P7 - Design an A/B Test

Student: Peter Carsten Petersen

Experiment Design

Metric Choice

Invariant Metric	**Reason**
Number of Cookies	Given that the test we are analyzing takes place after a click, we are looking for Invariant Metrics which take place before the click, since these should theoretically be comparable between Control and Experiment, if all else is equal. These 3 specific Metrics are all taking place before Click and therefore ideal as Invariant metrics.
Number of Clicks	
Click Through Probability	

Evaluation Metric	**Reason**
Gross Conversion **(dmin = 0.01)**	Given that our test is to measure the impact on **Enrollments** in **Free Trial** as well as impact on **Continued enrollment and payment** beyond **Free Trial**, from adding a layer between Click and Enrollment, the Gross and Net Conversion on clicks on “start free trial” button are the logical choices as Evaluation metrics, as these are directly influenced by the test, and can be measured between Control group (not experiencing the change) and Experiment group (experiencing the change).
Net Conversion **(dmin = 0.0075)**	

Experiment Launch: I would expect the result of this experiment, to show a negative effect on the **Gross Conversion,** as some potential students will be discouraged by the prompt for considering necessary time needed to complete the course.

For **Net Conversion** I would also expect some negative effect, but to a lesser extent than **Gross Conversion** as the pool of discouraged students from **Gross Conversion** were already likely candidates to leave before end of Free Trial.

In order to recommend a launch of the change post experiment, I will be looking for no more change in Gross Conversion, than a 95% confidence interval centered around the dmin of 0.01 and a Net Conversion with no change or a change less than dmin of 0.0075.

Metrics excluded from Analysis	**Reason**
Number of User Id's	Could potentially have been chosen as Invariant, but User Id's are only registered if they enroll, omitted if they do not enroll, i.e. data is incomplete.
Retention	Could potentially have been chosen as Evaluation, but since this is only calculated on User Id's it has same shortcomings as "Number of User Id's".

Measuring Standard Deviation

Evaluation Metric	**Standard Deviation**
Gross Conversion	0.0202
Net Conversion	0.0156

Analytical vs. Empirical: I would definitely expect both Metrics too have higher Empirical variance than Analytical, and more so Gross Conversion than Net Conversion.

Gross Conversion can be influenced by a lot of variables, such as time of year (e.g. in Winter more likely to start new online study than when warmer), quality and amount of competition to same type of course, word of mouth or positive mention in media are all examples of underlying variables which can have influenced traffic and Gross conversion rates and variability empirically.

For **Net Conversion** I would expect less difference between Analytical and Empirical, unless the initial course material has been changed many times in which case these changes can also have resulted in increased empirical variability.

Sizing

Number of Samples vs. Power

Description	Use method Yes/No	Reason
Bonferroni Method	No	Bonferroni Method not used as Evaluation Metrics are expected to be correlated, both being conversion rates on Clicks, and consequently using Bonferroni would lead to an overly conservative result, and possibly pointing towards a “Non launch” conclusion where “Launch” would be beneficial.

Evaluation Metric	Parameters	Pageviews Needed for analysis
Gross Conversion	prob. 20.63% / α = 0.05 / β = 0.2 / dmin = 1%	647,325
Net Conversion	prob. 10.93% / α = 0.05 / β = 0.2 / dmin = 0.75%	685,275 ✓

For pageviews needed in the A/B test, we must choose the highest number calculated on each of the evaluation metrics. I.e. 685,275 pageviews will be needed.

Duration vs. Exposure

Pageviews Needed	Fraction of traffic diverted	Length of experiment in Days
685,275	0.50	35

Reason and Risk: The 0.5 fraction of traffic is chosen in order to run the experiment within a reasonable amount of time, so that other factors may not influence the result. It is however not without risk to run on such a significant portion of the traffic. If this was a major change, I would not recommend this high of a fraction, as if experienced as negative by users could have a severely negative effect on the business. However given that this is “merely” an added help for potential students to consider before enrolling, i think that the fraction is within reasonable range and risk.

Experiment Analysis

Sanity Checks

Invariant Metric	Lower Bound	Upper Bound	Observed	Passed
Number of Cookies	0.4988	0.5012	0.5006	✓
Number of Clicks	0.4959	0.5041	0.5005	✓
Click Through Probability	0.0812	0.0830	0.0822	✓

All Sanity Checks on Invariant Metrics are passed, and point towards a successfully run test.

Result Analysis

Effect Size Tests

Evaluation Metric	Lower Bound	Upper Bound	dmin	Statistical Significant	Practical Significant
Gross Conversion	-0.0291	-0.0120	+/- 0.01	✓	✓
Net Conversion	-0.0116	0.0019	+/-0.0075	%	%

Sign Tests

Evaluation Metric	Days where experiment > control	Total Days	two-tailed p value	Statistical Significant
Gross Conversion	4	23	0.0026	✓
Net Conversion	10	23	0.6776	%

Summary

Bonferroni Method not used as Evaluation Metrics are expected to be correlated, both being conversion rates on Clicks, and consequently using Bonferroni would lead to an overly conservative result, and possibly pointing towards a “Non launch” conclusion where “Launch” would be beneficial.

No discrepancies between result of Effect Size and Sign tests.

Recommendation

Recap of goal of the tested change: *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.*

Based on the A/B Test result I would make a recommendation **not to implement this change.**

The **Gross Conversion** is showing a both statistical and Practical Significant change in the negative direction with a confidence interval between -0.012 and -0.029, this is perhaps not surprising given the nature of the change which is actively discouraging some potential students from enrolling, and based on this result alone the recommendation would be to **not change since it is above the dmin of -0.01.**

I am conscious to the fact that the goal of not significantly reducing students continuing past the free trial is being met, as **Net Conversion** is not showing any sign of change being both statistically and practically insignificant, but the size of the **Gross Conversion** result makes me recommend keeping these potentially frustrated students in the free trial, and instead searching for other methods, for either encouraging to stay in course or alternatively other ways of discouraging but with a lower rate.

Finally it is noted that the Data Material provided did not allow for a complete test result, as only 23 days of full data provided vs. a calculated need of 35 days of data. Therefore calculations should be updated with a full dataset if possible, or run again before final recommendation can be made.

Follow-Up Experiment

In order to increase the **Net Conversion** of students enrolling in Free Trial, I would recommend testing, **post Free Trial Enrollment,** an added course segment consisting of a series of small explanatory Video’s outlining for each segment of course:

1. High Level description of specific course segment and presentation of instructors.
2. Description and visualization of tools mastered post each segment completion.
3. Real life examples of course segments.
4. A maximum duration of 3-5 minutes per video.

The Hypothesis would be an increase of **Net Conversion** as students become more aware of the interesting path ahead, and the benefits of completing and acquiring new competencies.

Invariant Metrics: Will be the same as for the experiment analysed in this A/B Test, i.e. **Number of Cookies, Number of Clicks, Click Through Probability,** as this test will also take place post Click on Free Trial Button. **Gross Conversion** can also be used as Invariant for this particular test as experiment should not influence click on “Free Trial”.

Evaluation Metrics: Will in this case only be the **Net Conversion (user id’s staying on after 14 day Free trial Boundary divided by Unique cookies clicking Start Free Trial Button) ,** as we are only testing the rate of students who stay on after end of Free Trial.

dmin should be calculated before Testing, as a product of minimum uplift needed in **Net Conversion** in order to recover cost of producing video content within a reasonable timeframe. This calculated **dmin** should furthermore pass a simple feasibility consideration before possible testing is initiated.

Resources: www.wikipedia.org / stats.stackexchange.com
www.evanmiller.org/ab-testing/sample-size.html / graphpad.com/quickcalcs/binomial1/