Metric Choice

Invariant Metrics

- Number of cookies: It should not change, as the numbers of persons visiting the page is the same whether they enroll in a paid course or not.
- Number of clicks: It will not change, as it is measured before the new step is shown to the user.
- Click-through-probability: As the number of clicks, it should not change, since this step is previous to the new step being evaluated.

Variant Metrics

- Number of user-ids: Less users are expected to enroll in free trial, so number of user-ids is expected to decrease.
- Gross conversion: If the hypothesis is true, gross conversion will be lower, since users who click on free trial are discouraged to complete checkout if they can't keep the expected time commitment.
- Retention: Retention is given by number of user-ids to remain enrolled after the trial divided by number of user-ids that completes checkout. If the hypothesis is true, less users are expected to complete checkout, but the number of users to remain enrolled after free trial should not decrease, therefore increasing retention.
- Net conversion: net conversion is given by user ids that remain enrolled past 14-day boundary divided by cookies that click on "start free trial" button. If the hypothesis is true, the change will not affect the number of users who remain enrolled after the free trial, neither the number of users who click on start free trial, since the disclaimer is given after clicking on free trial. So net conversion should not decrease.

Evaluation metrics

The hypothesis is that the changes affect the application by:

1. setting clearer expectations for students upfront, reducing the number of frustated students who leave because they don't have enough time
2. without significantly reducing the number of students to continue past the free trial

We can measure the reduction of the number of frustrated students leaving the course by using the gross conversion. In the given hypothesis, the user would give up after clicking the start free trial and before completing the checkout. In this case, for the hypothesis to be true, we expect a reduction in the gross conversion, that is given by number of user-ids to complete checkout divided by number of unique cookies that click on "start free trial".

The second part of the hypothesis is true if the number of students to continue past the free trial is unaffected. We can test this using the net conversion, that is given by user ids that remain enrolled past 14-day boundary divided by cookies that click on "start free trial" button. If the

hypothesis is true, the net conversion should not decrease even though the gross conversion decreases.

Measuring Standard Deviation

- Gross Conversion: the unit of diversion is click. For 5000 unique cookies to view page, we have 400 who click. Standard deviation is given by the square root of p(1-p)/N, which equals 0.0202.
- Net Conversion: the unit of diversion is click. For 5000 unique cookies to view page, we have 400 who click. Standard deviation equals 0.0156.

Both metrics unit of analysis are unique cookies, which is the same unit of diversion used in the experiment. Hence it is likely that the the empirical computed variability will not differ from the analytical computed variability, so it is not required to compute the empirical estimate of the variability.

Sizing

The minimum number of unique cookies to view page calculated are listed below:

Metric	***sample size***	***corresponding page views***	***experiment + control***
Gross Conversion	25,839	322,988	645,975
Net Conversion	27,411	342,638	685,275

The required number of page views is given by the evaluation metric with the highest requirement of sample size. In this case, is the net conversion, that requires 685,275 page views.

The test does not seem to impose any risk to the user, as it is crafted only to discourage users who might not be committed, but does not stops the user from signing up if he wants to. There is also no chance of the user getting hurt. The data is not sensitive, hence there are no privacy issues related to the data being used in the experiment. Based on this risk evaluation, we can decide to divert 100% of the traffic to the experiment in order to expedite the completion of the A/B test.

Given that we have 40,000 page views per day, we can conclude that it would at take approximately 18 days (685,275 / 40,000) to complete the experiment.

Sanity Checks

All the sanity checks for the selected invariant metrics passes. The experiments results are within the boundaries of the confidence intervals. The results and its calculations are depicted below:

Number of cookies:

SE	0.0006
Margin of Error	0.0012
difference	0.5000
CI lower bound	0.49882
CI upper bound	0.50118
Observed	0.50064

Number of clicks on "Start Free Trial":

SE	0.0021
Margin of Error	0.0041
difference	0.5000
CI lower bound	0.49588
CI upper bound	0.50412
Observed	0.50047

Click-through probability:

control	0.08213
standard error	0.00047
margin of error	0.00092
CI lower bound	0.08121
CI upper bound	0.08304
Observed	0.08218

Check for Pratical and Statistical Significance

Bonferroni correction is not applied in this case. We have two metrics, gross conversion and net conversion, and each metric applies to one part of the hypothesis. So we need to consider the results of both metrics in order to evaluate the hypothesis. Bonferroni would be applied in cases where there are multiple metrics, but just one would need to be relevant for the hypothesis to be true.

The computed confidence intervals for the evaluation metrics and its calculations are below:

Gross Conversion

control	0.218875
experiment	0.198320
difference	-0.020555
pooled probability	0.208607
pooled standard error	0.004372
margin of error	0.008568
confidence interval - lower bound	-0.029123
confidence interval - upper bound	-0.011986

Net Conversion

control	0.117562
experiment	0.112688
difference	-0.004874
pooled probability	0.115127
pooled standard error	0.003434
margin of error	0.006731
confidence interval - lower bound	-0.011605
confidence interval - upper bound	0.001857

The results show that the gross conversion difference is statistically significant and practically significant, as its absolute upper bound is greater than 0.01. On the other hand, the net conversion difference is not statistically significant nor practically significant. However, its absolute confidence interval lower bound (0.012) is greater than the practical significance boundary (0.075), which implies a risk for launching the experiment.

Sign Tests

The table below shows the number of days in which the control or experiment group had a higher value, and the calculated two-tail p-value:

	Gross Conversion	Net Conversion
Control Higher	19	13
Experiment Higher	4	10
Two-tail p-value	0.0026	0.6776

The given alpha is 0.05. For the gross conversion, the calculated p-value is below the alpha value, confirming the results found in the evaluation test. There is only a 0,26% probability of the observed distribution to have occurred by chance.

As for the net conversion, the calculated p-value is much higher than the alpha value, which also confirms our previous findings that net conversion differences are not statistically or practically significant. The observed differences are likely to have occurred by chance.

Make a Recommendation

Gross conversion results shows that it went down by a difference greater than the practical significance boundary, as expected, what would be a positive indicator to launch the experiment.

Gross Conversion
dmin (-0.010) > upper bound (-0.012) > difference (-0.021) > lower bound (-0.029)

However, net conversion results confidence interval lower bound is lower than the practical significance boundary, meaning there is a possibility that the change reduces net conversion (we want it not to decrease).

Net Conversion
lower bound (-0.012) < dmin (-0.0075) < difference (-0.005) < upper bound (0.002)

As it stands, if there is time available, I would repeat the experiment with a larger sample before deciding whether or not to launch the change. If there is no time available to repeat the experiment, I would not launch the change. There is a risk that the company is losing customers, which is a risk that the business might not be willing to take.

Follow-Up Experiment: How to Reduce Early Cancellations

Experiment

Provide half an hour of one-to-one coaching, by video, to students who are struggling with the material or the platform at the first month of the course. The goal of the coaching is to clarify doubts, give the student an overall perspective of the subject, show common pitfalls, and give tips and motivation for the student to succeed.

Hypothesis

The new feature would decrease the number of users who cancelled the enrollment in the first month.

Metrics

For evaluation metric, we can use an extended retention, that is given by the number of users that remain enrolled after the first month divided by the number of users that completes check-out and enroll in free trial.

As an invariant metric we can use number of user-ids (number of users who enroll in free trial). Both net conversion and retention can vary, since the coaching may occur at any time after the student enrolls, so they are not suitable to be invariant metrics.

Unit of Diversion

The ideal unit of diversion is user-id, since we are tracking each user individually. It would also be advisable to segregate by cohort - as an example, if we run the experiment for a month, we can filter only students who have enrolled in the past 15 days.