

Experiment Design

Metric Choice

Invariant Metrics

Number of cookies

Number of clicks

Click-through-probability

Evaluation Metrics

Gross conversion

Net conversion

Unused

Number of user-ids

Retention

Invariant metrics help sanity check our experiment, so we expect their values to be the same between control and experiment groups. Viewing the course overview page and clicking the “Start free trial” button occur *before* we show the experimental free trial screener, we expect the three metrics listed above to be roughly the same between both groups.

Evaluation metrics help us decide if the change was successful or not. As a result, they should differ between groups and be useful in seeing the effect of our change. Our hypothesis has two parts: 1) The free trial screener sets clearer expectations and as a result, some students will self-select out, and 2) The students who take the free trial will have a better experience, because of clearer expectations and increased coaching availability, and will thus finish the course.

For (1), gross conversion measures the number of enrollments divided by the number of “Start free trial” clicks, this is a good evaluation metric.

For (2), we’d ideally want to choose a metric that tracks whether or not a student has completed the course, however, this would probably take too long. Tracking whether or not a student remains enrolled past the 14-day boundary is a good proxy, which net conversion tracks. For the unused metrics, number of user-ids is both a bad invariant and evaluation metric. Since we expect the screener to reduce enrollments, this will be a poor invariant metric. As an evaluation metric, the number of user-ids will differ between groups, but the proportion of users who click “Start free trial” and then enroll is more relevant than the actual number. This is better tracked by gross conversion.

In order to launch the experiment, I would expect to a decrease in gross conversion while net conversion to stay roughly the same.

Measuring Standard Deviation

Metric	Standard Deviation
Gross conversion	0.0202
Net conversion	0.0156

The unit of diversion is the same as the unit of analysis, so I expect the analytic and empirical estimates to be roughly the same.

Sizing

Number of Samples vs. Power

I will not use the Bonferroni correction in the analysis phase. I will need 685326 pageviews.

Duration vs. Exposure

The change is a simple UI change that does not deal with sensitive data. The experiment is not risky, so we can run with 100% of our traffic, requiring 18 days to run the experiment.

Experiment Analysis

Sanity Checks

Metric	Lower bound	Upper bound	Observed	Passes
Number of cookies	0.4988	0.5012	0.5006	Yes
Number of clicks on “Start free trial”	0.4959	0.5041	0.5005	Yes
CTR on “Start free trial”	0.4988	0.0013	0.0001	Yes

Result Analysis

Metric	Lower bound	Upper bound	Statistical	Practical
--------	-------------	-------------	-------------	-----------

			significance	significance?
Gross conversion	-0.02912	-0.01199	Yes	Yes
Net conversion	-0.0116	0.00186	No	No

Sign Tests

Metric	p-value	Statistical significance
Gross conversion	0.0026	Yes
Net conversion	0.6776	No

Summary

I did not use the Bonferroni correction, since we need to consider both metrics in order to launch. Bonferroni is applicable when there are many metrics and only one of them needs to match our expectations in order to launch.

There do not appear to be any discrepancies between the sign tests and the hypothesis tests.

Recommendation

Make a recommendation and briefly describe your reasoning.

Finally, make a recommendation. Would you launch this experiment, not launch it, dig deeper, run a follow-up experiment, or is it a judgment call? If you would dig deeper, explain what area you would investigate. If you would run follow-up experiments, briefly describe that experiment. If it is a judgment call, explain what factors would be relevant to the decision.

I would not launch the experiment and would dig deeper into why students are cancelling early in the course.

The data suggest that the screener does indeed reduce enrollments. However, we see a *reduction* in net conversion, although the change is statistically insignificant. The confidence interval includes the practical significance boundary, so it's possible that there has been a drop in net conversion that we care about. This suggests that we don't know if the screener improves the experience for the students who enroll.

To dig deeper, I'd look at the response data for the questionnaire students fill out when they cancel early. The questionnaire gives the following choices as reasons for cancelling early:

- Not able to dedicate enough time to this program
- Not happy with quality of content
- Not happy with level of support received
- Content is not rigorous enough
- Don't have required background for course
- Don't feel program was worth cost
- Other

I'd like to look at if students particularly cite not being able to dedicate enough time to this program or not being happy with the level of support received. Given the data we have so far, I'd expect the responses for these to be similar between the experiment and control groups. If that's true, we need explore how we can make the student experience better.

Follow-Up Experiment

Give a high-level description of the follow up experiment you would run, what your hypothesis would be, what metrics you would want to measure, what your unit of diversion would be, and your reasoning for these choices.

List ideas:

- Make one-on-one appointments button more prominent

I would try making the one-on-one feature more prominent, because I believe students aren't aware of the support they can get during the course.

Problem

When I enrolled up for the Data Analyst Nanodegree, I was vaguely aware that there were coaching resources. However, it took me a while to find in the top right to actually sign up for a one-on-one appointment.

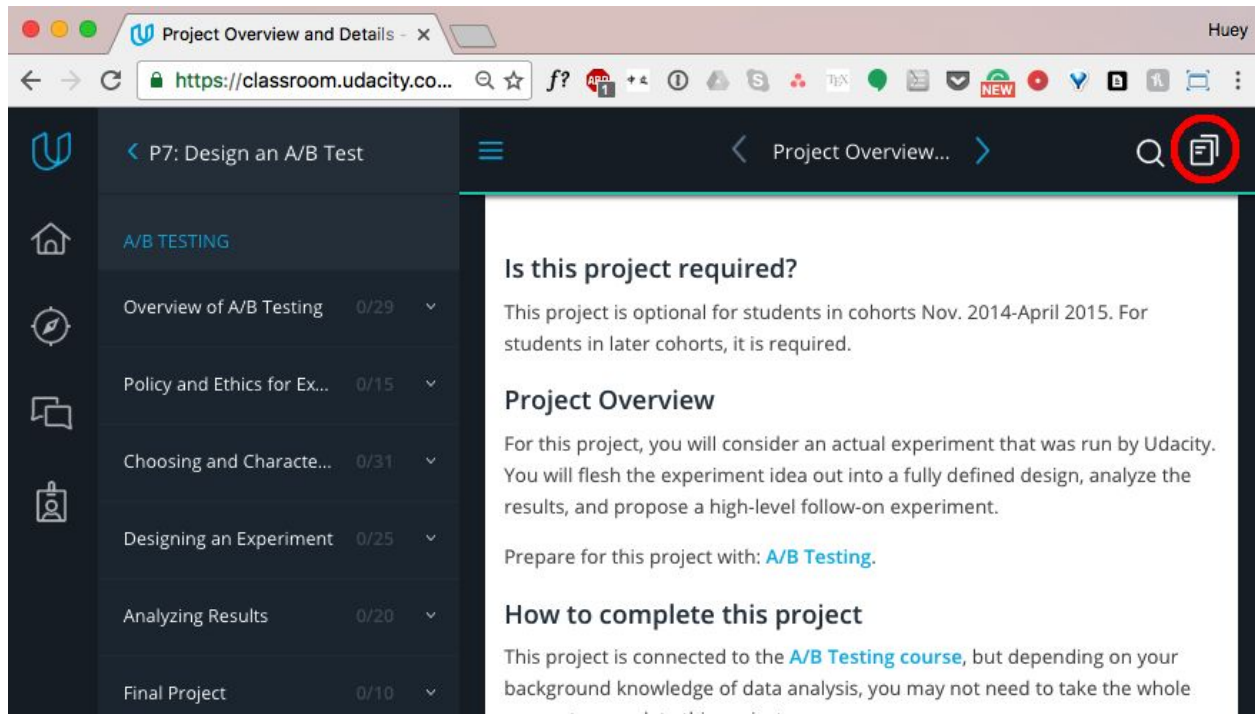


Figure 1. Resources button in the Project Classroom.

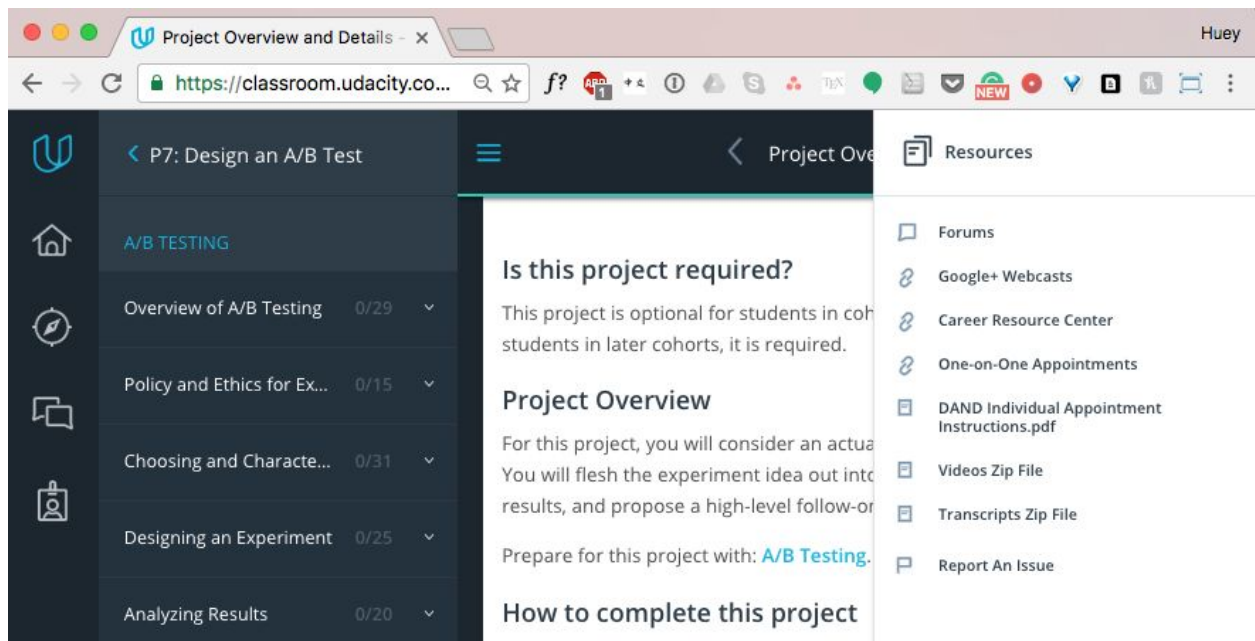


Figure 2. List of Resources, including the One-on-One Appointments Link.

My hypothesis is that students aren't aware of the help they can get when taking a course, and that if they were more aware, they would be more likely to use help and continue the course.¹

¹ I am assuming that a major reason students cancel is because they are unhappy with the level of support received. The only data I have to support this assumption is that the first experiment assumed that increased

UI Change

I would add a button in the left navbar that would allow students to schedule a One-on-One appointment.

Unit of Diversion

The unit of diversion would be user-ids and would divert users into groups upon enrollment in the course.

Metrics

I would use number of students enrolled in a course as an invariant metric. Since the experiment begins after enrollment, we should expect the number of students enrolled in courses to be the same between groups.

I would use retention as an evaluation metric, i.e. the number of user-ids enrolled past the 14-day boundary divided by number of user-ids to complete checkout.

coaching capacity would help students complete the course, and that this reason appears in the questionnaire. Before running this experiment, I'd look through the questionnaire response data to validate this assumption.