Potential Program Insights:
An Experiment Analytics Report

Abstract

In this report, our analysts used student score data from several districts to analyze the impact of different versions of the Skills Advantage program. We used the statistical tool regression analysis to confirm the significance of the project's impact. Through analysis, it was found that Program A and Program B have their own strengths in different skills, and the potential benefit is relatively high. We recommend choosing Program A and Program B to update the Skills Advantage program based on the different paths the organization wants to develop in the future.

Introduction

Survey on Course Preferences:
The CEO wants to learn more about the views and demographics of their clients in order to make informed judgments about growing their online educational offerings. This comes after reviewing the team's study of past test score data from their Refresh and Tutoring courses. After the holidays in January, a survey is scheduled. The group is entrusted with creating three to five inquiries concerning the opinions and demographics of the clientele, together with the question structure, categories for responses, and guidelines for participants. On November 30, these queries and a succinct explanation of each should be brought before management.

Analysis of the SkillsAdvantage Curriculum Experiment:
Three possible new educational programs for the SkillsAdvantage course have been selected by the Curriculum Review Committee. A test market consisting of several districts that only used one of the new curricula was used in the experiment. The results and supporting paperwork from the experiment, which are attached to the memo, are to be analyzed by the team. They are supposed to offer an overview of the findings and suggest an instructional plan. All of the districts that were served by the trial were administered, but the team is free to choose which districts to include in their analysis.

Based on your objective to identify the most promising among three prospective versions of the SkillsAdvantage course, we conducted a comprehensive analysis of student enrollment and post-program performance data spanning from August to October 2023, as provided by your organization.

Given the availability of performance metrics for program versions A, B, C, and the current version across eight school districts, our primary aim was to assess the potential benefits of these programs and offer strategic insights.

To commence our analysis, we initiated data cleaning procedures, ensuring data integrity and accuracy. Subsequently, we assessed whether student scores exhibited any patterns among the school districts. This assessment informed our decision-making process regarding the inclusion of specific districts and the consideration of random assignment for program versions.

During the analysis, our initial focus was on evaluating the average performance improvements in each district. Following this, we delved deeper into the distinct improvements observed in the four key skill areas. Based on our comprehensive analysis, we have formulated strategic recommendations to guide your decision-making process.

Pre-experiment Preparation

Before we began designing our experiment, we did some pre-experiment questioning in order to make sure that our experiment took into account the factors that could affect the accuracy and validity of the results.

Pre-experiment Question 1: **Which districts to include in the experiment?**
Since programs are being assigned at the district level, the programs are intended to impact the individual student's score. We begin by analyzing if all districts must be included in the analysis.

Method:
For this, we test the relationship between the total scores and the district.
For this, we run a regression where total_scores = reading score + writing score + math No Calc score + math Calc. Scores for each student are the dependent variable and district, which is categorical in nature, is taken as the independent variable.

Regression Results:

```
Call:
lm(formula = diff_totalscore ~ district, data = merged_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-63.102  -9.802  -0.567   9.741  66.898 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)          44.5862     0.4170 106.916   <2e-16 ***
districtCamas        65.6703     0.6962  94.322   <2e-16 ***
districtLake Oswego  48.9811     0.6137  79.815   <2e-16 ***
districtRidgefield   48.0346     0.6095  78.814   <2e-16 ***
districtRiverdale   -36.3271     0.7232 -50.233   <2e-16 ***
districtSherwood     60.5156     0.5773 104.817   <2e-16 ***
districtSt. Paul     -1.5803     0.6758  -2.338   0.0194 *  
districtWilsonville -18.7842     0.5880 -31.948   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.53 on 9027 degrees of freedom
  (9492 observations deleted due to missingness)
Multiple R-squared:  0.8378,	Adjusted R-squared:  0.8376 
F-statistic:  6659 on 7 and 9027 DF,  p-value: < 2.2e-16
```

Interpretation and Usage:
Since the estimates (given in comparison to the benchmark district) are significant for all districts in the analysis, we conclude that the relationship between total scores and the districts is significant for all districts. Thus, we include all districts in the experiment.

Pre-experiment Question 2: **Were the different educational programs (A/B/C/Current) randomly assigned to the districts?**
We understand that the average score at the intake of students is different for different districts, we thus know that there is a difference in the average intelligence levels of districts already. It is important to ensure that there is no relationship between the district performance and the program assigned to it.

Method:
For this purpose, we calculate average total scores across districts and the educational programs assigned and try to observe if there exists any pattern.

Results:

district	average_score	educational_program
Sherwood	1565.723	B
Ridgefield	1641.781	A
Camas	1662.496	B
Wilsonville	1673.442	C
St. Paul	1676.189	Current
Lake Oswego	1676.303	A
Riverdale	1690.316	C
Beaverton	1750.123	Current

Interpretation and Usage: We observe no definite pattern between the average score of each district and the educational program assigned to them, thus concluding that programs are randomly assigned and are not biased by initial average performance.

After seeking results for the pre-experiment questions we perform an **exploratory analysis to see if the programs are effective; We also see if different programs have different impacts (measured through average score improvement).**

Method:
We calculate the average improvement in total score for each student and average it over the district. We then see if there is any evident pattern in the program assigned to each district and their average improvement.

Results:

district	average_score_improvement	educational_program
Camas	110.256443	B
Sherwood	105.101785	B
Lake Oswego	93.567227	A
Ridgefield	92.620803	A
Beaverton	44.586157	Current
St. Paul	43.005862	Current
Wilsonville	25.801994	C
Riverdale	8.259045	C

Interpretation and Usage:
We find that the performance of districts is related to programs since the top two performing districts belong to program B, followed by the next two performing districts belonging to A, the next two performers in descending order belong to the current program and finally, the last two improvement scores belong to program C. This pattern is a hint at the different impacts of the programs but we need advanced analysis to firmly conclude the relationship.

This is the end of our exploratory analysis and it lays the foundation for our main experiment design.

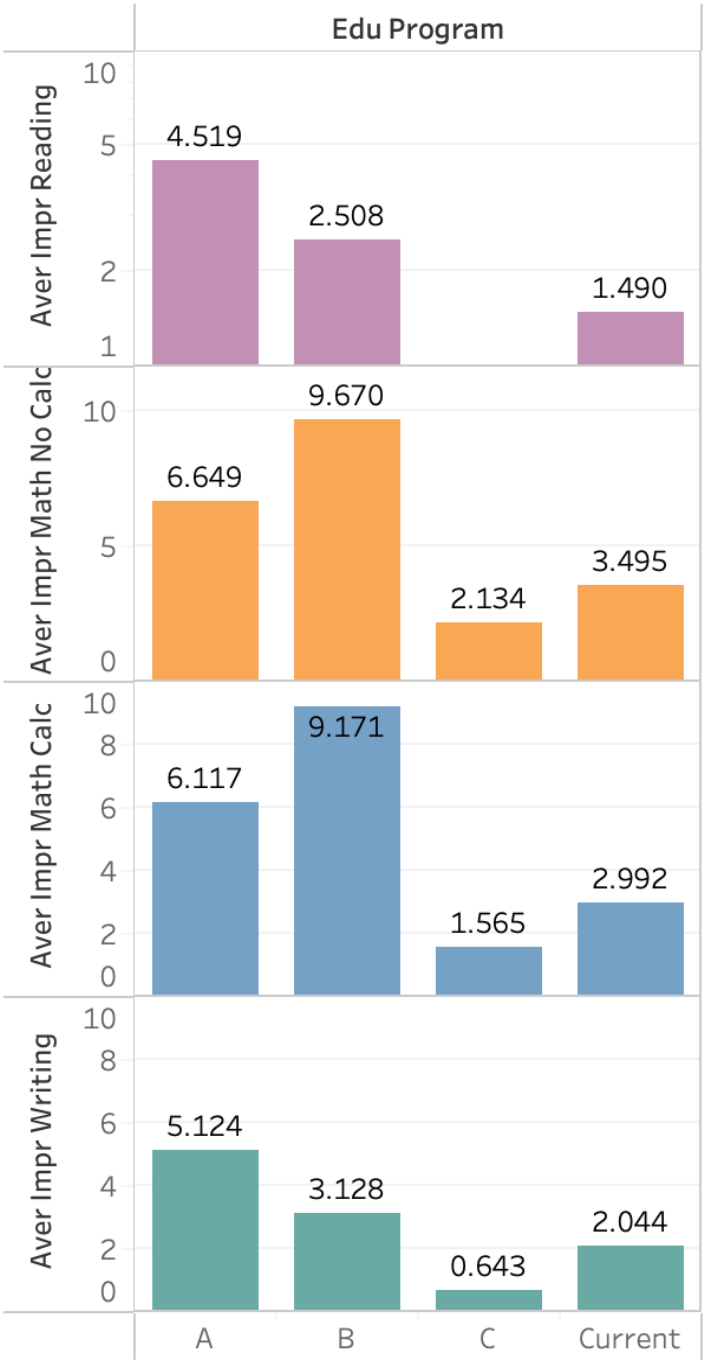
Experimental Methodology

Experiment Question: **For 4 specific skills (Reading, Writing, "calculator" math, "no calculator" math), what impact do different versions of the project have on them?**

Method:
We took an in-depth look at the impact of the four versions of the program on the four skills. First, subtract each student's intake score from their post-project score and divide by the intake score to find improvement. As we analyzed in the pre-experiment question, we will use student scores in all districts as participant groups. Analyze at the individual and skills level. In order to prove the significance of the results, we used the individual skill scores of all students as dependent variables and the four programs A/B/C/Current as independent variables for regression.

Results

The picture below shows the average score improvement of different skills under different versions of A/B/C/Current.



The results below show the regression results of four skills scores for different projects.

```
Call:
lm(formula = imprReading ~ eduProgram, data = imprScore)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0675 -0.7026  0.0056  0.7031  3.7300 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        4.51915    0.02148  210.39   <2e-16 ***
eduProgramB       -2.01137    0.03078  -65.35   <2e-16 ***
eduProgramC       -4.36348    0.03150 -138.51   <2e-16 ***
eduProgramCurrent -3.02895    0.03095  -97.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.055 on 9031 degrees of freedom
Multiple R-squared:  0.6964,	Adjusted R-squared:  0.6963 
F-statistic:  6906 on 3 and 9031 DF,  p-value: < 2.2e-16
```

Reading Skills

```
Call:
lm(formula = imprWriting ~ eduProgram, data = imprScore)

Residuals:
   Min     1Q Median     3Q    Max 
-7.310 -0.689 -0.124  0.986 47.956 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        5.12392    0.03390  151.13   <2e-16 ***
eduProgramB       -1.99628    0.04858  -41.09   <2e-16 ***
eduProgramC       -4.48063    0.04972  -90.11   <2e-16 ***
eduProgramCurrent -3.07945    0.04886  -63.03   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.665 on 9031 degrees of freedom
Multiple R-squared:  0.4925,	Adjusted R-squared:  0.4923 
F-statistic:  2921 on 3 and 9031 DF,  p-value: < 2.2e-16
```

Writing Skills

```
Call:
lm(formula = imprMathCalc ~ eduProgram, data = imprScore)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3526 -0.7356  0.0201  0.7391  4.5446 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        6.11679    0.02230  274.26   <2e-16 ***
eduProgramB        3.05456    0.03196   95.58   <2e-16 ***
eduProgramC       -4.55209    0.03271 -139.17   <2e-16 ***
eduProgramCurrent -3.12474    0.03214  -97.23   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.095 on 9031 degrees of freedom
Multiple R-squared:  0.8761,	Adjusted R-squared:  0.876 
F-statistic: 2.128e+04 on 3 and 9031 DF,  p-value: < 2.2e-16
```

"Calculator" Math Skills

```
Call:
lm(formula = imprMathNoCalc ~ eduProgram, data = imprScore)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6716 -0.7175  0.0130  0.7250  4.6159 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        6.64882    0.02199  302.41   <2e-16 ***
eduProgramB        3.02096    0.03150   95.89   <2e-16 ***
eduProgramC       -4.51433    0.03224 -140.00   <2e-16 ***
eduProgramCurrent -3.15356    0.03168  -99.54   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.08 on 9031 degrees of freedom
Multiple R-squared:  0.8779,	Adjusted R-squared:  0.8778 
F-statistic: 2.164e+04 on 3 and 9031 DF,  p-value: < 2.2e-16
```

"No Calculator" Math Skills

Discussion

We can see that in terms of reading skills, program A helps students the most, while program C has little help (Almost 0). Through regression analysis, the same results can be obtained. The baseline in the regression is program A, which means that if converted to program B, students' progress in reading skills will be reduced by 2.01, by 3.92 by the current program, and by 4.36 by program C.

For writing skills, program A also brings the greatest improvement. Our return confirmed this. However, it is worth noting that program B's performance in mathematics skills is that program B can maximize students' performance in both "Calculator" Math and "No Calculator" Math skills. In the regression, we can also see that the coefficient of program B is positive, which means that compared to program A, program B can bring greater improvement, and the result is statistically significant at the 5% level. This means that program B does bring greater improvement than program A.

One potential source of bias can be due to the absence of information regarding the class size of the programs. We say this because the number of students enrolled in a program A/B/C/Current is not equally divided between the states and is extremely unequal.

district	distinct_student_ids	district	distinct_student_ids	district	distinct_student_ids	district	distinct_student_ids
Lake Oswe	1309	Camas	845	Riverdale	699	Beaverton	1469
Ridgefield	1290	Sherwood	1566	Wilsonville	1432	St. Paul	882
Total A	2599	Total B	2411	Total C	2131	Total Current	2351

Since there is no information about the class size of each program or any standard set for the student-faculty ratio by CLC, the test scores of the students can also be affected based on the class size within a district. This could be a potential source of bias in our analysis.

Conclusion

Key Findings

Program A delivered the highest improvement results in Reading and Writing, while Program B delivered the highest improvement results in both Math with and without a calculator. The Current program had mediocre results in all subjects. The significance of these findings is that we can now understand the key strengths of the programs and optimize one to perform best on all counts. It would also give us an insight into what seems to be working for each subject.

Research Question: **For 4 specific skills (Reading, Writing, "Calculator" Math, "No Calculator" Math), what impact do different versions of the project have on them?**

Our analysis answers the research question as it tells us that Program version A has the highest impact on Reading and Writing while B has the highest on both kinds of Math courses.

Recommendations

We would like to give our recommendations. We have suggestions in three directions. First, if you want to be in an institution where mathematics is your strength in tutoring, please choose program B. Second, if you want to become a tutoring organization that specializes in reading and writing, please choose program A. Finally, if you want all-round development, please choose program A and further study how to transfer the mathematical strengths of program B to program A to make it a relatively all-round tutoring program.

Appendices

Part 1: Survey questions

Before answering the survey:

This survey is intended for both students and parents. If you are a parent completing this form, please provide responses based on your child's information.

Client Demographics

Q1: What is your gender?

Response Categories:

- Female
- Male
- Non-binary
- Prefer not to say
- Other (please specify)

Instructions: Please select the option that best describes your gender identity. Your response will help us understand the demographic diversity of our student population.

Explanation: This question is designed to gather the gender identity of students. This data is valuable for tailoring educational resources and marketing strategies for CLC ‘s online courses.

Q2: What is your current education level?

Response Categories:

- 7th Grade
- 8th Grade
- 9th Grade
- 10th Grade
- 11th Grade
- 12th Grade
- Other (please specify)

Instructions: Select the option that best describes your current level of education. If your education level is not listed, please select 'Other' and provide a brief description.

Explanation: We chose to replace the common ‘age question’ with education level. This is because students at the same level of education may be of different ages. And CLC only needs to know what the customer needs to learn, not their age.

Q3: Which school district do you attend the most?

Response Categories:

- St. Paul
- Beaverton
- Lake Oswego
- Ridgefield
- Camas
- Sherwood
- Riverdale
- Wilsonville
- I only enrolled in the online courses.

Instructions: Please select the school district you attend most often from the list, and if there are two or more, choose the one you prefer.

Explanation: We currently have data on each student's district. However, we do not have the number of students who only take online courses. This question can help us find out this aspect.

Q4: How many children are there in your household?

Response Categories:

- 1
- 2
- 3
- 4
- 5 or more

Instructions: Please select the total number of children in your household, including yourself. This includes any stepsiblings, half-siblings, or adopted siblings living in the same household and in grades 7 to 12.

Explanation: This question aims to gather information about the size of student households in terms of the number of children. This information can be an asset for CLC in making their marketing decisions for online courses.

Client attitudes

Q1: Are you enrolled in our online program or center-based program?

Response Categories:

- Online Program
- Center-Based Program
- Both

Instructions: If you are enrolled in both the online and center-based programs, select "Both" and please answer all questions below accordingly.

Explanation: This question can help us ask questions for students in different segmentation, improving the efficiency of the survey.

<u>If Online:</u>

1. **Why did you choose online courses?** (Select all that apply.)

Response Categories:

- Flexibility in scheduling
- No commuting required
- Accessibility from any location
- Preference for digital learning environments
- Course availability
- Health and safety concerns
- Other (please specify)

Explanation: Understanding why current students choose online courses can help CLC understand how to transition more students to online programs.

<u>If Center-Based:</u>

1. **How willing are you to take online programs?**

Response Categories:

- Very willing
- Somewhat willing
- Neutral
- Somewhat unwilling

- Not willing at all

Explanation: This information is crucial for making informed decisions about resource allocation between online and center-based programs.

2. **On average, how long is your commute to our learning center?**

Response Options:

- Less than 15 minutes
- 15 to 30 minutes
- 31 to 60 minutes
- Over an hour

Explanation: Shorter commute times might indicate higher accessibility and potentially greater satisfaction with center-based learning, whereas longer commutes might reveal a preference or need for online learning alternatives. This data can be instrumental in making decisions about the location of future centers, scheduling, and the allocation of resources between center-based and online course offerings.