Experiment Design

Metric Choice

1. INVARIANT
    a. # of cookies

    The number of unique cookies to view the course overview page is considered invariant as the course overview page is not subject to the experimental condition and is therefore not likely to be influenced by it.

    b. # of clicks

    The number of unique cookies that click on the Start Free Trial button is considered invariant as like the course overview page, these clicks sit prior to the experimental condition and are not likely to be influenced by it. As the uniqueness of cookies is determined by day, it is possible that a user revisits the page and exhibits different behavior after the first time through.

    c. Click through probability

    The CTP is a combination of the above two terms and thus represents a ratio of our two invariants.

2. EVALUATION
    a. Gross conversion

    The number of user-ids that complete checkout, enrolling in the free trial divided by the number of unique cookies that click on the Start Free Trial button is an evaluation metric because the experimental design is set-up to explicitly measure the effectiveness of an intervention after the Start Free Trial button is clicked but before signup. This immediate signup is important to measure as the intervention may yield immediate effects but no long term effects.

    b. Net conversion

    Net conversion serves as a combination of gross conversion and retention. The number of user-ids that remain after the 14 day boundary are divided by the total number of unique cookies that click on the Start Free Trial button.

3. EXCLUDED
    a. Retention

    Retention was initially included but conferred too high a time cost for running the experiment with enough power to determine an effect. The maximum amount of time for the experiment I would like is roughly 30 days. However including retention would require over one year of experiment time.

    b. # of user-ids

    Looking at the number of user-ids allows for a clear understanding of the effect of the experimental condition. The number of user-ids in the experimental condition is hypothesized to be less than the number of user-ids in the control condition. However, this effect is also captured by the gross conversion metric which has the advantage of

being normalized in the same fashion as the net conversion metric. As such, the number of user-ids is excluded from the analysis.

I will be looking specifically for significant deviations in the invariant metrics to make sure that these metrics are truly invariant. Otherwise I will need to reevaluate the experimental design and/or the metrics chosen. Specifically I will be looking at the number of unique cookies to click the Start Free Trial button as this could yield unreliable data if a user is exposed to the experimental condition and a day or more later revisits the site. A user that is exposed to the experimental condition may initially decline sign up and reconsider later rather than declining and not visiting again. This may confound some of the results but I will keep an eye out for any unusual patterns.

The experiment is designed to test whether clearer expectations in time cost for completing the course will reduce the number of students who left the free trial without significantly reducing the number of students to continue past the free trial and complete the course. The gross conversion metric will capture any effect the experimental condition may have on the number of students leaving the free trial while the net conversion metric will capture any effect the experimental condition may have on the number of students paying for the course after the free trial expires. In order to launch the experiment, the gross conversion metric must result in a significant reduction in the rate of sign-ups in the experimental condition, and the net conversion must show no significant difference between conditions.

Measuring Standard Deviation

EVALUATION METRICS
Gross Conversion   SD: .0202
Net Conversion   SD: .0156

Since the unit of diversion is a cookie in this experiment, the analytic and empirical variance will be comparable for both gross conversion and net conversion as their units of analysis are cookies as well. Given a mismatch, the empirical variability will not be comparable and will be much larger than the analytic variability.

Sizing
Number of Samples vs. Power

Given I will be looking at both effects to determine whether or not to proceed with the experiment, I will not use the Bonferroni correction. Bonferroni correction is helpful when looking for any of a number of hypotheses to be supported such that the resulting Type I error of rejecting the null hypothesis by chance can be reduced. In this case, since both gross

conversion and net conversion are required to launch the experiment, the Type II error is increased. The Bonferroni correction is not appropriate for reducing the likelihood of failing to reject the null hypothesis. The number of pageviews I will need to power the experiment appropriately is 679300.

Duration vs. Exposure

I would divert the entirety of Udacity's traffic to this experiment. In doing so, in order to achieve the required number of pageviews, the duration of the experiment must be at least 17 days. There is little chance that the experiment could harm any participant in the experiment as the time cost is low and the information provided merely introduces an added piece of information about the nature of the course. Additionally, no sensitive information is required as part of the experiment. Given the potential upside of cleaning enrollment figures to those users who are likely to remain in the course and offer payments, it is a valuable experiment that can aid revenue analysts greatly.

Experiment Analysis

Sanity Checks

CONFIDENCE INTERVALS for invariant metrics

Number of Cookies	Lower: .4988	Upper: .5012	Observed: .5006	PASSED
Number of Clicks	Lower: .4959	Upper: .5041	Observed: .5005	PASSED
CTP	Lower: -0.0013	Upper: .0013	Observed: 0	PASSED

Result Analysis

Effect Size Tests

CONFIDENCE INTERVALS for evaluation metrics

Gross conversion	Lower: -.0291	Upper: -.0120	Diff: -.0206
Net conversion	Lower: -.0116	Upper: .0019	Diff: -.0049

The gross conversion group difference is both statistically significant as the confidence interval does not include 0. Gross conversion's d_{min} is 0.01 and this value is outside of the confidence interval yielding practical significance to the result as well. Net conversion is not statistically significant as its confidence interval includes zero. However, net conversion's d_{min} is 0.0075 and the negative of this is included in the confidence interval.

Sign Tests

Gross conversion	p-value: .0026
Net conversion	p-value: .6776

Gross conversion showed a $p < .05$ statistical significance through the sign test. Net conversion was not statistically significant ($p = .6776$).

Summary

The Bonferroni correction was not used as the two evaluation metrics in this experiment were relevant to the decision to continue the experiment. Relying on the effects of both metrics in this case increases the Type II error while the Bonferroni correction is used in reducing the Type I error.

No discrepancies between effect size and sign test results are observed.

Recommendation

In order to launch, as stated above, the experiment must show that the number of students enrolled in the free trial are reduced in the experimental condition while the number of students that continue past the free trial must remain statistically unchanged. In the data I find that the gross conversion rate is significantly reduced in the experimental condition while the net conversion rate is not statistically different across conditions. However, the negative of net conversion's practical significance boundary is included which indicates that it is possible that the number of net conversions decreased by an amount that matters to the business albeit statistically insignificant. Because of net conversions possibility of a decrease in the experimental condition, the experiment needs further testing before launch.

Follow-Up Experiment

The previous experiment takes advantage of framing on user behavior. Namely, positioning the paid version of the course as something that takes at least 5 hours a week disincentivizes those who are not as committed but willing to go through a free trial and does not dissuade those who are up for the challenge and know what is expected of them. However, this particular intervention ignores whether or not a user has completed a course in the past. My assumption is that past courses completed successfully would serve as a predictor for the probability of current course completion. Further, if a user is asked whether or not they have completed a course in the past, an intervention can be uniquely targeted to the group that is presumably in need of expectation setting. A user acknowledging that they have not taken a course may be put in a particular frame such that new information about how to successfully complete a course will become more salient.

The proposed experiment will continue with the placement of the current intervention. However, upon clicking the Start Free Trial button, users will be asked whether they have successfully completed a Udacity course. Answering 'yes' prompts a pleasant greeting such as 'Welcome back!' and the opportunity to enroll. Answering 'no' prompts a notice that successful completion of the course requires greater than 5 hours a week and subsequently advises the user to proceed only when willing to make this time commitment. Again this is followed by the opportunity to enroll. My hypothesis is that users in the experimental condition that have not completed a course and are presented with the above prompts will enroll at a reduced rate compared to the control group. The net conversion rate of the experimental group is hypothesized not to be significantly different from the control group. Contrary to the results of the above experiment, I also hypothesize that given this tweaked experimental paradigm, the net conversion effect size confidence interval will not include the practical significance boundary.

Much like the above experiment, the unit of analysis is the number of unique cookies that click the Start Free Trial button. As this experiment evaluates the effect of an intervention before a user-id is captured, another measure must be used. A unique cookie allows the closest proxy. The experiment can then track the gross conversion and net conversion and test in the same manner as the previous experiment. The unit of diversion is also unique cookies that click the Start Free Trial button for the same reason. A user-id is not necessarily available at this stage especially for those that have not taken a course before. An event unit of diversion is not appropriate either as this particular experiment is extremely visible to the user. The evaluation metrics can remain gross and net conversion, the results of which are expected to trend in the same manner as the above experiment. A difference test can then be performed between the two experiments identifying which experimental condition produced the most significant drop in enrollments while not significantly changing payments.