	-0.0116
dhat + m	-0.0120	0.0019
Statistical Significance	Yes	No
Practical Significance	Yes	NO

Sign Tests

	Gross conversion	Net conversion
Number of successes	4	10
Number of Trial	23	23
Probability	0.5	0.5
Two-tailed P value	0.0026	0.6776
Significance	Yes	No

Summary

There were no discrepancies between the sign and the effect size tests. I did not use the Bonferroni correction because I need both of my evaluation metrics to be significant in order to launch the experiment.

Recommendation

Gross conversion turned out to be statistically, practically significant and negative. Therefore, we were able to reduce the number of people who would sign in for a free trial but would not have the time to work on their courses and in the end would not pass the first payment.Net conversion on the other hand turned out to be statistically and practically insignificant.Also, that might have a negative impact on the revenue since we have negative numbers in our confidence interval. Taking that into consideration performing additional test would be helpful to make a decision for launching or not the change.

Follow-Up Experiment

Udacity introduced the Teams feature before over 4 months but sadly they decided to retire the program due to low attendance. I believe that the feature was really useful to people who needed support and wanted to socialize.The fact that exactly those people think that canceling the program isn't the right way to go supports my thinking. As a Team Guide of 11 teams, I can confer that the active people in the groups would like to continue to socialise and they appreciate the support that they get. A lot of them wanted to join other groups where they can find more people with interests similar to theirs.One thing that Udacity could change in order to imrpove the Teams feature is to ask the students upon enrolling weather they want to be part of a team or they do not. Because we can not assume that all the students who enroll in the

nanodegree want to be part of a team. Not everyone likes to socialize, seek help or give one, but for those who do Teams is a great place to be part of.Also, it is a great way to reduce early cancellations.

My hypothesis is that by grouping people who like to socialize, really need support and are willing to provide one in return, Retention will increase by practically significant amount.

Upon enrollment students will randomly be assigned to a control and an experiment group. Students in the experiment group will have the option to join a team and they will be introduced to the responsibilities and benefits of that being a team member. The experience for the users in the control group will not change.

The Unit of Diversion will be the user - id because any impact can be seen after the user enrolls into the nanodegree.

The Invariant metric will be Number of users-ids because the option for joining a team will be present after enrolling into the nanodegree.

Retention will be the Evaluation metric. To have a positive outcome, we need retention to increase and to be practically significant. If this is the case we can launch the feature.