# Houston, Ready for Launch?

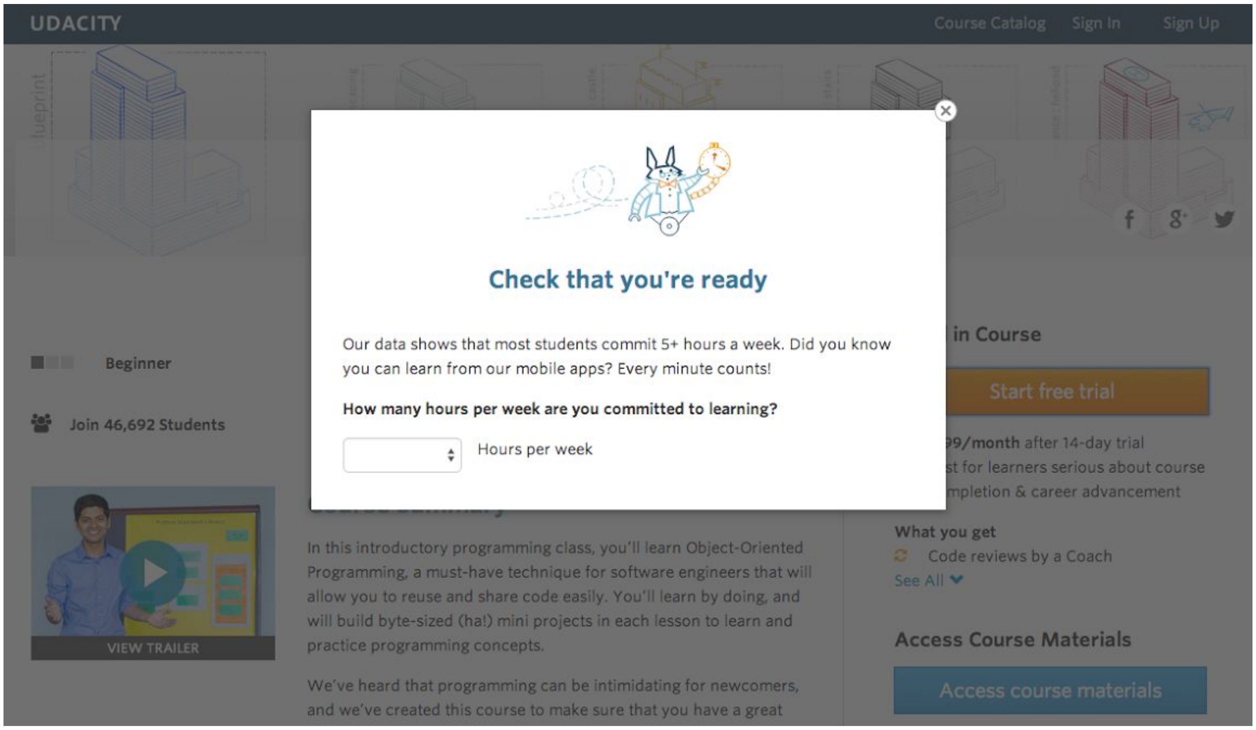
*A Udacity A/B Test*

By: Nicholas Cica

## Experiment Overview: Free Trial Screener

From the Project Description:

> 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.



Check that you're ready

Our data shows that most students commit 5+ hours a week. Did you know you can learn from our mobile apps? Every minute counts!

How many hours per week are you committed to learning?

Hours per week

99/month after 14-day trial
st for learners serious about course
mpletion & career advancement

In this introductory programming class, you'll learn Object-Oriented Programming, a must-have technique for software engineers that will allow you to reuse and share code easily. You'll learn by doing, and will build byte-sized (ha!) mini projects in each lesson to learn and practice programming concepts.

We've heard that programming can be intimidating for newcomers, and we've created this course to make sure that you have a great

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

For this experiment, each metric will need to be defined as either a **Invariant Metric**, **Evaluation Metric** or removed from the experiment.

Invariant metrics occur before the change in the pipeline and perform some kind of **Sanity Check** before running the experiment and verify whether there is a similar distribution between the control and experimental groups. If the sample distribution between the control and experimental group differs, then the experiment should not be performed because some factor is influencing the distribution. This could include variables like time of year, day of the week, or even time of day. Evaluation metrics, on the other hand, occur after the change in the pipeline and will inform the hypothesis.

## Invariant Metrics

### Number of Cookies

- Number of unique cookies to view the course overview page.
- This is an **Invariant Metric** because it occurs before the experimental factor and will inform the distribution of the control and experimental group. It will not be used as an evaluation metric because viewing the course overview page occurs before the change in the pipeline.

### Number of Clicks

- Number of unique cookies to click the “Start free trial” button (which happens before the free trial is triggered).
- This is an **Invariant Metric** because it occurs before the experimental factor and will inform the distribution of the control and experimental group. It will not be used as an evaluation metric because viewing the course overview page occurs before the change in the pipeline.

### Click-through-probability

- Number of unique cookies to click the “start free trial” button divided by the number of unique cookies to view the course overview page.
- This is an **Invariant Metric** because it occurs before the experimental factor and will inform the distribution of the control and experimental group. It will not be used as an evaluation metric because viewing the course overview page occurs before the change in the pipeline.

## Evaluation Metrics

### Gross Conversion

- 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.
- This is an **Evaluation Metric** because it occurs after the modification and it will inform the experiment’s hypothesis. It will not be used as an invariant metric because it occurs after the change in the pipeline being evaluated by the experiment.

### Retention

- 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.

- This is an **Evaluation Metric** because it occurs after the modification and it will inform the experiment’s hypothesis. It will not be used as an invariant metric because it occurs after the change in the pipeline being evaluated by the experiment.

### Net Conversion

- Number of user-ids 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.
- This is an **Evaluation Metric** because it occurs after the modification and it will inform the experiment’s hypothesis. It will not be used as an invariant metric because it occurs after the change in the pipeline being evaluated by the experiment.

## Excluded Metrics

### Number of User-Ids

- Number of Users who enroll in the free trial.
- This could be used as an Evaluation Metric because it could be used to track whether the number of students who continue past the free trial has changed however, is not the best metric as it is not normalized.

In order to launch the experiment, one or more of the following should be observed:

- Increased Retention
- Decreased Gross Conversion without a decrease in Net Conversion

## Measuring Standard Deviation

| Evaluation Metric | Standard Deviation |
|---|---|
| Gross Conversion | 0.0202 |
| Retention | 0.0549 |
| Net Conversion | 0.0156 |

For Gross Conversion and Net Conversion, the analytic estimate would be **comparable** to the empirical variability because the unit of diversion is equal to the unit of analysis.

For Retention, the analytic estimate would be **not be comparable** to the empirical variability because the unit of diversion is not equal to the unit of analysis..

### Sizing

#### Number of Samples vs. Power

The Bonferroni Correction will not be used during the analysis because it increases the probability of producing **false negatives** which could lead to launching the experiment under false pretenses.

The size of the experiment is dependent on which evaluation metrics are included. If Retention is included, **4,741,212** pageviews are needed. Removing Retention and changing the focus to only Gross Conversation and Net Conversion would instead require approximately **685,325** pageviews to power the experiment.

#### Duration vs. Exposure

Udacity receives 40,000 pageviews per day and this number will be used to approximate how long the experiment will run.

Normally, if **100%** of the traffic is diverted to this experiment, it would take **119** days to run. This seems like an excessive amount of time for an experiment since the results would not be known for approximately four months. If they are extremely negative, this could have devastating effects. If Retention is removed as an evaluation metric, and only Gross Conversion and Net Conversion are used, it would take **18** days to run the experiment.

Another factor that must be considered if the amount of risk. Minimal risk is defined as the probability and magnitude of harm that a participant would encounter in normal daily life. Since the experiment is neither health or financial related, the experiment proposes minimal risk. Therefore, the proposal is to divert **100%** of the traffic toward the experiment for **18** days.

## Experiment Analysis

### Sanity Checks

For each of the invariant metrics, the 95% confidence intervals and their actual observed values are as follows:

#### Number of Cookies:

- CI Lower Bound: 0.4988
- CI Upper Bound: 0.5011
- Observed: 0.5006
- Sanity Check: Pass

#### Number of Clicks:

- CI Lower Bound: 0.4959
- CI Upper Bound: 0.5042
- Observed: 0.5005
- Sanity Check: Pass

### Click Through Probability on “Start Free Trial”:

- CI Lower Bound: 0.0013
- CI Upper Bound: 0.0013
- Observed: 0.0001
- Sanity Check: Pass

All three invariant metrics **passed** the sanity check.

## Result Analysis

### Effect Size Tests

For each of the evaluation metrics, a 95% confidence intervals around the difference between the experimental and control groups are below. Whether each metric is statistically and practically significant is also indicated. Retention has been removed to reduce the duration of the experiment.

### Gross Conversion

- CI Lower Bound: -0.0291
- CI Upper Bound: -0.0120
- Statistically/Practically Significant: Both

### Net Conversion

- CI Lower Bound: -0.0116
- CI Upper Bound: 0.0019
- Statistically/Practically Significant: Neither

### Sign Tests

For each evaluation metric a sign test was performed using the day-by-day data and reported the p-value of the sign test and whether the result is statistically significant.

### Gross Conversion

- P-Value for Sign Test: 0.0026
- Statistically significant: Yes

### Net Conversion

- P-Value for Sign Test: 0.6776

- Statistically significant: No

### Summary

After analysis, the **Invariant Metrics** had equal distribution in the control and experimental groups and passed the sanity check with a 95% confidence interval. The **Evaluation Metrics** were reduced from 3 to 2, and **Gross Conversion** was found to be both statistically and practically significant at a 95% confidence interval. Therefore, the NULL hypothesis was rejected. Finally, **Net Conversion** was found to be neither statistically or practically significant at a 95% confidence interval.

As stated above, the **Bonferroni Correction** was not used in the experiment because it increases the probability of producing false negatives. Since the criteria for launch is to have all criteria pass, the chance of a false negative impacting that decision is too high to include the Bonferroni Correction.

### Recommendation

Houston, we have a problem.

The purpose of the Udacity experiment was to determine whether presenting the incoming student with more information about time commitments would improve the overall learning experience. There has been no statistically significant change, but the confidence interval does include the negative practical significance boundary. That is, it is possible that the number of students continuing past the 14 day free trial decreased by an amount that would influence the business.

Therefore, it is **not recommend** to launch the change and instead, the recommendation is to conduct additional follow-up experiments.

## Follow-Up Experiment

The main goal of this experiment was to try and understand why students canceled early. Since this experiment has eliminated the students not understanding the course time commitment, the list of possible explanations can be expanded. Additional experiments will test new hypothesis.

The next experiment should evaluate whether additional information about the classes have an affect on student retention. For example, providing a complete syllabus that outlines the project scope and complexity may sway student behavior. There may be some subset of students who are enrolling and dropping because of a lack of understanding what coursework that will be covered in the class and end up merely “shopping” the class.

For the experiment, the **Course Information** screen will be changed and updated to include detailed information about the lessons and projects in the course. The course information page will be available before enrollment to every student.

The **NULL Hypothesis** is that the group participating in the experiment will not see an significant increase in the number of students enrolled after the 14 day free trial ends. This way we can accept or reject the NULL hypothesis based on the experimental findings.

The **Unit of Diversion** will be the Cookies as the experiment will need to be able to include non-logged in traffic to compute some specific metrics of interest.

The **Invariant Metric** will be the number of unique Cookies to view the course information screen. This user action occurs before the experiment and will inform the distribution of the control and experimental group.

The **Evaluation Metric** will be Retention. If it is statistically and practical significant, then the experiment will show that the change to the course information screen had an impact on the number of students who “cancelled early”.

Regardless of the results, additional experiments should be performed continuously to increase the amount of student retention.