Experiment Overview: Free Trial Screener

Background information: Udacity courses had two options on the course overview 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.

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.

Metric Choice

The following Invariant & Evaluation metrics were chosen

Invariant Metrics

Invariant metrics were chosen on their characteristic that they should not vary or change between the control and experiment.

1. **Number of Cookies:** This is the number of unique cookies to view the course overview page. This is the unit of diversion and should be evenly distributed between the control and the experiment groups. Hence, this would serve the purpose of being an invariant metric

2. **Number of Clicks:** The number of unique cookies to click the "Start free trial" button. This metric should not change between control & experiment groups because until this point, no change has been implemented for the users. Thus it should be an invariant metric

3. **Click-through-probability:** The number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page. This metric should not change between control & experiment groups because until this point, no change has been implemented for the users. Thus it should be an invariant metric

Evaluation Metrics

Evaluation metrics were chosen as they may be affected by the experiment, and hence show a different value between the experiment and control groups.

1. **Gross Conversion:** This is 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. This metric may be affected by the experiment where the user is additionally warned about the 5 hours' time requirement, and is hence an evaluation metric. **To launch the experiment, this metric should show a statistically & practically significant reduction.**

2. **Retention:** This is the 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 metric may be affected by the experiment where the user is additionally warned about the 5 hours' time requirement, and is hence an evaluation metric. . **To launch the experiment, this metric should not show any statistically or practically significant change or show a statistically & practically significant improvement.**

3. **Net Conversion:** This is the 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. This metric may be affected by the experiment where the user is additionally warned about the 5 hours' time requirement, and is hence an evaluation metric. **To launch the experiment, this metric should not show any statistically or practically significant change or show a statistically & practically significant improvement.**

Neither Invariant nor evaluation

Number of user-ids: The number of users to enroll in the free trial.

1. This was not chosen as an invariant metric as enrollment might be affected by the “5 hours’ warning”, hence could have different values between the control and experiment group.

2. This was not chosen as an evaluation metric, as this an absolute number and is not capturing the number of clicks on the “Start Trial”. Thus it could vary also with the overall amount of the traffic on the site on that day and not just because of the experiment. Hence, I did not choose this metric for evaluation of the experiment.

Measuring Standard Deviation

The analytical standard deviation is calculated as std = sqrt(p * (1-p) / N).
For the analytical estimates of standard deviation to match the empirical standard deviation, the unit of analysis should be the same as the unit of diversion.

1. **Gross conversion**

- Baseline probability for gross conversion is p = 0.20625
- N, number of users who see the "start free trial" page = 5000 * 0.08 = 400
- Thus, standard deviation is std = sqrt(p * (1-p) / N) = 0.0202

The unit of analysis and the unit of diversion can be assumed to be the same (Unique cookies). Hence empirical estimate should closely match the analytical estimate.

2. **Retention**

- Baseline retention rate is p = 0.53
- N, number of users who enrolled in the free trial = 5000 * 0.08 * 0.20625 = 82.5
- Thus, standard deviation is std = sqrt(p * (1-p) / N) = 0.0549

For retention, as the units of diversion and unit of analysis are different (user-id & cookies), the analytical standard deviation and the empirical standard deviation may not closely match.

3. **Net conversion**

- Baseline net conversion rate is p = 0.1093125
- N, number of users who see the "start free trial" page = 5000 * 0.08 = 400
- Thus, standard deviation is std = sqrt(p * (1-p) / N) = 0.0156

The unit of analysis and the unit of diversion can be assumed to be the same (Unique cookies). Hence empirical estimate should closely match the analytical estimate

Number of Samples vs. Power

The Bonferroni correction was not used as it would be a very conservative approach, given that the metrics are not independent.

- The online calculator (http://www.evanmiller.org/ab-testing/sample-size.html) was used to find the results with alpha = 0.05 & beta = 0.8
- The baseline conversion rate and minimum detectable effect (d_min) are listed individually in the table below
- The sample size would need to be multiplied by two as there are two groups, control & experiment
- Gross & net conversion sample size would need to be divided by 0.08 (clicks/pageview)
- Retention sample size would need to be divided .0165 (enrollments/pageview)

Sn	Metric	Baseline conversion rate	d_min	Sample size	Conversion to Page view	Page Views
1	Gross conversion	0.20625	0.01	25835	25835 / 0.08 * 2	645875
2	Retention	0.53	0.01	39115	39115 / 0.08 / 0.20625 * 2	4741212
3	Net conversion	0.1093125	0.0075	27413	27413 / 0.08 * 2	685325

Taking the maximum of the pageview values, it would need 4,741,212 page views. Assuming 100% diversion and 40,000 page views per day, the experiment would take ~ 119 days. This is a

rather long time period, and it may not be feasible or desired. Hence, retention can be dropped at his stage, which significantly reduces the required page views to 685,325.

Duration vs. Exposure

Assuming 100% of the traffic is diverted to this experiment with 40,000 page views per day, the length of the experiment would be 685325 / 40000 = 17.13, which would effectively be 18 days.

100% of the traffic was chosen to be diverted as

- The experiment does not represent a business risk as there is no major change in the user experience
- Experimental message can be ignored; hence this experiment does not forcibly impose any decision on the user
- The experiment does not have the potential to harm anyone
- The experiment does not does not collect any sensitive information

Experiment Analysis

Sanity Checks

1. **Number of cookies:** This was modeled as a Bernoulli distribution with 0.5 probability

- control group total = 345543
- experiment group total = 344660
- standard deviation = sqrt(0.5 * 0.5 / (345543 + 344660)) = 0.0006018
- margin of error = 1.96 * 0.0006018 = 0.0011796
- lower bound = 0.5 - 0.0011797 = 0.4988
- upper bound = 0.5 + 0.0011797 = 0.5012
- observed = 345543 / (345543 + 344660) = 0.5006

The observed value is within the upper & lower bounds, and **thus the sanity check is passed**

2. **Number of clicks on "start free trial":** This was modeled as a Bernoulli distribution with 0.5 probability

- control group total = 28378

- experiment group total = 28325
- standard deviation = sqrt(0.5 * 0.5 / (28378 + 28325)) = 0.0021
- margin of error = 1.96 * 0.0021 = 0.0041
- lower bound = 0.5 - 0.0041 = 0.4959
- upper bound = 0.5 + 0.0041 = 0.5041
- observed = 28378 / (28378 + 28325) = 0.5005

The observed value is within the upper & lower bounds, and **thus the sanity check is passed**

3. **Click-through-probability on "start free trial":** The standard deviation for the experiment group was found using the probability for the control group (p_cnt) and the sample size,

- standard deviation = sqrt(p_cnt * (1 - p_cnt) / N_exp)
- p_cnt= 0.0821258
- standard deviation = sqrt(0.0821258 * (1-0.0821258) / 344660) = 0.000468
- margin of error = 1.96 * 0.000468 = 0.00092
- lower bound = 0.0821258 - 0.00092 = 0.0812
- upper bound = 0.0821258 + 0.00092 = 0.0830
- experiment value = 0.0821824

The observed value is within the upper & lower bounds, and **thus the sanity check is passed**

Result Analysis

Effect Size Tests

Formulae Definition: The formulae below were used in the calculations

Pooled probability (p_pooled) & pooled standard error were calculated as below

- p_pooled = (X_cnt + X_exp) / (N_cnt + N_exp)
- se_pooled = sqrt(p_pooled * (1-p_pooled) * (1./N_cnt + 1./N_exp))

where

- X_cnt= Control Target samples
- X_exp = Experiment Target Samples
- N_cnt=Total control samples

- N_exp=Total experiment samples
- The probability difference(d) = X_exp / N_exp - X_cnt / N_cnt
- lower bound= d - se_pooled
- upper bound = d + se_pooled

1. **Gross conversion:** Total samples are the clicks for "start free trial" & target samples are enrolled users

- N_cnt = clicks_controlled = 17293
- X_cnt = enroll_controlled = 3785.
- N_exp = clicks_experiment = 17260.
- X_exp = enroll_experiment = 3423
- p_pooled = (X_cnt + X_exp) / (N_cnt + N_exp) = 0.2086
- se_pooled = sqrt(p_pooled * (1-p_pooled) * (1./N_cnt + 1./N_exp)) = 0.00437
- d = X_exp / N_exp - X_cnt / N_cnt = -0.02055
- lower bound= d - se_pooled = -0.0291
- upper bound = d - se_pooled = -0.0120

- As 0 does not lie between the lower bound & upper bound, the **metric is statistically significant.**
- As positive or negative d_min (0.01 or -0.01) does not lie between the lower bound & upper bound, **this metric is also practically significant**.

2. **Net conversion:** Total samples are the clicks of "start free trial" & target samples (numerator) are paid users.

- N_cnt = clicks_controlled = 17293
- X_cnt = pay_controlled = 2033
- N_exp = enroll_experiment = 17260.
- X_exp = pay_experiment = 1945
- p_pooled = (X_cnt + X_exp) / (N_cnt + N_exp) = 0.1151
- se_pooled = sqrt(p_pooled * (1-p_pooled) * (1./N_cnt + 1./N_exp)) = 0.00343
- d = X_exp / N_exp - X_cnt / N_cnt = -0.0048
- lower bound= d - se_pooled = -0.0116
- upper bound = d + se_pooled = 0.0019

- As 0 lies between the lower bound & upper bound, **the metric is not statistically significant.**

- As negative d_min (-0.0075) lies between the lower bound & upper bound, **the metric is also not practically significant**

Sign Tests

The online calculator (https://www.graphpad.com/quickcalcs/binomial1.cfm) was used to perform sign test.

1. **Gross conversion**
    - Number of days with reduction in experiment group: 19
    - Total Days of experiment: 23
    - Probability 0.5
    - p-value 0.0026,

With alpha = 0.05, **the change is statistically significant**

2. **Net conversion**
    - Number of days with improvement in experiment group: 10
    - Total days of experiment: 23
    - Probability 0.5
    - p-value 0.6776,

With alpha = 0.05, **the change is not statistically significant**

Summary

The Bonferonni correction was not used this analysis, as it is a method to limit errors in multiple independent metrics comparisons. In this case, the two metrics, gross conversion and net conversion are not independent. Also, since we need all the metrics to move in the desired direction (instead of any one) before we launch the change, using Bonferonni correction would be an extremely conservative approach and not suitable in this case.

The effective size hypothesis tests and sign tests both showed that the change will reduce gross conversion in a statistically significant manner, but not affect the net conversion rate in a statistically significant manner. Hence there was no discrepancy amongst their result.

Recommendation

The experiment was successful in reducing the number of frustrated students who left the free trial, as indicated by the gross conversion metric results.

The net conversion metric also showed that the experiment had no statistically significant impact on the ratio of the number of students who completed the course to those who clicked on the "start free trial" button".

However "net conversion", though not statistically significant, **showed a high likelihood of a negative impact from the experiment (**-0.0116 to 0.0019**). This negative impact, if true, will directly affect the number of paying students and hence the revenue of Udacity**.

Given the risk to Revenues, **I would recommend against launching the change and suggest further experiments to study this effect in greater detail**.

Follow-Up Experiment

A useful follow-up experiment could be to track the time spent by new users to complete the first few lessons. Based on historical data of time spent on these lessons by students who successfully graduated, the new users can be classified as either "Acceptable" or "At risk of cancelling" depending on the difference between their time spent Vs historical time spent by successful students.

The users who are 'At risk of cancelling" could be then shown a message where additional resources (e.g. forum links/detailed explanation videos /tutoring options) could be offered.

Some high level characteristics of the experiment could be

- Hypothesis: Students who struggle/cannot find time/are de-motivated in the beginning of the trial can be persuaded to continue the course (and make a payment) by offering them extra support resources like forum links/detailed explanation videos/tutoring options.

- Unit of diversion: User-id for newly enrolled students, as this study only concerns this group. They could then be further divided into control and experiment

- Invariant Metric: Number of Newly enrolled User IDs only. The number of newly enrolled User IDs should not change across experimental or control groups as it is the unit of diversion. Hence this could be used as an invariant metric

- Evaluation Metric: Ratio of users who continued past the free trial (hence paid) to users who enrolled for the free trial . This ratio should be directly affected by the experiment, persuading users who are at risk of cancelling to continue the course and hence make a payment.

A statistically & practically significant increase in the evaluation metric (ratio of users who continued past the free trial to users who enrolled for the free trial) should lead to launching this change

References

1. https://www.udacity.com/
   Experiment data, tutorials and templates
2. https://www.graphpad.com/quickcalcs/binomial1.cfm
   Online Sign test calculator
3. http://www.evanmiller.org/ab-testing/sample-size.html
   Sample size calculator
4. https://en.wikipedia.org/wiki/Sign_test
   Sign Test Details
5. https://en.wikipedia.org/wiki/A/B_testing
   A B tests information