Experiment Design

Metric Choice

The invariant metrics chosen for this test were:
- Number of cookies (The number of unique cookies to view the course overview page)
- Number of clicks (The number of unique cookies to click the "Start free trial" button)
- Click-through-probability (Number of Clicks/Number of cookies)

All three of the invariant metrics were chosen because they are recorded prior to the testing protocol – they could not have been affected by the test procedure.

The evaluation metrics chosen for this test were:
- Gross conversion (The 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
- Net Conversion (Number of user-ids to remain enrolled past the 14-day boundary (and make at least one payment) divided by number of unique cookies to click "Start free trial" button.

Gross conversion was chosen because it is first point of diversion – were some users swayed to attempt free classes rather than start the trial. If the change is working, and users with less than five hours per week are being diverted to free class rather than the free trial, we would expect this metric to decrease in our test group vs the control group.

Net conversion was chosen because it is expected to vary with the addition of the time estimate prompt. My assumption is that Net conversion will not decrease if the change is working as expected.

Number of user-ids was not used as a metric because the data is very similar to gross conversion, but more affected by the daily data collection process.

Retention was originally chosen as an evaluation metric, but the power required to use this metric was going to require a prohibitively long test time, so it was dropped.

Measuring Standard Deviation

I believe the analytical estimate of the standard deviation for gross conversion and net conversion are adequate due their unit of analysis being "unique cookies to click "Start free trial" button". So both are unit of analysis and unit of diversion are cookies and it will not be necessary to calculate empirical estimates.

	Standard Deviation
Gross Conversion	0.0202
Net Conversion	0.0156

Sizing

Number of Samples vs. Power

I will not be using the Bonferroni correction, as the three evaluation metrics I chose are likely covariant, and the pageviews I would require would unnecessarily extend the experiment time frame. Sans Bonferroni correction, this experiment will require 685,325 pageviews.

Duration vs. Exposure

For this experiment I will divert 50% of Udacity's traffic for a duration of 35 days. This experiment time frame will allow us to view potential Simpson effects while maintaining a compact experimental period.

There is no inherent risk to this test, as no sensitive data is being dealt with, and the chance of someone being injured is not an issue. 100% of the traffic could be safely diverted without issue.

Experiment Analysis

Sanity Checks

Below is a table of the calculated statistics from the sanity check. Observed values fall within the confidence intervals for both gross and net conversion, and the check thus passes.

	Pageviews	Clicks
probability	0.5	0.5
SE	0.0006019	0.002100
ME	0.0011797	0.004116
Lower CI bound	0.4988	0.4959
Upper CI bound	0.5012	0.5041
Observed	0.5007	0.5005

Result Analysis

Effect Size Tests

Below is a table of the experimental results and as series of statistics calculated for each of our evaluation metrics. The confidence interval for gross conversion indicates our result is both statistically and practically significant. The experimental value for net conversion is neither statistically or practically significant.

Result Analysis					
	Clicks	Enrollments	Gross conversion	Payments	Net conversion
Control	17293	3785	0.21887	2033	0.11756
Experiment	17260	3423	0.19832	1945	0.11269
p			0.20860		0.11513
d			-0.02055		-0.00487

Empirical SE			0.02020	0.01560
SE			0.00437	0.00343
m			0.00857	0.00673
Upper CI			-0.01199	0.00186
Lower CI			-0.02912	-0.01160

Sign Tests

Next, a sign test was conducted and a p value calculated for each of our evaluation metrics.

	p value
Gross Conversions	0.0026
Net Conversions	0.6776

These p values indicate the change observed in gross conversions is statistically significant, and the change seen in net conversions is not.

Summary

I did not use the Bonferroni correction in this experiment to keep the required pageviews down. No discrepancies arose between the effect size hypothesis tests and sign tests.

Recommendation

I recommend that Udacity not move forward to implement this change. From this analysis it does not appear to effect net conversions (and hence revenue) in a statistically significant way, while it does significantly reduce gross conversions. However, the confidence interval around Net Conversion does include the negative of the practical significance boundary, and may thus have gone down in a way that is practically significant to the business.

Follow-Up Experiment

I believe that the original experiment was conducted along the correct line of thinking, and I believe further iterations on the pop up would be useful. I suggest in addition to the five hour per week popup, any easy to use weekly calendar is included. Students can be asked to fill in which five hours per week they plan to commit to the program.

Setup: Same as above, with the addition of the above mentioned calendar.

Null hypothesis: Planning one week of time for classwork does not reduce gross conversion rate.

Unit of diversion: Number of clicks (The number of unique cookies to click the "Start free trial" button)

Invariant Metrics, the same as above:

- Number of cookies (The number of unique cookies to view the course overview page)
- Number of clicks (The number of unique cookies to click the "Start free trial" button)
- Click-through-probability (Number of Clicks/Number of cookies)

All three of the invariant metrics were chosen because they are recorded prior to the testing protocol – they could not have been affected by the test procedure.

Evaluation metrics, the same as above:

- Gross conversion (The 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
- Net Conversion (Number of user-ids to remain enrolled past the 14-day boundary (and make at least one payment) divided by number of unique cookies to click "Start free trial" button.

Gross conversion was chosen because it is first point of diversion – were some users swayed to attempt free classes rather than start the trial. If the change is working, and users with less than five hours per week are being diverted to free class rather than the free trial, we would expect this metric to decrease in our test group vs the control group.

Net conversion was chosen because it is expected to vary with the addition of the time estimate prompt. Net conversion will not decrease if the change is working as expected.

Launch criteria: If a statistically significant reduction was seen in gross conversions, without a statistically or practically significant reduction in net conversions was observed the change would be launched.

If the above was well received, I believe it would be valuable for Udacity to implement a calendar integration system with in class time tracking to the site. I am a min/max guy (and a data nerd) and would love to see my weekly engagement in real time.

References

http://graphpad.com/quickcalcs/binomial2/
http://stattrek.com/estimation/confidence-interval.aspx