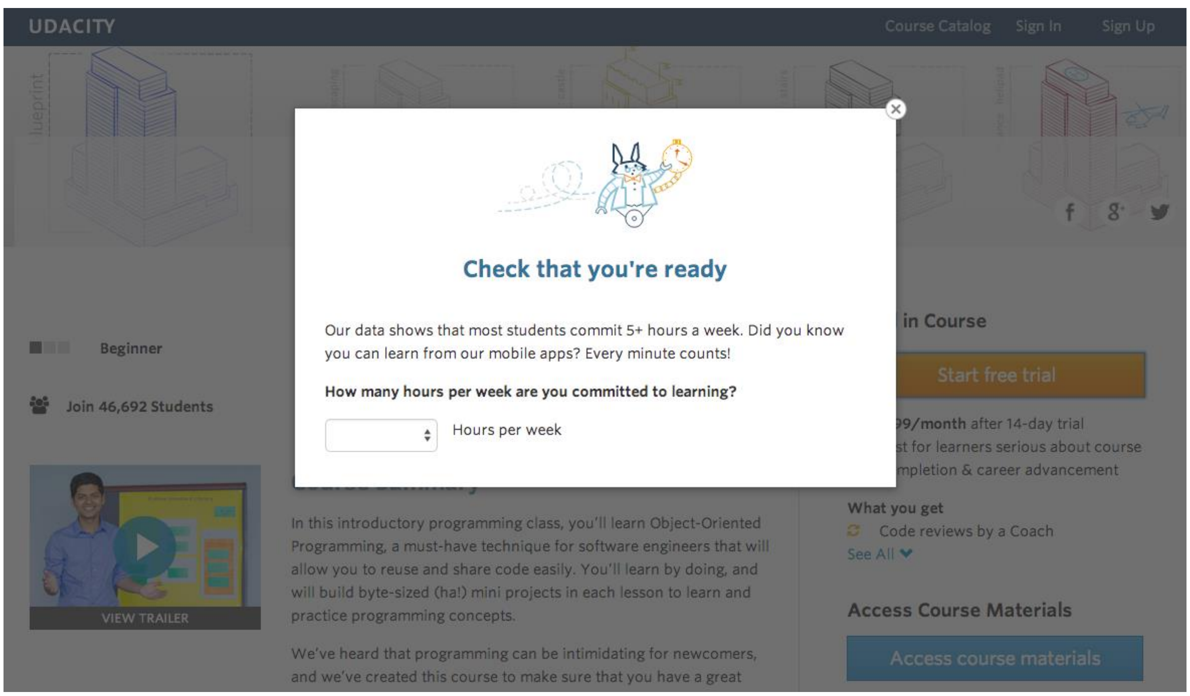
Experiment Overview: Free Trial Screener

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.



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.

Project Baseline Values

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

Design Process and Calculations

Experiment Design

Metric Choice

Invariant Metrics:

1. Number of cookies: Number of unique cookies to view the course overview page
2. Number of clicks: Number of unique cookies to click the "Start free trial" button
3. Click-through-probability: number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page

Evaluation Metrics:

1. 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

2. Net Conversion: 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

Other metrics:

1. 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

2. Number of user-ids: Number of users who enroll in the free trial

Choice of metrics:

1. **Number of cookies**: Since the unit of diversion is a cookie, the unique cookies are being randomly assigned to experiment and control group. So we should have roughly the same number of cookies in each group and can choose it as an invariant metric.

   Since the unique cookies to view the course overview page will not see the screener, the metric will not reflect the effect of the experiment and hence will not be a good choice for an evaluation metric.

2. **Number of clicks**: We would want an equal split between the number of users who click the button and are shown the screener, and the number of users who are not shown the screener, to measure the effect of experiment. So it makes a good choice for an invariant metric.

   Since users click the "Start free trial" button before the screener, the number of clicks on the button will not reflect the effect of introduction of screener. Therefore, it will not be a good choice for evaluation metric.

3. **The click-through-probability** is the ratio of number of unique cookies to click the button to the number of unique cookies to view the course overview page. So we expect the click-through-probability to be invariant as it is the ratio of two invariant metrics.

   Also since viewing the course overview page and clicking the "Start free trial" button occur before the user is shown the screener, we do not expect the click-through-probability to reflect the effect of experiment. So it will not be a good choice as an evaluation metric.

4. **Gross Conversion:** Since the users will see the screener after clicking "Start free trial" button, gross conversion should be affected by the change. The screener might avoid enrollment by students who might not be able to give the appropriate time to the course. So it should be kept as an evaluation metric.

Since the gross conversion might be different between the experiment and control group due to introduction of the screener, it will not be a good choice for invariant metric.

5. **Retention:** Retention could be chosen as an evaluation metric, but on further analysis the number of page views required to measure retention was too large. So I did not keep retention as an evaluation metric.

   Moreover, retention might be affected after the change is introduced, since it is measured after the intervention, so it will not be a good choice as an invariant metric.

6. **Net Conversion:** The aim of introducing the screener is to set the expectations for time commitment and reduce the number of students who drop out during the free trial period. We would launch the experiment only if there is no significant reduction in the number of students to continue past the free trial and eventually complete the course. Therefore, we would like to keep net conversion as an evaluation metric.

   Since the users will see the screener after clicking "Start free trial", net conversion might be affected after implementing the change. Therefore, it would not be a good choice as an invariant metric.

7. **Number of user-ids** is not ideal both as an invariant and evaluation metric. The number of user ids to enroll would be affected by the introduction of screener and we would expect to see different values in the experiment and control group, so it would not be a good invariant metric.

   Since the number of user-ids refers to the raw enrollment counts, it could get affected by the variability in distribution between control and experiment group. A small difference in group size could impact the accuracy, so a raw count metric like user-id would not be a good evaluation metric. In comparison, gross conversion which gives the probability to enroll given a click on the "start free trial" button is a ratio. It would be able to handle minor variations in split between control and experiment group.

Results required for launch:
We need to observe both the conditions given below in order to launch the experiment:

- Practically significant decrease in gross conversion
- Net conversion doesn't decrease below practically significant boundary (increase or no change would be fine)

Measuring Standard Deviation

Sample size of 5000 cookies visiting the course overview page:

$$Gross\ Conversion = \frac{Number\ of\ user - ids\ to\ enroll}{Number\ of\ unique\ cookies\ to\ click\ "Start\ free\ trial"}$$

$$Net\ Conversion = \frac{Number\ of\ user - ids\ to\ make\ a\ payment}{Number\ of\ unique\ cookies\ to\ click\ "Start\ free\ trial"}$$

$$S.D.(Gross\ Conversion) = \sqrt{\frac{0.020625(1-0.20625)}{3200/8}} = 0.0202$$

$$S.D.(Net\ Conversion) = \sqrt{\frac{0.109325(1-0.109325)}{3200/8}} = 0.0156$$

Gross Conversion: 0.0202
Net Conversion: 0.0156

Comparison between empirical and analytic estimate for variability:
The unit of analysis (denominator) for both the gross conversion and net conversion is the number of unique cookies to click the "Start free trial" button. The unit of diversion for the experiments is a cookie. Since the unit of analysis and unit of diversion are the same, the analytical variability will match the empirical variability.

Sizing

Number of Samples vs. Power

Number of pageviews needed (using alpha=0.05 and beta=0.2): 685325
(http://www.evanmiller.org/ab-testing/sample-size.html)

Bonferroni correction was not used since it is important to use the correction if a change will be launched if **any** of the metrics show a significant difference, as it is likely that one of the metrics will be falsely positive if the number of metrics is high.

In the present case, the change will be launched only if **all** of the evaluation metrics show a significant change, so there is no need to use the Bonferroni correction.

Duration vs. Exposure

The experiment isn't risky as it does not involve any backend changes or UI changes which could frustrate the users. The experiment also does not involve any sensitive data such as personal medical history. Therefore, the experiment is very low risk and we can divert a majority of the traffic. Therefore, in order to keep the duration of experiment optimal, I would choose to divert 80% of the traffic to the experiment. This would take approximately 3 weeks to finish the experiment and would be appropriate since it would include an equal amount of weekdays and weekends.

Fraction of Udacity's traffic to divert to this experiment: 0.8
Given this, the number of days Udacity will need to run the experiment: 22

$$Number\ of\ days = \frac{685325}{20000} = 21.416 \cong 22$$

Experiment Analysis
Sanity Checks

95% confidence interval for invariant metrics and sanity checks based on observed values

	Lower Bound	Upper Bound	Observed	Passes
Number of cookies	0.4988	0.5012	0.5006	Yes
Number of clicks on "Start free trial"	0.4958	0.5041	0.5004	Yes
Click-through-probability on "Start free trial"	-0.0013	0.0013	0.00006	Yes

Number of Cookies:

$$Standard\ Error = \sqrt{\frac{0.5 \times 0.5}{345543 + 344660}} = 0.000602$$

$m = z \times Standard\ Error = 1.96 \times 0.000602$ = 0.00118
95% Confidence Interval = (0.5-m, 0.5+m) = (0.4988, 0.5012)

Number of clicks on "Start free trial":

$$Standard\ Error = \sqrt{\frac{0.5 \times 0.5}{28378 + 28325}} = 0.00209$$

$m = z \times Standard\ Error = 1.96 \times 0.00209$ = 0.00411
95% Confidence Interval = (0.5-m, 0.5+m) = (0.4958, 0.5041)

Click-through-probability on "Start free trial":

$$\hat{p}_{pool} = \frac{X_{control} + X_{exp.}}{N_{control} + N_{exp.}} = \frac{28378 + 28325}{345543 + 344660} = 0.08215$$

$$SE_{pool} = \sqrt{0.08215 \times (1 - 0.08215) \times (\frac{1}{345543} + \frac{1}{344660})} = 0.00066$$

$m = z \times Standard\ Error = 1.96 \times 0.00066$ = 0.001296
95% Confidence Interval: (-0.0013, 0.0013)

Since all the sanity checks pass, we can proceed with the rest of the analysis.

Result Analysis

Effect Size Tests

95% Confidence interval around the difference between the experiment and control groups for Evaluation metrics.

	Lower Bound	Upper Bound	Statistical Significance	Practical Significance
Gross Conversion	-0.0291	-0.0119	Yes	Yes
Net conversion	-0.0116	0.0018	No	No

Gross Conversion:

$$\hat{p}_{exp} = \frac{Enrollments}{Clicks} = \frac{3423}{17260} = 0.1983$$

$$\hat{p}_{control} = \frac{Enrollments}{Clicks} = \frac{3785}{17293} = 0.2188$$

$$\widehat{d} = \hat{p}_{exp} - \hat{p}_{control} = -0.02055$$

$$\hat{p}_{pool} = \frac{X_{control} + X_{exp.}}{N_{control} + N_{exp.}} = \frac{3785 + 3423}{17293 + 17260} = 0.20861$$

$$SE_{pool} = \sqrt{0.20861 \times (1 - 0.20861) \times (\frac{1}{17293} + \frac{1}{17260})} = 0.004372$$

$m = z \times Standard\ Error = 1.96 \times 0.004372$ = 0.008568
95% Confidence Interval: (-0.0291, -0.0116)

Net Conversion:

$$\hat{p}_{exp} = \frac{Enrollments}{Clicks} = \frac{1945}{17260} = 0.11268$$

$$\hat{p}_{control} = \frac{Enrollments}{Clicks} = \frac{2033}{17293} = 0.11756$$

$$\hat{d} = \hat{p}_{exp} - \hat{p}_{control} = -0.00487$$

$$\hat{p}_{pool} = \frac{X_{control} + X_{exp.}}{N_{control} + N_{exp.}} = \frac{2033 + 1945}{17293 + 17260} = 0.11512$$

$$SE_{pool} = \sqrt{0.11512 \times (1 - 0.11512) \times (\frac{1}{17293} + \frac{1}{17260})} = 0.003434$$

$m = z \times Standard\ Error = 1.96 \times 0.003434$ = 0.00673
95% Confidence Interval: (-0.0116, 0.0018)

Sign Tests

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

	p-value	Statistical Significance
Gross Conversion	0.0026	Yes
Net conversion	0.6776	No

Summary

Bonferroni correction was not used since it is critical to use the correction if a change will be launched if **any** of the metrics show a significant difference, as it is likely that one of the metrics will be falsely positive if the number of metrics is high.

In current experiment, the change will be launched only if **all** of the evaluation metrics show a significant change, so the Bonferroni correction was not used. Moreover, the number of evaluation metrics for the experiment was only two which is not very high.

The effect size tests and the sign tests for both the evaluation metrics agree.

Recommendation

I would recommend to do a follow-up experiment. The decrease in gross conversion was statistically and practically significant which agrees with one of the requirements to launch.
A decrease in net conversion below 0.75% is not acceptable for the business. The confidence interval (-0.0116, 0.0018) does include the negative of the practical significance boundary

(-0.0075), so it's possible that this number went down by an amount that would matter to the business. This would not be an acceptable risk so my suggestion is to not launch the experiment.

Follow-Up Experiment

As a follow up experiment, Udacity could do a weekly –check in with the enrolled students by sending them an e-mail and asking them to respond to a short online questionnaire. The questions could be:

1. Do you have any questions/concerns that Udacity can help answer?
2. Do you need to access the additional support offered by the nanodegree subscription?

Based on the student's responses Udacity team could reach out to the student, and hence reduce the chance that a student would drop out of the program due to frustration.

Choice for metrics:
Unit of diversion: user-id
Since the change will be introduced after the user logs in, the unit of diversion should be the user-id.

Invariant metric: user-id
Since the unit of diversion is user-id, the unique user-ids will be randomly assigned to experiment and control group. So we should have roughly the same number of user-ids in each group and can choose it as an invariant metric.

Evaluation metric:
The metric I will measure is retention. The change should result in a statistical and practically significant increase in the retention in order to launch the change.
(Net conversion will not be a good choice as an evaluation metric as it is cookie based and less stable than metrics based on user-id.)

The hypothesis statements are:
Null Hypothesis: The new weekly check-in feature has no impact on the retention of enrolled students.
Alternative Hypothesis: The new weekly check-in feature results in a statistically significant increase in the retention.